ISSN 2346-8092

ა. რაზმაძის მათემატიკის ინსტიტუტის შრომები

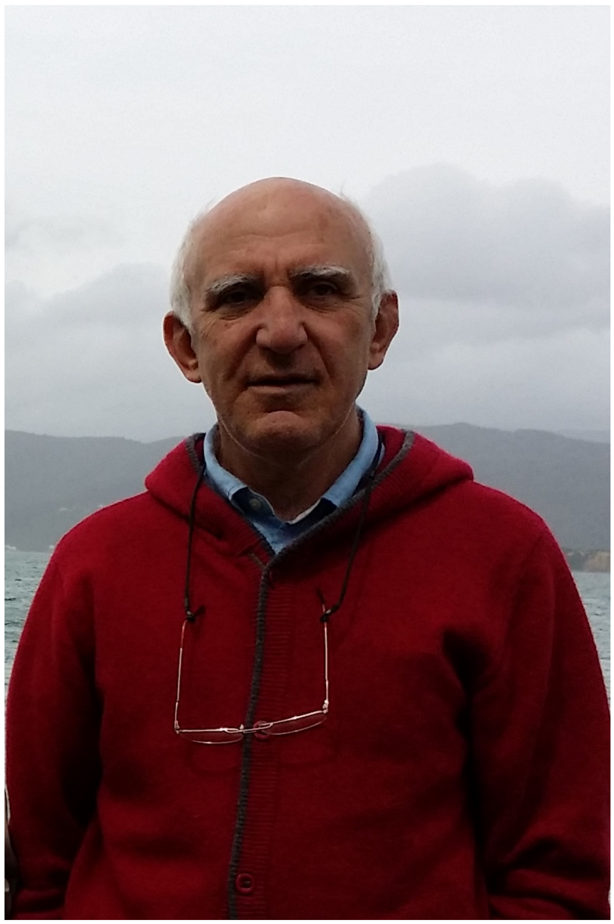
ივანე ჯავახიშვილის სახელობის თბილისის
სახელმწიფო უნივერსიტეტი

ტომი 174, N2, 2020

Transactions of A. Razmadze Mathematical Institute is a continuation of Travaux de L' Institut Mathematique de Tbilisi, Vol. 1–15 (1937–1947), Trudy Tbilisskogo Matematicheskogo Instituta, Vol. 16–99 (1948–1989), Proceedings of A. Razmadze Mathematical Institute, Vol. 100–169 (1990–2015), Transactions of A. Razmadze Mathematical Institute (published by Elsevier), Vol. 170–172 (2016–2018).

CONTENTS

This Volume is dedicated to 75-th
Anniversary of Professor Estate Khmaladze,
Foreign Member of the National Georgian Academy of Sciences,
Fellow of the Royal Society (Academy) of New Zealand,
Fellow of the Institute of Mathematical Statistics



Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 127–134

HIGHLIGHTS OF RESEARCH WORK OF ESTATE KHMALADZE

ROBERT M. MNATSAKANOV

The goodness of fit problem and scanning martingales. Here we briefly outline one of the central problems of mathematical statistics, the difficulties which remained open there from the mid 50s to the early 80s and the way they were overcome using very unexpected "martingale approach" developed by Khmaladze (1981), as well as its nontrivial extension to the case of multidimensional time.

Let F_n be an empirical distribution function of a sequence of n independent scalar random variables with distribution function F. The normalised difference $\sqrt{n}(F_n - F) = v_n$, one of the basic processes in statistical theory, is called the empirical process.

Certainly, both the distribution of the process v_n for finite n and its limiting distribution as $n \to \infty$ depend on F (the limiting process v is called F-Brownian bridge). However, the amazingly simple Kolmogorov's transformation (Kolmogorov (1933)), $u_n = v_n \circ F^{-1}$ with the condition that F is continuous, maps v_n into the so-called uniform empirical process u_n with the standard (and independent of F) distribution. This opens an extremely important possibility to use asymptotic theory for u_n *only* in asymptotic statistical inference concerning *any* continuous F. This "asymptotic distribution freeness" of u_n became one of the basic facts in nonparametric statistics and in the theory of the so called goodness of fit tests.

In the late 1950's and early 1960's it was discovered (see Kac, Kiefer and Wolfowitz (1955) or Gikhman (1953)) that in most practical cases, where $F = F_\theta$ is known only up to a finite dimensional parameter θ to be estimated from the data, the process $v_{n,\hat{\theta}} = \sqrt{n}(F_n - F_{\hat{\theta}})$ has the asymptotic distribution not only different from that of F-Brownian bridge, but also such that $v_{n,\hat{\theta}} \circ F_{\hat{\theta}}^{-1}$ is no longer asymptotically distribution free. The bibliography on this subject is huge; one review paper is Durbin (1973). Chibisov (1971) and Moore and Spruill (1975) demonstrated that the chi-square statistics with estimated parameter is, in general, also not asymptotically chi-square distributed. All developments in the late 1960's and throughout the 1970's persuaded statisticians that this was an unavoidable complication which needed to be lived with.

However, Khmaladze's paper (1981) changed this stereotype completely. It was shown that using a different point of view on v_n, a transformation of $v_{n,\hat{\theta}}$ can be found, $w_{n,\hat{\theta}} = \sqrt{n}(F_n - K_{\hat{\theta},n})$, which converges to F_θ-Brownian motion, and therefore $w_{n,\hat{\theta}} \circ F_{\hat{\theta}}^{-1}$, is asymptotically a standard Brownian motion on $[0, 1]$, and thus asymptotically distribution free. The process $w_{n,\hat{\theta}}$ can be thought of as the innovation martingale of the process $v_{n,\hat{\theta}}$ with respect to the natural filtration of the later. In this way, the whole beauty and usefulness of asymptotically distribution free procedures were restored.

Further work of E. Khmaladze developed similar transformations in the difficult case of empirical processes based on multidimensional random variables. As is know, the existing theory of martingales in multidimensional time is complicated and involves restrictive conditions, not satisfied by many important processes with multidimensional time (cf., e.g., Wong and Zakai (1974), Cairoli and Walsh (1975), Hajek and Wong (1980), Gikhman (1982), Nualart (1983)). Therefore a new approach to the stochastic calculus of Gaussian random processes with multidimensional time was required. The notion of "scanning martingales", suggested by Khmaladze (1988a, 1993), provides such an approach and leads to an elegant and simple theory of innovation martingales in multidimensional spaces. In Khmaladze (1988a), the general goodness of fit problem for simple hypothesis was formulated for the first time, and Khmaladze (1993) gives the solution of this problem in a completely general setting and opens a way to distribution free "model testing" in multidimensional spaces, the possibility, previously nonexistent in the statistical theory.

Mathematically, the paper establishes new connections between the goodness of fit problem of statistics and empirical processes with functional time on one side, and the theory of stochastic differential equations for measure-value processes and Volterra decompositions of Hilbert-Schmidt operators on the other side.

The paper of Einmahl and Khmaladze (2001) gives a similar solution in $\mathbb{R}^d$ for another classical problem of statistics, the so- called two-sample problem in a multidimensional space.

Sequential ranks. The sequential rank S_k of a random variable X_k is its rank among random variables $X_1, X_2, \ldots, X_k$ as compared with the "ordinary" rank R_{kn} which is the rank of X_k amongst all n "available" random variables $X_1, X_2, \ldots, X_k, \ldots, X_n$, with $k \leq n$. Sequential ranks are practically very convenient when observations arrive one-by-one. However, their asymptotic theory meets with certain difficulties; it was not known how to study the efficiency of statistical procedures based on sequential ranks. Consequently, this theory fell into disuse, whereas the theory of "ordinary" ranks received considerable attention in the 1960's through to the 1980's. For example, in Sen (1978), although primarily devoted to sequential problems, it was necessary for the authors to work with "ordinary" ranks, which are inconvenient in this setting, because asymptotic methods were unavailable for sequential ranks.

On the other hand, if $X_1, X_2, \ldots, X_n$ are independent and identically distributed, the sequential ranks are of very nice behaviour: $S_1, S_2, \ldots, S_n$ are independent and each S_k has uniform distribution on integers $1, 2, \ldots, k$.

The difficulties connected with efficiency of the tests based on sequential ranks were overcome in the papers of Khmaladze and Parjanadze (1986), Pardzhanadze and Khmaladze (1986), which established an asymptotic theory of sequential ranks in the same basic framework as the existing theory for "ordinary" ranks. This was possible due to the development of asymptotic methods, not normally applied in the theory of rank statistics. Namely, it was shown that the partial sum processes based on (functions of) "ordinary" ranks and those based on sequential ranks may be asymptotically connected through a linear stochastic differential equation and hence the properties of one can be carried over into the properties of another.

In particular, it was shown that asymptotic distributions of linear statistics

$$\sum_{k=1}^{n} c_k a(R_{kn}/n)$$

from "ordinary" ranks and linear statistics

$$\sum_{k=1}^{n} \left(c_k - \sum_{m \leq k} c_m / k \right) a(S_k/k)$$

from sequential ranks have the same asymptotic distribution under *all* contiguous alternatives. Equivalently, statistics

$$\sum_{k=1}^{n} c_k a(S_k/k) \quad \text{and} \quad \sum_{k=1}^{n} \left(c_k - \sum_{m \geq k} c_m / m \right) a(R_{kn}/n)$$

have the same limit distributions under all contiguous alternatives. Thus, whenever one of them is optimal against some contiguous alternative, the other is also optimal for the same alternative.

Multinomial distributions of increasing dimension. The research in this field may be of interest to the colleagues in discrete mathematics.

The sequence of multinomial distributions $\mathcal{M}(\cdot, p_N, n)$, where $p_N = \{p_{in}\}_1^N$, $p_{in} > 0$ and $\sum_1^N p_{in} = 1$, which have the number of different possible outcomes $N = N_n$ increasing with number of trials n, form a surprisingly rich class of distributions. They reflect and illustrate a very large number of interesting problems found in other parts of statistics, such as:

• the asymptotic behaviour and properties of statistics like the classical χ^2-statistic when $N_n \to \infty$ as $n \to \infty$ are sharply different from those which one can deduce when first letting $n \to \infty$ and then $N \to \infty$;

• statistical problems with increasing numbers of parameters, like the problem of estimating spectra of matrices of increasing dimension or problems with "fine" partitions, are very similar to what one meets in the asymptotic analysis of $\mathcal{M}(\cdot, p_N, n)$ — the normalised probabilities np_{in}, $i = 1, \ldots, N_n$, are these parameters;

• the class of asymptotic laws of which the famous Zipf - Mandelbrot's law is the most remarkable representative, are highly connected with the sequences $\mathcal{M}(\cdot, p_N, n)$ where $n \to \infty$. We comment on Zipf – Mandelbrot's law separately below.

The paper of Khmaladze (1983) completely modified the tools and approaches used in this field. Instead of considering sums, called "divisible statistics",

$$\sum_{i=1}^{N_n} g(\nu_{in}, np_{in})$$

and limit theorems for each sum, the paper studied partial sums

$$\sum_{i=1}^{k} g(\nu_{in}, np_{in}),\ k = 1, 2, \ldots, N_n,$$

and derived limit theorems for these processes. They are treated as semimartingales associated not with its natural filtration, but with richer filtration $\mathcal{F}_k = \sigma\{\nu_{1n}, \ldots, \nu_{kn}\}$, $k = 1, 2, \ldots, N_n$, based on underlying frequencies. The point of it is that the conditional distribution of ν_{in} given previous frequencies ν_{jn}, $j = 1, 2, \ldots, i-1$, is much simpler object, than conditional distribution of $g(\nu_{in}, np_{in})$ given previous summands $g(\nu_{jn}, np_{jn})$, $j = 1, 2, \ldots, i-1$.

It showed how new at a time limit theorems for semimartingales could be utilised and lead to general functional limit theorems for the basic statistics of the field – the so-called additively divisible statistics (statistics of increasing numbers of small, separate frequencies). The paper demolished an unnecessary partition between different parts of asymptotic statistics (for a better picture, see Ivchenko and Levin's review paper (1996)). It led to similar advances in the theory of general spacings (see Borovikov (1987)) and in the analysis of the so-called "very rare events" (see Mnatsakanov (1985) or Prakasa Rao (1987)).

Large number of rare events (LNRE) theory. The text of Dante's "Divina Comedia" is in length some 100,000 words. Approximately 13,000 of these words are different, that is, the vocabulary of "Divine Comedia" is only 13,000 words. It would be, however, very incorrect to suppose that each word was used by Dante approximately 8 times. There is certainly nothing like an even usage of words, few words were used hundreds of times, while about 6,000 words (half the vocabulary) were used only once and about 2,000 words were used only twice.

This is the typical situation in a surprisingly large number of applied statistical problems, not only in all sorts of large texts, but also in studies of the number of species in an environment, opinions in a survey, chemical analysis, income distributions, distribution of languages, etc.

According to Zipf's law, if $\mu_n(m)$ is the number of words (species, opinions, etc.) which occurred m times in a sample of size n and if μ_n is the number of all different words (species, opinions, etc.) in the same sample, then

$$\frac{\mu_n(m)}{\mu_n} \to \frac{1}{m(m+1)} \quad \text{as} \quad n \to \infty.$$

Its slightly modified form

$$\frac{\mu_n(m)}{\mu_n} \to \frac{1}{(a+bm)^q} \quad \text{as} \quad n \to \infty$$

is called Zipf - Mandelbrot's law. In the words of Mandelbrot (1953), "The form of Zipf's law is so striking and also so very different from any classical distribution of statistics that it is quite widely felt that it "should" have a basically simple reason, possibly as "weak" and general as the reasons which account for the role of Gaussian distribution. But, in fact, these laws turn out to be quite resistant to such an analysis. Thus, irrespective of any claim as to their practical importance, the "explanation" of their role has long been one of the best defined and most conspicuous challenges to those interested in statistical laws of nature".

The present interest in Zipf's law is, perhaps, characterised by the increase of interest in the more general concept of LNRE distributions, which was introduced and first systematically studied in an unpublished paper of Khmaladze (1988b), partly reproduced by Khmaladze and Chitashvili (1989) and called and treated as "fundamental" in the monograph of Baayen (2001).

"Chimeric" contiguous alternatives. The theory of contiguity of probability measures is a main tool in asymptotic statistics to study the efficiency and power of statistical procedures. Contiguous distributions or contiguous alternatives (to a given distribution) form a class of alternatives which are, heuristically speaking, most difficult to detect. It is well known that if $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are two distributions, then n-fold direct products $\mathbb{P}^{(n)}$ and $\tilde{\mathbb{P}}^{(n)}$ are either asymptotically singular as $n \to \infty$ or coincide (alternative of Kakutani). In order for $\tilde{\mathbb{P}}^{(n)}$ to be contiguous to $\mathbb{P}^{(n)}$, the distribution $\tilde{\mathbb{P}}$ must depend on n in such a way, basically, that

$$\left(\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\right)^{1/2} = 1 + \frac{1}{\sqrt{n}} h_n \quad \text{with} \quad \limsup_{n\to\infty} \|h_n\|_{L_2(\mathbb{P})} < \infty$$

(see, e.g., Oosterhof and van Zwet (1979)). Practically all papers which use contiguity theory replace the latter condition by $h_n \overset{L_2(\mathbb{P})}{\longrightarrow} h$ and for very good reasons. Nevertheless, the paper of Khmaladze (1998) studies such contiguous alternatives that $\|h_n\| \geq 1$, but $h_n \overset{\omega}{\to} 0$, that is, h_n has no limiting points in $L_2(\mathbb{P})$. It is clear that no classical goodness of fit test based on empirical process can detect any such "chimeric" alternative. Yet the paper of Khmaladze (1998) shows that new versions of empirical processes can be constructed and a goodness of fit theory can be developed which is no less rich than that which exists for converging contiguous alternatives.

The paper also shows that although they look exotic, "chimeric" alternatives can frequently be found in real problems. After all, the existence of our civilisation is itself an enormous "chimeric" alternative.

Change-set problem (spacial change-set problem). The idea of transferring the range of problems usually unified by the term "change-point problem" for real line to finite-dimensional Euclidean space was entertained and discussed by E. Khmaladze at the end of 80-ies, while still in Moscow, in particular, at the Moscow Seminar of Young Statisticians. But he started working himself only in 1996.

With the help of the concept of the local covering numbers, the papers of Khmaladze, Mnatsakanov and Toronjadze (2006a, 2006b) investigated the convergence of statistical estimators of the change-set and finally obtained the correct rate of n^{-1}. This is the rate of convergence of what is called "super-efficient" estimators in statistics. The smoothness of the boundaries were not required – only that the class of possible change-sets was locally compact.

Differentiation of set-valued functions. The research work of E. Khmaladze here has a story. Heuristically, the idea came from the work on the change-set problem. In this problem, the object of interest is a set, say $A \in \mathbb{R}^d$ as a hypothesis, and a sequence of sets $B_n \in \mathbb{R}^d$, converging in Hausdorff metric to A as a sequence of contiguous alternatives. What a statistician observes is a point process N_n in $\mathbb{R}^d$, and, as $n \to \infty$, the intensity of this point process increases, so that there appear more and more points, and symmetric difference $A\Delta B_n$ shrinks towards the boundary of A at the same time. Since the number of points increases, it is not necessary that their number in $A\Delta B_n$ decreases to zero. In the most interesting cases this number becomes a Poisson random variable. So, the corresponding points do not disappear; but where do they eventually "live"? The first intuition was that they must "live" on the boundary of A. But later the feeling grew that this should not be true. Some sort of "differentiation" seems to be lurking behind the scenes.

All this were talks and guesses, and some reading for several years. The actual work started after 2004. In Karlsruhe, a very good colleague and very highly regarded geometer, Wolfgang Weil, came one day to the small temporary library room, where Estate was accommodated, and put down the famous book of J-P. Aubin and E. Frankowska (1990) on a set-valued analysis. The book contained two chapters on differentiation of set-valued functions. And this meant that secure, but not very original work, lied ahead. W. Weil was not too enthusiastic about this prospect. And in spite of joint paper of Khmaladze and Weil (2008), which was then the work in progress, Estate was on his own.

A year later, the draft of the paper on differentiation of set-valued functions was ready and in 2007 the paper Khmaladze (2007) was published. One short corollary of the new notion of a derivative is the following:

• if, as $t \to 0$, the symmetric difference $A\Delta B_t$ is differentiable at the boundary ∂A, then for any absolutely continuous distribution $\mathbb{P}$ in $\mathbb{R}^d$ there exists an absolutely continuous distribution $\mathbb{Q}$ on the normal cylinder of ∂A such that

$$\frac{d}{dt}\mathbb{P}(A\Delta B_t) = \mathbb{Q}\Big(\frac{d}{dt}B_t\Big).$$

Later, a review article with W. Weil was invited to the Annals of the Institute of Statistical Mathematics, Khmaladze and Weil (2018), where the derivative was given a name of "fold-up derivative" and was defined in more general class of situations. Before that, a paper with John Einmahl established CLT for the point process N_n on classes of sets in shrinking neighbourhoods of ∂A (see Einmahl and Khmaladze (2011)).

In a personal letter to E. Khmaladze, J.-P. Aubin calls the work of fold-up derivatives a "mathematical virtuosity".

Questionnaires – the problem of diversity in spaces of increasing dimension. A person is asked q binary questions: "yes" or "no". The person fills in this questionnaire and this is one "opinion". Altogether 2^q different opinions are possible. Lots of people, N, are asked to fill this questionnaire. So, there are lots of questionnaires with many possible opinions expressed in them. Exactly, how many different opinions will be found in the sample? How many opinions will be unique? These and all other similar questions have been answered by Khmaladze (2011). But the answers did not come without surprise also for the author.

Imagine, again, that we are in $[0,1]^q = [0,1] \times \cdots \times [0,1]$, and we divide the first interval $[0,1]$ in proportion $a_1 : 1-a_1$, the second $[0,1]$ as $a_2 : 1-a_2$, and so on. In this way one will obtain 2^q elementary cubes. In one-dimensional space, each of subintervals $[0,a_1]$ and $[a_1,1]$ is divided as $a_2 : 1-a_2$, then each of the resulting four are divided as $a_3 : 1-a_3$, and so on, q times. The first impression was that this would be some other version of random partition of a "stick" into 2^q subintervals: if $0 < U_1 < U_2 < \cdots < U_{2^q-1} < 1$ are uniformly distributed random variables, arranged in increasing order, then the spacings $[U_i, U_{i+1}]$ are forming this random partition. If we now throw N random points on $[0,1]$, or in $[0,1]^q$, and count frequencies of these points in each subinterval, or small cubes, what will be their behaviour?, how many subintervals will remain empty?, how many will contain just one point?, etc. For a random partition, the answers are more or less known and initially Estate wanted an analogue of this.

However, the behaviour of these frequencies turn out to be very different. Very uneven. Behaviour of spacings is, strictly speaking, also "uneven", but not so far from being even. But sizes of intervals, or volumes of cubes, obtained through these a_i-s are sharply uneven. And behaviour of the frequencies of random points in them, consequently, is also uneven. First of all, the number of cells with some filling turns to be $o(N)$, i.e., much smaller than the number of points thrown; or the number of different opinions is much smaller, than the number of persons asked. The fraction of cells with one, two, and in general k points in them, relative to the number of all non-empty cells, follows some "law", which "almost" does not depend on the choice of a_i-s. Yet, this law is not the famous Zipf's law, which many of us could have heard about.

A complete description of the situation is given by Khmaladze (2011) and partly in the 18-years earlier paper of Khmaladze and Tsigroshvili (1993). This was a strong step forward within the theory of diversity and occupation problem. Division in more than two subintervals at each step is a fascinating problem for the future.

Unitary operators. The last several years a new development took place in the direction of distribution free testing theory. The main idea can be explained as follows.

The empirical process with estimated parameter $v_{n,\hat{\theta}}$, or estimated empirical process for short, is not just a process with different limit distribution from the empirical process v_n, it has the specific structure – its limit distribution is that of the projection of F_θ-Brownian bridge, orthogonal to the score function $\dot{f}_\theta/f_\theta$. Thus the distribution of the projected F_θ-Brownian bridge is dependent on this score function. This asymptotic phenomenon was first described by Khmaladze (1979). It implies that for any regular parametric family of distributions $G_\theta, \theta \in \Theta$, as a limit of the estimated empirical process, one will obtain again a projection of G_θ-Brownian bridge, orthogonal to a corresponding score function. However, if the dimension of parameters in both families is the same, than with the help of unitary mappings one projection can be mapped to another projection thus rendering the two testing problems equivalent, in the sense that one can be transformed into other and the other way around.

Convenient framework for application of operators on empirical processes is provided by the function-parametric version of empirical processes

$$v_{n,\hat{\theta}}(\phi) = \int \phi(x) v_{n,\hat{\theta}}(dx), \quad \phi \in L_2(F_\theta),$$

because then the operator U on $v_{n,\hat{\theta}}$ can be naturally defined as the adjoint operator U^* on $L_2(F_\theta)$:

$$(Uv_{n,\hat{\theta}})(\phi) = v_{n,\hat{\theta}}(U^*\phi).$$

However, this notion of equivalence creates very wide classes of equivalence, and in each class one needs only one representative, for which the distributional work for test statistics should be carried through; this is no different to assuming that the sample came from uniform distribution while testing simple hypothesis.

The projections, as a result of estimation of parameters, are ubiquitous. They appear in situations where so far nobody considered testing problems. Estimation of parameters - yes, but not testing, in particular, not goodness of fit testing. From the families of discrete distributions, for which the goodness of fit testing theory appeared only in 2013 (see Khmaladze (2013)), to empirical processes in regression, and now testing models for point processes (see the article of Khmaladze (2020) in this issue), testing parametric hypothesis for Markov chains and for Markov diffusion processes, like the Ornstein – Uhlenbeck process, all are work in progress.

In this account of scientific contribution of Estate Khmaladze in statistics and stochastic models we do not comment on the other fields of his research such as

– kernel density estimators,
– asymptotic of non-crossing probabilities with moving boundaries,
– formulation of the strong law of large numbers for Voronoi tessalation,
– extreme value theory and record processes

and various others. One can find these, e.g., in Mnacakanov and Hmaladze (1981), Kotel'nikova and Khmaladze (1982), Khmaladze, Nadareishvili and Nikabadze (1997), Khmaladze and Shinjikashvili (2001), Khmaladze and Toronjadze (2001) (see also Schneider and Weil (2008)), Can, Einmahl, Khmaladze and Laeven (2015).

References

1. J. -P. Aubin, H. Frankowska, *Set-valued Analysis*. Birkhäuser Boston, MA, 1990.
2. R. H. Baayen, *Word Frequency Distributions*. Kluwer Academic Publishers, Dordrecht, 2001.
3. V. P. Borovikov, Some limit theorems for statistics which are partial sums of functions of spacings. *Theory Probab. Appl.* **32** (1987), no. 1, 86–97.
4. R. Cairoli, J. B. Walsh, Stochastic integrals in the plane. *Acta Math.* **134** (1975), 111–183.
5. S. U. Can, J. H. J. Einmahl, E. V. Khmaladze, R. J. A. Laeven, Asymptotically distribution-free goodness-of-fit testing for tail copulas. *Ann. Statist.* **43** (2015), no. 2, 878–902.
6. D. M. Chibisov, Certain chi-square type tests for continuous distributions. *Theory Probab. Appl.* **16** (1971), no.1, 1-22.
7. J. Durbin, *Distribution Theory for Tests Based on the Sample Distribution Function.* Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, no. 9. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1973.
8. J. H. J. Einmahl, E. Khmaladze, The two-sample problem in $\mathbb{R}^m$ and measure-valued martingales. *IMS Lecture Notes Monogr. Ser.* **36** (2001), 434-463.
9. J. H. J. Einmahl, E. Khmaladze, Central limit theorems for local empirical processes near boundaries of sets. *Bernoulli* **17** (2011), no. 2, 545–561.

10. J. I. Gikhman, Some remarks on A. N. Kolmogorov's criterion of fit. (Russian) *Doklady Akad. Nauk SSSR* **91** (1953), 715–718.
11. J. I. Gikhman, Biparametric martingales. (Russian) *Uspekhi Mat. Nauk* **37** (1982), 215–237.
12. B. Hajek, E. Wong, Representation and transformation of two-parameter martingales under a change of measure. *Z. Wahrsch. Verw. Gebiete* **54** (1980), no. 3, 313–330.
13. G. I. Ivchenko, V. V. Levin, Processes in discrete schemes (A survey, Part I and II). *Math. Methods Statist.* **2** (1993), no. 1, 72–84; no. 2, 115–129.
14. M. Kac, J. Kiefer, J. Wolfowitz, On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Statist.* **26** (1955), no. 2, 189–211.
15. E. V. Khmaladze, The use of ω^2-tests for testing parametric hypotheses. *Theory Probab. Appl.* **24** (1979), no. 2, 283–301.
16. E. V. Khmaladze, A martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** (1981) no. 2, 240–257.
17. E. V. Khmaladze, Martingale limit theorems for divisible statistics, *Theory Probab. Appl.* **28** (1983), no. 3, 530–549.
18. E. V. Khmaladze, A. M. Parjanadze, Functional limit theorems for linear statistics from sequential ranks. *Probab. Theory Related Fields* **73** (1986), no. 4, 585–595.
19. E. V. Khmaladze, An innovation approach to goodness-of-fit tests in $\mathbf{R}^m$. *Ann. Statist.* **16** (1988a), no. 4, 1503–1516.
20. E.V. Khmaladze, The statistical analysis of a large number of rare events. *Technical Report MS-R8804, CWI, Amsterdam*, 1988b.
21. E. V. Khmaladze, R. Ya. Chitashvili, Statistical analysis of a large number of rare events and related problems. (Russian) Probability theory and mathematical statistics. (Russian) *Trudy Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. SSR* **92** (1989), 196–245.
22. E. V. Khmaladze, Goodness of fit problem and scanning innovation martingales. *Ann. Statist.* **21** (1993), no. 2, 798–829.
23. E. V. Khmaladze, Z. P. Tsigroshvili, On polynomial distributions with a large number of rare events. *Math. Methods Statist.* **2** (1993), no. 3, 240–247.
24. E. Khmaladze, M. Nadareishvili, A. Nikabadze, Asymptotic behaviour of a number of repeated records. *Statist. Probab. Letters* **35** (1997), no.1, 49-58.
25. E. V. Khmaladze, Goodness of fit tests for "himeric" alternatives. *Statist. Neerlandica* **52** (1998), no. 1, 90–111.
26. E. Khmaladze, E. Shinjikashvili, Calculation of noncrossing probabilities for Poisson processes and its corollaries. *Adv. in Appl. Probab.* **33** (2001), no. 3, 702–716.
27. E. Khmaladze, N. Toronjadze, On almost sure coverage property of Voronoi tessellation: The $\mathbf{R}^1$ case. *Adv. Appl. Probab.* **33** (2001), no. 4, 756–764.
28. E. Khmaladze, R. Mnatsakanov, N. Toronjadze, The change-set problem for Vapnik-Červonenkis classes. *Math. Methods Statist.* **15** (2006a), no. 2, 224–231.
29. E. Khmaladze, R. Mnatsakanov, N. Toronjadze, The change set problem and local covering numbers. *Math. Methods Statist.* **15** (2006b), no. 3, 289–308.
30. E. V. Khmaladze, Differentiation of sets in measure. *J. Math. Anal. Appl.* **334** (2007), no. 2, 1055–1072.
31. E. V. Khmaladze, Wolfgang Weil, Local empirical processes near boundaries of convex bodies. *Ann. Inst. Statist. Math.* **60** (2008), no. 4, 813–842.
32. E. V. Khmaladze, Convergence properties in certain occupancy problems including the Karlin-Rouault law. *J. Appl. Probab.* **48** (2011), no. 4, 1095–1113.
33. E. V. Khmaladze, Note on distribution free testing for discrete distributions. *Ann. Statist.* **41** (2013), no. 6, 2979–2993.
34. E. Khmaladze, W. Weil, Fold-up derivatives of set-valued functions and the change-set problem: a survey. *Ann. Inst. Statist. Math.* **70** (2018), no. 1, 1–38.
35. E. V. Khmaladze, Distribution free testing for linear regression. Extension to general parametric regression. *Ann. Inst. Statist. Math.* (2020) (under review).
36. A. N. Kolmogorov, Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.* **4** (1933), 83–91.
37. V. F. Kotel'nikova, E. V. Khmaladze, Calculation of the probability of an empirical process not crossing a curvilinear boundary. (Russian) *Teor. Veroyatnost. i Primenen.* **27** (1982), no. 3, 599–607.
38. B. Mandelbrot, An informational theory of the statistical structure of language. *Communication theory* **84** (1953), 486–502.
39. R. M. Mnacakanov, E. V. Hmaladze, On L_1-convergence of statistical kernel estimators of distribution densities. *Soviet Math. Dokl.* **23** (1981), no. 3, 633-636.
40. R. M. Mnatsakanov, A functional limit theorem for additively separable statistics in the case of very rare events. *Theory Probab. Appl.* **30** (1986), no. 3, 622–626.
41. D. S. Moore, M. C. Spruill, Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* **3** (1975), 599–616.
42. D. Nualart, Two-parameter diffusion processes and martingales. *Stochastic Process. Appl.* **15** (1983), no. 1, 31–57.
43. J. Oosterhoff, W. van Zwet, A Note on Contiguity and Hellinger Distance. In: *Contributions to Statistics*, pp. 157–166, Reidel, Dordrecht-Boston, Mass.-London, 1979.

44. A. M. Pardzhanadze, E. V. Khmaladze, On the asymptotic theory of statistics of sequential ranks. *Theory Probab. Appl.* **31** (1986), no. 4, 669-682.
45. B. L. S. Rao Prakasa, *Asymptotic Theory of Statistical Inference.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1987.
46. R. Schneider, W. Weil, *Stochastic and Integral Geometry.* Springer, New York, 2008.
47. P. K. Sen, Invariance principles for linear rank statistics revisited. *Sankhya Ser. A* **40** (1978), no. 3, 215–236.
48. E. Wong, M. Zakai, Martingales and stochastic integrals for processes with a multi-dimensional parameter. *Z. Wahrsch. Verw. Gebiete* **29** (1974), 109–122.

Department of Mathematics, West Virginia University, P.O. Box 6310, Morgantown, WV 26506, USA

A. Razmadze Mathematical Institute of I. Javakhishvili Tbilisi State University, 6 Tamarashvili Str., Tbilisi 0177, Georgia

E-mail address: Robert.Mnatsakanov@mail.wvu.edu

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 135–146

ON THE TESTING HYPOTHESIS OF EQUALITY OF TWO BERNOULLI REGRESSION FUNCTIONS

PETRE BABILUA AND ELIZBAR NADARAYA

Abstract. We establish the limit distribution of the square-integrable deviation of two nonparametric kernel-type estimations for the Bernoulli regression functions. The criterion of testing the hypothesis of two Bernoulli regression functions is constructed. The question as to its consistency is studied. The power asymptotics of the constructed criterion is also studied for certain types of close alternatives.

Assume that random variables $Y^{(i)}$, $i=1,2$, take two values: 1 and 0 with probabilities p_i ("success") and $1-p_i$ ("failure"), $i=1,2$, respectively. Assume that the probability of "success" p_i is a function of an independent variable $x\in[0,1]$, i.e., $p_i=p_i(x)=\mathbb{P}\{Y^{(i)}=1 \mid x\}$ (see [2,3,8]). Let t_j, $j=1,\dots,n$, be points of a partition of the segment $[0,1]$:

$$t_j=\frac{2j-1}{2n}\,,\ \ j=1,\dots,n.$$

Let $Y_i^{(1)}$ and $Y_i^{(2)}$, $i=1,\dots,n$, be mutually independent Bernoulli random variables with

$$\mathbb{P}\{Y_i^{(k)}=1 \mid t_i\}=p_k(t_i)\ \text{ and }\ \mathbb{P}\{Y_i^{(k)}=0 \mid t_i\}=1-p_k(t_i),\quad i=1,\dots,n,\ \ k=1,2.$$

Using the samples $Y_1^{(1)},\dots,Y_n^{(1)}$ and $Y_1^{(2)},\dots,Y_n^{(2)}$, it is required to test the hypothesis

$$H_0:\ p_1(x)=p_2(x)=p(x),\ \ x\in[0,1],$$

against a sequence of "close" alternatives:

$$H_{1n}: p_1(x)=p(x),\ \ p_2(x)=p(x)+\alpha_n u(x)+o(\alpha_n),$$

where α_n tends to 0 in a suitable way, $u(x)\neq 0$, $x\in[0,1]$, and the third term is $o(\alpha_n)$ uniformly with respect to $x\in[0,1]$.

The problem of comparison of two Bernoulli regression functions may appear in some applications, e.g., in the quantum bioanalyses carried out in pharmacology. In this case, x is a dose of medicine and $p(x)$ is the probability of efficiency of the dose x [3,6].

To test the hypothesis H_0 we use the statistic:

$$\begin{aligned} T_n&=\frac{1}{2}\,nb_n\int\limits_{\Omega_n(\tau)}\big[\widehat{p}_{1n}(x)-\widehat{p}_{2n}(x)\big]^2p_n^2(x)\,dx\\ &=\frac{1}{2}\,nb_n\int\limits_{\Omega_n(\tau)}\big[p_{1n}(x)-p_{2n}(x)\big]^2\,dx,\\ &\qquad \Omega_n(\tau)=[\tau b_n,1-\tau b_n],\ \ \tau>0,\end{aligned}$$

2020 *Mathematics Subject Classification.* 62G10, 62G20.

Key words and phrases. Bernoulli regression function; Power; Consistency; Limiting distribution.

where

$$\widehat{p}_{in}(x) = p_{in}(x)p_n^{-1}(x),$$
$$p_{in}(x) = (nb_n)^{-1}\sum_{j=1}^{n} K\Big(\frac{x-t_j}{b_n}\Big)Y_j^{(i)}, \quad i=1,2,$$
$$p_n(x) = (nb_n)^{-1}\sum_{j=1}^{n} K\Big(\frac{x-t_j}{b_n}\Big),$$

$K(x)$ is a distribution density, $b_n \to 0$ is a sequence of positive numbers, and $\widehat{p}_{in}(x)$ is a kernel estimate for the regression function [6, 9].

We assume that the kernel $K(x) \geq 0$ is chosen so that it is a function with bounded variation satisfying the following conditions: $K(x) = K(-x)$, $K(x) = 0$ for $|x| \geq \tau > 0$ and

$$\int K(x)\,dx = 1,$$

By $H(\tau)$, we denote the class of such functions.

We also introduce the following notation:

$$T_n^{(1)} = \frac{1}{2}\,nb_n \int\limits_{\Omega_n(\tau)} \big[\widetilde{p}_{1n}(x) - \widetilde{p}_{2n}(x)\big]^2\,dx,$$
$$\widetilde{p}_{in}(x) = p_{in}(x) - \boldsymbol{E}p_{in}(x), \quad i=1,2.$$

It is clear that

$$T_n^{(1)} = H_n + \frac{1}{2nb_n}\sum_{i=1}^{n}\varepsilon_i^2 Q_{ii}, \quad H_n = \frac{1}{nb_n}\sum_{1\leq i<j\leq n}\varepsilon_i\varepsilon_j Q_{ij},$$
$$\varepsilon_i = \varepsilon_{1i} - \varepsilon_{2i}, \quad \varepsilon_{ki} = Y_i^{(k)} - p_k(t_i), \quad k=1,2, \quad i=1,\dots,n,$$
$$Q_{ij} = \psi_n(t_i,t_j), \quad \psi_n(u,v) = \int\limits_{\Omega_n(\tau)} K\Big(\frac{x-u}{b_n}\Big)K\Big(\frac{x-v}{b_n}\Big)\,dx.$$

It is easy to see that

$$\sigma_n^{-1}(T_n^{(1)} - \Delta_n) = \sum_{k=1}^{n}\xi_k^{(n)} + \frac{1}{2nb_n\sigma_n}\sum_{i=1}^{n}(\varepsilon_i^2 - \boldsymbol{E}\varepsilon_i^2)Q_{ii},$$
$$\Delta_n = \boldsymbol{E}T_n^{(1)}, \quad \sigma_n^2 = \boldsymbol{Var}\,H_n = (nb_n)^{-2}\sum_{k=2}^{n} d_k\sum_{i=1}^{k-1} d_i Q_{ik}^2,$$
$$d_i = d(t_i) = \boldsymbol{Var}\,\varepsilon_i, \quad i=1,\dots,n,$$
$$\xi_k^{(n)} = \sum_{i=1}^{k-1}\eta_{ik}^{(n)}, \quad k=2,\dots,n, \quad \xi_1^{(n)} = 0, \quad \xi_k^{(n)} = 0, \quad k>n,$$
$$\eta_{ij}^{(n)} = \frac{\varepsilon_i\varepsilon_j Q_{ij}}{nb_n\sigma_n}, \quad \mathcal{F}_k^{(n)} = \sigma(\varepsilon_1,\dots,\varepsilon_k),$$

i.e., $\mathcal{F}_k^{(n)}$ is a σ-algebra generated by random variables $\varepsilon_1,\dots,\varepsilon_k$ and $\mathcal{F}_0^{(n)} = (\varnothing,\Omega)$ in what follows, for simplicity, we use the notation $\xi_k^{(n)}$, $\eta_{ij}^{(n)}$ and $\mathcal{F}_k^{(n)}$ instead of ξ_k, η_{ij} $\mathcal{F}_k$.

Lemma 1. *The stochastic sequence $(\xi_k,\mathcal{F}_k)_{k\geq 1}$ is a martingale difference*

Lemma 2 ([7])**.** *Let $K(x) \in H(\tau)$ and $p(x)$, $0 \leq x \leq 1$, be a function of bounded variation. If $nb_n \to \infty$, then*

$$\frac{1}{nb_n}\sum_{i=1}^{n} K^{\nu_1}\Big(\frac{x-t_i}{b_n}\Big)K^{\nu_2}\Big(\frac{y-t_i}{b_n}\Big)p^{\nu_3}(t_i)$$

$$= \frac{1}{b_n} \int\limits_0^1 K^{\nu_1}\Big(\frac{x-u}{b_n}\Big) K^{\nu_2}\Big(\frac{y-u}{b_n}\Big) p^{\nu_3}(u)\,du + O\Big(\frac{1}{nb_n}\Big),$$

uniformly with respect $x,\, y \in [0,1]$, *where* $\nu_i \in N \cup \{0\}$, $i = 1,2,3$.

Lemma 3. *Let* $K(x) \in H(\tau)$, $p(x) \in C^1[0,1]$ *and let* $u(x)$ *be a continuous function on* $[0,1]$. *If* $nb_n^2 \to \infty$ *and* $\alpha_n = n^{-1/2} b_n^{-1/4}$, *then, for the hypothesis* H_{1n}

$$b_n^{-1}\sigma_n^2 \longrightarrow \sigma^2(p) = 2\int\limits_0^1 p^2(x)(1-p(x))^2\,dx \int\limits_{|x|\le 2\tau} K_2^2(x)\,dx \tag{1}$$

and

$$b_n^{-1/2}(\Delta_n - \Delta(p)) = O(b_n^{1/2}) + O(\alpha_n b_n^{-1/2}) + O\Big(\frac{1}{nb_n^{3/2}}\Big), \tag{2}$$

where

$$\Delta_n = \boldsymbol{E} T_n^{(1)}, \quad \Delta(p) = \int\limits_0^1 p(x)(1-p(x))\,dx \int\limits_{|x|\le\tau} K^2(u)\,du, \quad K_2 = K * K,$$

and $*$ *denotes the operation of convolution.*

The following statement is true:

Theorem 1. *Let* $K(x) \in H(\tau)$ *and* $p(x), u(x) \in C^1[0,1]$. *If* $nb_n^2 \to \infty$ *and* $\alpha_n = n^{-1/2} b_n^{-1/4}$, *then, for the hypothesis* H_{1n},

$$b_n^{-1/2}(T_n - \Delta(p))\sigma^{-1}(p) \xrightarrow{d} N(a,1),$$

where $\Delta(p)$ *and* $\sigma^2(p)$ *are defined in Lemma* 3, $\xrightarrow{d}$ *denotes the convergence in distribution,* $N(a,1)$ *is a random variable having normal distribution with parameters* $(a,1)$, *and*

$$a = \frac{1}{2\sigma(p)} \int\limits_0^1 u^2(x)\,dx.$$

Proof. We have

$$T_n = T_n^{(1)} + L_n^{(1)} + L_n^{(2)},$$

where

$$L_n^{(1)} = nb_n \int\limits_{\Omega_n(\tau)} \big[\widetilde{p}_{1n}(x) - \widetilde{p}_{2n}(x)\big]\big[\boldsymbol{E}p_{1n}(x) - \boldsymbol{E}p_{2n}(x)\big]\,dx,$$

$$L_n^{(2)} = \frac{1}{2}\, nb_n \int\limits_{\Omega_n(\tau)} \big[\boldsymbol{E}p_{1n}(x) - \boldsymbol{E}p_{2n}(x)\big]^2\,dx.$$

By virtue of Lemma 2, we conclude that

$$b_n^{-1/2} L_n^{(2)} = \frac{1}{2}\, nb_n^{1/2}\alpha_n^2 \int\limits_{\Omega_n(\tau)} \bigg\{\frac{1}{b_n}\int\limits_0^1 K\Big(\frac{x-t}{b_n}\Big)u(t)\,dt + O\Big(\frac{1}{nb_n}\Big)\bigg\}^2\,dx. \tag{3}$$

Since $\big[\frac{x-1}{b_n}, \frac{x}{b_n}\big] \supset [-\tau,\tau]$ for all $x \in \Omega_n(\tau)$, it follows from (3) that

$$b_n^{-1/2} L_n^{(2)} = \frac{1}{2}\, nb_n^{1/2}\alpha_n^2 \int\limits_{\Omega_n(\tau)} \bigg[\int\limits_{-\tau}^{\tau} K(t)u(x-b_n t)\,dt + O\Big(\frac{1}{nb_n}\Big)\bigg]^2\,dx. \tag{4}$$

Further, since $u(x) \in C^1[0,1]$, in view of (4), we get

$$b_n^{-1/2} L_n^{(2)} \longrightarrow \frac{1}{2} \int\limits_0^1 u^2(t)\,dt. \tag{5}$$

We now show that

$$b_n^{-1/2} L_n^{(1)} \xrightarrow{\boldsymbol{P}} 0.$$

Thus, we have

$$b_n^{-1/2} L_n^{(1)} = \frac{1}{2}\, n b_n^{1/2} \int\limits_{\Omega_n(\tau)} \widetilde{p}_{1n}(x)\big(\boldsymbol{E}p_{1n}(x) - \boldsymbol{E}p_{2n}(x)\big)\,dx$$
$$- \frac{n b_n^{1/2}}{2} \int\limits_{\Omega_n(\tau)} \widetilde{p}_{2n}(x)\big(\boldsymbol{E}p_{1n}(x) - \boldsymbol{E}p_{2n}(x)\big)\,dx = I_n^{(1)} + I_n^{(2)}. \tag{6}$$

It is clear that

$$\begin{aligned}
\boldsymbol{E}|I_n^{(1)}| &\le \big(\boldsymbol{E}(I_n^{(1)})^2\big)^{1/2} \\
&= \frac{1}{2}\, n b_n^{1/2} \bigg[\boldsymbol{E}\bigg(\int\limits_{\Omega_n(\tau)} \widetilde{p}_{1n}(x)\big(\boldsymbol{E}p_{1n}(x) - \boldsymbol{E}p_{2n}(x)\big)\,dx\bigg)^2\bigg]^{1/2} \\
&= \frac{1}{2}\, n b_n^{1/2} \bigg[\int\limits_{\overline{\Omega}_n(\tau)} \operatorname{cov}\big(p_{1n}(x_1), p_{1n}(x_2)\big)\big(\boldsymbol{E}p_{1n}(x_1) - \boldsymbol{E}p_{2n}(x_1)\big) \\
&\qquad \times \big(\boldsymbol{E}p_{1n}(x_2) - \boldsymbol{E}p_{2n}(x_2)\big)\,dx_1\,dx_2\bigg]^{1/2}, \quad \overline{\Omega}_n(\tau) = \Omega_n(\tau) \times \Omega_n(\tau).
\end{aligned}$$

It is easy to see that

$$\operatorname{cov}\big(p_{1n}(x_1), p_{1n}(x_2)\big)$$
$$= \frac{1}{(nb_n)^2} \sum_{i=1}^n K\Big(\frac{x_1 - t_i}{b_n}\Big) K\Big(\frac{x_2 - t_i}{b_n}\Big) p_1(t_i)(1 - p_1(t_i)).$$

By virtue of Lemma 2, we find

$$\operatorname{cov}\big(p_{1n}(x_1), p_{1n}(x_2)\big)$$
$$= n^{-1} b_n^{-2} \int\limits_0^1 K\Big(\frac{x_1 - u}{b_n}\Big) K\Big(\frac{x_2 - u}{b_n}\Big) p_1(u)(1 - p_1(u))\,du + O\Big(\frac{1}{(nb_n)^2}\Big).$$

Hence,

$$\begin{aligned}
\boldsymbol{E}|I_n^{(1)}| \le \frac{1}{2}\, n b_n^{1/2} \bigg\{ \int\limits_{\overline{\Omega}_n(\tau)} \bigg[\frac{1}{n b_n^2} \\
\times \int\limits_0^1 K\Big(\frac{x_1 - u}{b_n}\Big) K\Big(\frac{x_2 - u}{b_n}\Big) p_1(u)(1 - p_1(u))\,du + \frac{1}{(nb_n)^2}\bigg] \\
\times \big(\boldsymbol{E}p_{1n}(x_1) - \boldsymbol{E}p_{2n}(x_1)\big)\big(\boldsymbol{E}p_{1n}(x_2) - \boldsymbol{E}p_{2n}(x_2)\big)\,dx_1\,dx_2 \bigg\}^{1/2} \\
\le c_3 \sqrt{n}\, b_n^{1/2} \alpha_n = c_3 \frac{1}{\sqrt{n}\,\alpha_n} \longrightarrow 0,
\end{aligned}$$

because

$$\sqrt{n}\,\alpha_n = \frac{1}{b_n^{1/4}} \longrightarrow \infty.$$

Therefore, $I_n^{(1)} \xrightarrow{\boldsymbol{P}} 0$. Similarly, we prove that $I_n^{(2)} \xrightarrow{\boldsymbol{P}} 0$.

By using (6), we get

$$b_n^{-1/2} L_n^{(1)} \xrightarrow{\boldsymbol{P}} 0. \tag{7}$$

To prove the theorem, it remains to show that

$$\frac{T_n^{(1)} - \Delta_n}{\sigma_n} \xrightarrow{d} N(0,1).$$

We have

$$\frac{T_n^{(1)} - \Delta_n}{\sigma_n} = K_n^{(1)} + K_n^{(2)},$$

where

$$K_n^{(1)} = \sum_{k=1}^{n} \xi_k, \quad K_n^{(2)} = \frac{\sum\limits_{i=1}^{n} (\varepsilon_i^2 - \boldsymbol{E}\varepsilon_i^2) Q_{ii}}{2nb_n\sigma_n}\,.$$

We now show that $K_n^{(2)} \xrightarrow{\boldsymbol{P}} 0$. Indeed,

$$\boldsymbol{Var}(K_n^{(2)}) = (2nb_n\sigma_n)^{-2} \sum_{i=1}^{n} \boldsymbol{Var}\varepsilon_i^2 Q_{ii}^2$$

$$= (2nb_n\sigma_n)^{-2} \sum_{i=1}^{n} \left(\sum_{k=1}^{2} p_k(t_i)(1-p_k(t_i))\big[1 - 3p_k(t_i)(1-p_k(t_i))\big] \right) Q_{ii}^2.$$

Since $Q_{ii} \le c_4 b_n$ $\;b_n^{-1}\sigma_n^2 \longrightarrow \sigma^2(p)$ as $n \to \infty$, this yields

$$\boldsymbol{Var}(K_n^{(2)}) \le c_5\, \frac{1}{nb_n}\,.$$

Thus, $K_n^{(2)} \xrightarrow{\boldsymbol{P}} 0$.

We now prove that $K_n^{(1)} \xrightarrow{d} N(0,1)$. To this end, we show that it is possible to apply Corollaries 2 and 6 of Theorem 2 in [4]. It is necessary to check the validity of conditions imposed in these statements and guaranteeing the asymptotic normality of a square-integrable martingale difference and to take into account the fact that, according to Lemma 1, the sequence $\{\xi_k, \mathcal{F}_k\}_{k\ge1}$ is, in fact, a square-integrable martingale difference.

It is easy to see that $\sum\limits_{k=1}^{n} \boldsymbol{E}\xi_k^2 = 1$. The asymptotic normality of $K_n^{(1)}$ is realized whenever

$$\sum_{k=1}^{n} \boldsymbol{E}\Big[\xi_k^2 I\big(|\xi_k| \ge \varepsilon\big) \mid \mathcal{F}_{k-1}\Big] \xrightarrow{\boldsymbol{P}} 0 \tag{8}$$

and

$$\sum_{k=1}^{n} \xi_k^2 \xrightarrow{\boldsymbol{P}} 1 \tag{9}$$

as $n \to \infty$. In [4], it is proved that, under the conditions

$$\sup_{1\le k\le n} |\xi_k| \xrightarrow{\boldsymbol{P}} 0$$

and (9), condition (8) is also satisfied.

Note that, for $\varepsilon > 0$, we have

$$P\bigg\{ \sup_{1\le k\le n} |\xi_k| \ge \varepsilon \bigg\} \le \varepsilon^{-4} \sum_{k=1}^{n} \boldsymbol{E}\xi_k^4.$$

Hence, by virtue of relation (11) presented in what follows, in order to prove

$$K_n^{(1)} \xrightarrow{d} N(0,1)$$

it remains to check the validity of condition (9). To this end, it suffices to show that

$$\boldsymbol{E}\Big(\sum_{k=1}^n \xi_k^2 - 1\Big)^2 \longrightarrow 0 \ \text{ as } \ n \to \infty$$

i.e., since $\sum\limits_{k=1}^n \boldsymbol{E}\xi_k^2 = 1$, we get

$$\boldsymbol{E}\Big(\sum_{k=1}^n \xi_k^2\Big)^2 = \sum_{k=1}^n \boldsymbol{E}\xi_k^4 + 2\sum_{1\le k_1<k_2\le n} \boldsymbol{E}\xi_{k_1}^2\xi_{k_2}^2 \longrightarrow 1. \tag{10}$$

We now prove (10). Taking into account the definitions of η_{ik} and ξ_k, we obtain

$$\sum_{k=1}^n \boldsymbol{E}\xi_k^4 = I_n^{(1)} + I_n^{(2)},$$

where

$$I_n^{(1)} = \frac{1}{(nb_n)^4\sigma_n^4}\sum_{k=2}^n \boldsymbol{E}\varepsilon_k^4 \sum_{j=1}^{k-1} \boldsymbol{E}\varepsilon_j^4 Q_{jk}^4,$$
$$I_n^{(2)} = \frac{3}{(nb_n)^4\sigma_n^4}\sum_{k=2}^n \sum_{i\ne j} \boldsymbol{E}\varepsilon_j^2 \boldsymbol{E}\varepsilon_i^2 Q_{jk}^2 Q_{ik}^2.$$

Since

$$Q_{ij} \le c_6 b_n, \quad \boldsymbol{E}\varepsilon_j^4 \le 8\sum_{k=1}^2 p_k(t_j)\big(1-p_k(t_j)\big)\Big[1 - 3p_k(t_j)\big(1-p_k(t_j)\big)\Big] \le 4,$$
$$\boldsymbol{E}\varepsilon_j^2 \le \frac{1}{2}, \quad |\boldsymbol{E}\varepsilon_j^3| \le \sum_{k=1}^2 p_k(t_j)\big(1-p_k(t_j)\big)\Big[\big(1-p_k(t_j)\big)^2 + p_k^2(t_j)\Big] \le 1$$

and $b_n^{-1}\sigma_n^2 \longrightarrow \sigma^2(p)$, we find

$$I_n^{(1)} = O\Big(\frac{1}{(nb_n)^2}\Big), \quad I_n^{(2)} = O\Big(\frac{1}{nb_n^2}\Big).$$

Hence,

$$\sum_{k=1}^n \boldsymbol{E}\xi_k^4 \longrightarrow 0 \ \text{ for } \ n \to \infty. \tag{11}$$

Further, it follows from the definition of ξ_i that

$$\xi_{k_1}^2\xi_{k_2}^2 = B_{k_1k_2}^{(1)} + B_{k_1k_2}^{(2)} + B_{k_1k_2}^{(3)} + B_{k_1k_2}^{(4)},$$

where

$$B_{k_1k_2}^{(1)} = \sigma_2(k_1)\sigma_2(k_2), \quad B_{k_1k_2}^{(2)} = \sigma_2(k_1)\sigma_1(k_2),$$
$$B_{k_1k_2}^{(3)} = \sigma_1(k_1)\sigma_2(k_2), \quad B_{k_1k_2}^{(4)} = \sigma_1(k_1)\sigma_1(k_2),$$
$$\sigma_1(k) = \sum_{1\le i\ne j\le k-1} \eta_{ik}\eta_{jk}, \quad \sigma_2(k) = \sum_{i=1}^{k-1} \eta_{ik}^2.$$

Therefore,

$$2\sum_{1\le k_1<k_2\le n} \boldsymbol{E}\xi_{k_1}^2\xi_{k_2}^2 = \sum_{i=1}^4 A_n^{(i)},$$

where

$$A_n^{(i)} = 2 \sum_{1\le k_1<k_2\le n} \boldsymbol{E} B_{k_1k_2}^{(i)}, \quad i=1,2,3,4.$$

We now consider $A_n^{(3)}$. By using the definition of η_{ij}, we can easily show that $\boldsymbol{E} B_{k_1k_2}^{(3)} = 0$ and, hence,

$$A_n^{(3)} = 0. \tag{12}$$

We now estimate $A_n^{(2)}$. We have

$$|\boldsymbol{E} B_{k_1k_2}^{(2)}| = \frac{1}{(nb_n\sigma_n)^4} \Big| \sum_{i=1}^{k_1-1} \boldsymbol{E}\varepsilon_i^3 \boldsymbol{E}\varepsilon_{k_1}^3 \boldsymbol{E}\varepsilon_{k_2}^2 Q_{ik_1}^2 Q_{ik_2} Q_{k_1k_2} \Big|.$$

Since $\boldsymbol{E}|\varepsilon_i^3| \le 1$ and $Q_{ij} \le c_6 b_n$, we get

$$|\boldsymbol{E} B_{k_1k_2}^{(2)}| \le c_6\, \frac{k_1-1}{(n\sigma_n)^4}\,.$$

Further, since

$$\sum_{1\le k_1<k_2\le n} (k_1-1) = O(n^3) \ \text{ and } \ b_n^{-1}\sigma_n^2 \longrightarrow \sigma^2(p) > 0,$$

we obtain

$$|A_n^{(2)}| \le c_7\, \frac{n^3}{n^4\sigma_n^4} = c_7\, \frac{1}{nb_n^2(b_n^{-1}\sigma_n^2)^2} = O\Big(\frac{1}{nb_n^2}\Big). \tag{13}$$

We now establish that $A_n^{(1)} \to 1$ as $n\to\infty$. It is clear that

$$A_n^{(1)} = 2 \sum_{1\le k_1<k_2\le n} \boldsymbol{E} B_{k_1k_2}^{(1)} = S_n^{(1)} + S_n^{(2)},$$

where

$$S_n^{(1)} = 2 \sum_{1\le k_1<k_2\le n} \Big(\sum_{i=1}^{k_1-1} \boldsymbol{E}\eta_{ik_1}^2\Big)\Big(\sum_{j=1}^{k_2-1} \boldsymbol{E}\eta_{jk_2}^2\Big),$$

$$S_n^{(2)} = 2\bigg(\sum_{k_1<k_2} \boldsymbol{E} B_{k_1k_2}^{(1)} - \sum_{k_1<k_2} \Big(\sum_{i=1}^{k_1-1} \boldsymbol{E}\eta_{ik_1}^2\Big)\Big(\sum_{j=1}^{k_2-1} \boldsymbol{E}\eta_{jk_2}^2\Big)\bigg).$$

It follows from the definition of σ_n^2 that

$$S_n^{(1)} = 1 - \sum_{k=2}^{n} \Big(\sum_{i=1}^{k-1} \boldsymbol{E}\eta_{ik}^2\Big)^2.$$

Furthermore,

$$\sum_{k=2}^{n} \Big(\sum_{i=1}^{k-1} \boldsymbol{E}\eta_{ik}^2\Big)^2 \le c_8\, \frac{b_n^4 n^3}{(nb_n)^4\sigma_n^4} = O\Big(\frac{1}{nb_n^2}\Big).$$

This yields

$$S_n^{(1)} = 1 + O\Big(\frac{1}{nb^2}\Big). \tag{14}$$

Further, we show that $S_n^{(2)} \to 0$. The quantity $S_n^{(2)}$ can be rewritten in the form

$$S_n^{(2)} = 2 \sum_{k_1<k_2} \bigg[\sum_{i=1}^{k_1-1} \operatorname{cov}(\eta_{ik_1}^2, \eta_{ik_2}^2) + \sum_{i=1}^{k_1-1} \operatorname{cov}(\eta_{ik_1}^2, \eta_{k_1k_2}^2)\bigg].$$

It is easy to see that

$$\operatorname{cov}(\eta_{ik_1}^2, \eta_{ik_2}^2) = O\Big(\frac{1}{n^4\sigma_n^4}\Big).$$

However, since

$$\sum_{1\le k_1<k_2\le n} (k_1-1) = O(n^3),$$

we conclude that

$$S_n^{(2)} = O\Big(\frac{1}{n\sigma_n^4}\Big) = O\Big(\frac{1}{nb_n^2}\Big). \tag{15}$$

Hence, according to (14) and (15), we find

$$A_n^{(1)} = 1 + O\Big(\frac{1}{nb_n^2}\Big). \tag{16}$$

Finally, we show that $A_n^{(4)} \to 0$ as $n \to \infty$. By using the definition of η_{ij} and the inequalities $Q_{ij} \geq 0$ and

$$\boldsymbol{E}\varepsilon_i^2 = d(t_i) \leq \frac{1}{2}\,,$$

we obtain

$$\boldsymbol{E}B_{k_1k_2}^{(4)} = 4 \sum_{1\leq t<s\leq k_1-1} \boldsymbol{E}\eta_{sk_1}\eta_{tk_1}\eta_{sk_2}\eta_{tk_2}$$

$$\leq \frac{c_8}{n^4b_n^4\sigma_n^4} \sum_{1\leq t<s\leq k_1-1} Q_{sk_1}Q_{tk_1}Q_{sk_2}Q_{tk_2}.$$

Thus,

$$A_n^{(4)} \leq \frac{c_9}{n^2b_n^4\sigma_n^4} \sum_{k_1<k_2} A_{k_1k_2}\,,$$

where

$$A_{k_1k_2} = \frac{1}{n^2} \sum_{1\leq t<s\leq k_1-1} Q_{sk_1}Q_{tk_1}Q_{sk_2}Q_{tk_2}.$$

At the same time,

$$\sum_{k_1<k_2} A_{k_1k_2} \leq \sum_{k_1,k_2=1}^{n} \Big(\frac{1}{n}\sum_{t=1}^{n} Q_{tk_1}Q_{tk_2}\Big)^2.$$

Therefore,

$$A_n^{(4)} \leq c_{10}\,\frac{1}{n^2b_n^4\sigma_n^4} \sum_{k_1,k_2=1}^{n} \bigg[\int\limits_{\Omega(\tau)}\int\limits_{\Omega_n(\tau)} K\Big(\frac{x-x_{k_1}}{b_n}\Big)K\Big(\frac{y-x_{k_2}}{b_n}\Big)$$

$$\times\frac{1}{n}\sum_{i=1}^{n} K\Big(\frac{x-x_i}{b_n}\Big)K\Big(\frac{y-x_i}{b_n}\Big)\,dx\,dy\Big)\bigg]^2. \tag{17}$$

Further, in view of Lemma 2, it follows from (17) that

$$A_n^{(4)} \leq \frac{c_{11}}{b_n^4\sigma_n^4} \sum_{k_1,k_2=1}^{n} \bigg\{\frac{1}{n}\int\limits_0^1 \int\limits_{\Omega_n(\tau)}\int\limits_{\Omega_n(\tau)} K\Big(\frac{x-x_{k_1}}{b_n}\Big)K\Big(\frac{y-x_{k_2}}{b_n}\Big)$$

$$\times K\Big(\frac{x-u}{b_n}\Big)K\Big(\frac{y-u}{b_n}\Big)\,du\,dx\,dy\bigg\}^2 + O\Big(\frac{1}{nb_n^2}\Big). \tag{18}$$

In relation (18), we now apply Lemma 2 once again. This yields

$$A_n^{(4)} \leq \frac{c_{12}}{b_n^4\sigma_n^4} \int\limits_0^1\int\limits_0^1\int\limits_0^1\int\limits_0^1 \psi_n(u_1,v_2)\psi_n(u_1,v_1)$$

$$\times\psi_n(u_2,v_1)\psi_n(u_2,v_2)\,du_1\,du_2\,dv_1\,dv_2, \tag{19}$$

where

$$\psi_n(x,y) = \int\limits_{\Omega_n(\tau)} K\Big(\frac{t-x}{b_n}\Big)K\Big(\frac{t-y}{b_n}\Big)\,dt.$$

We now estimate the integral in (19). Since $\left[\frac{x-1}{b_n}, \frac{x}{b_n}\right] \supseteq [-\tau,\tau]$ for all $x \in \Omega_n(\tau)$, we get

$$\begin{aligned}
&\int_0^1 \psi_n(u_1,v_2)\psi_n(u_1,v_1)\,du_1 \\
&= b_n \int_{\overline{\Omega}_n(\tau)} K\Big(\frac{t-v_2}{b_n}\Big)K\Big(\frac{z-v_1}{b_n}\Big)K_2\Big(\frac{z-t}{b_n}\Big)\,dt\,dz \\
&\le c_{13}b_n^3, \quad K_2 = K*K,\ \ \overline{\Omega}_n(\tau) = \Omega_n(\tau)\times\Omega_n(\tau).
\end{aligned}$$

Hence,

$$A_n^{(4)} \le c_{14}\,\frac{1}{b_n\sigma_n^4}\int_0^1\int_0^1\int_0^1 \psi_n(u_2,v_1)\psi_n(u_2,v_2)\,du_2\,dv_1\,dv_2 + O\Big(\frac{1}{nb_n^2}\Big). \tag{20}$$

Further, in a similar way, we derive the following result from (20):

$$\begin{aligned}
A_n^{(4)} &\le c_{15}\,\frac{b_n^4}{b_n\sigma_n^4} + O\Big(\frac{1}{nb_n^2}\Big) = O\Big(\frac{b_n^4}{b_n^3(b_n^{-1}\sigma_n^2)^2}\Big) + O\Big(\frac{1}{nb_n^2}\Big) \\
&= O(b_n) + O\Big(\frac{1}{nb_n^2}\Big).
\end{aligned} \tag{21}$$

Combining relations(12), (13), (16) and (21), we conclude that

$$2\sum_{1\le k_1<k_2\le n} \boldsymbol{E}\xi_{k_1}^2\xi_{k_2}^2 \longrightarrow 1.$$

In view of relation (11), this yields that

$$\boldsymbol{E}\Big(\sum_{k=1}^n \xi_k^2 - 1\Big)^2 \longrightarrow 0 \ \text{ for } \ n\to\infty.$$

Hence,

$$\frac{T_n^{(1)}-\Delta_n}{\sigma_n} \xrightarrow{d} N(0,1). \tag{22}$$

Further, by using the representation $T_n = T_n^{(1)} + L_n^{(1)} + L_n^{(2)}$, Lemma 3 and relations (5), (7), and (22) we get

$$b_n^{-1/2}\Big(\frac{T_n-\Delta(p)}{\sigma(p)}\Big) \xrightarrow{d} N\bigg(\frac{1}{2\sigma(p)}\int_0^1 u^2(x)\,dx, 1\bigg).$$

Theorem 1 is proved. □

Corollary 1. *Let $K(u)\in H(\tau)$ and $p(x)\in C^1[0,1]$. If $nb_n^2\to\infty$, then the following relation is true for the hypothesis H_0:*

$$b_n^{-1/2}(T_n-\Delta(p))\sigma^{-1}(p) \xrightarrow{d} N(0,1). \tag{23}$$

As an important application of Corollary 1, we construct a criterion for the testing of a simple hypothesis H_0 of equality of two Bernoulli regression functions $p_1(x) = p_2(x) = p(x)$, where the function $p(x)$ is completely defined. The critical domain is determined by the inequality

$$T_n \ge d_n(\alpha) = \Delta(p) + b_n^{1/2}\sigma(p)\lambda_\alpha,$$

where $\Phi(\lambda_\alpha) = 1-\alpha$ and $\Phi(\lambda)$ is the standard normal distribution.

Corollary 2. *Let $K(u)\in H(\tau)$ and $p(x), u(x)\in C^1[0,1]$. If $nb_n^2\to\infty$ and $\alpha_n = n^{-1/2}b_n^{-1/4}$, then the local behavior of the power $\boldsymbol{P}_{H_{1n}}(T_n\ge d_n(\alpha))$ has the form*

$$\boldsymbol{P}_{H_{1n}}(T_n\ge d_n(\alpha)) \longrightarrow 1-\Phi\Big(\lambda_\alpha - \frac{A(u)}{\sigma(p)}\Big),$$

where

$$A(u) = \frac{1}{2} \int_0^1 u^2(x)\,dx > 0.$$

We now assume that $p(x)$ is not defined by the hypothesis (i.e., we testing a ***composite*** hypothesis).

In this case, it is impossible to apply inequality (1) directly. First, it is necessary to replace the unknown parameters $\Delta(p)$ and $\sigma^2(p)$ appearing in (23) by certain estimates $\widetilde{\Delta}_n$ and $\widetilde{\sigma}_n^2$, respectively. As the estimates $\Delta(p)$ and $\sigma^2(p)$, we take the following statistics:

$$\widetilde{\Delta}_n = \int\limits_{\Omega_n(\tau)} \lambda_n(x)\,dx \int\limits_{|x|\le\tau} K^2(x)\,dx,$$

$$\widetilde{\sigma}_n^2 = 2 \int\limits_{\Omega_n(\tau)} \lambda_n^2(x)\,dx \int\limits_{|x|\le 2\tau} K_2^2(x)\,dx,$$

$$\lambda_n(x) = \frac{1}{2}\Big[p_{1n}(x)\big(p_n(x) - p_{1n}(x)\big) + p_{2n}(x)\big(p_n(x) - p_{2n}(x)\big)\Big].$$

We now show that

$$b_n^{-1/2}(\widetilde{\Delta}_n - \Delta(p)) \xrightarrow{\boldsymbol{P}} 0, \quad \widetilde{\sigma}_n^2 \xrightarrow{\boldsymbol{P}} \sigma^2(p). \tag{24}$$

Indeed, since

$$p_n(x) = 1 + O\Big(\frac{1}{nb_n}\Big)$$

uniformly with respect to $x \in \Omega_n(\tau)$ and $|p_{in}(x)| \le c_{16}$, $x \in [0,1]$, $i = 1,2$, we find

$$\begin{aligned}
&b_n^{-1/2}\boldsymbol{E}|\widetilde{\Delta}_n - \Delta(p)| \\
&\le c_{17} b_n^{-1/2}\Bigg[\int\limits_{\Omega_n(\tau)} \Big(\boldsymbol{E}\big(p_{1n}(x) - \boldsymbol{E}p_{1n}(x)\big)^2\Big)^{1/2} dx \\
&\quad + \int\limits_{\Omega_n(\tau)} \Big(\boldsymbol{E}\big(p_{2n}(x) - \boldsymbol{E}p_{2n}(x)\big)^2\Big)^{1/2} dx\Bigg] \\
&+ b_n^{-1/2} \int\limits_{\Omega_n(\tau)} \big|\boldsymbol{E}p_{1n}(x) - p(x)\big|\,dx + b_n^{-1/2} \int\limits_{\Omega_n(\tau)} \big|\boldsymbol{E}p_{2n}(x) - p(x)\big|\,dx.
\end{aligned}$$

Further, by using Lemma 2 and taking into account the facts that $p(x) \in C^1[0,1]$ and $\big[\frac{x-1}{b_n}, \frac{x}{b_n}\big] \supset [-\tau,\tau]$ for all $x \in \Omega_n(\tau)$, we immediately conclude that

$$\begin{aligned}
&b_n^{-1/2}\boldsymbol{E}|\widetilde{\Delta}_n - \Delta(p)| \\
&\le c_{18} b_n^{-1/2}\Bigg\{\int\limits_{\Omega_n(\tau)} \Bigg[\frac{1}{nb_n}\,\frac{1}{b_n}\int_0^1 K^2\Big(\frac{x-u}{b_n}\Big)p(u)(1-p(u))\,du + O\Big(\frac{1}{(nb_n)^2}\Big)\Bigg]^{1/2} \\
&\quad + O(b_n) + O\Big(\frac{1}{nb_n}\Big)\Bigg\} = O\Big(\frac{1}{\sqrt{n}\,b_n}\Big) + O(b_n^{1/2}) + O\Big(\frac{1}{nb^{3/2}}\Big).
\end{aligned}$$

Hence, $b_n^{-1/2}(\widetilde{\Delta}_n - \Delta(p)) \xrightarrow{\boldsymbol{P}} 0$. Similarly, we can show that $\widetilde{\sigma}_n^2 \xrightarrow{\boldsymbol{P}} \sigma^2(p)$.

Theorem 2. *Let $K(x) \in H(\tau)$ and $p_1(x) = p_2(x) = p(x) \in C^1[0,1]$. If $nb_n^2 \to \infty$, then, as $n \to \infty$,*

$$b_n^{-1/2}(T_n - \widetilde{\Delta}_n)\widetilde{\sigma}_n^{-1} \xrightarrow{d} N(0,1).$$

The *proof* follows from (23) and (24).

Theorem 2 enables us to construct an asymptotic criterion for the testing of the ***composite*** hypothesis

$$H_0: \quad p_1(x) = p_2(x), \quad x \in [0,1].$$

The critical domain for the testing of this hypothesis is given by the inequality

$$T_n \geq \widetilde{d}_n(\alpha) = \widetilde{\Delta}_n + b_n^{-1/2}\widetilde{\sigma}_n\lambda_\alpha, \quad \Phi(\lambda_\alpha) = 1 - \alpha. \tag{25}$$

Now let us investigate the asymptotic property of criterion (25) (i.e., the behavior of the power function as $n \to \infty$).

Theorem 3. *Let $K(x) \in H(\tau)$, $p_1(x), p_2(x) \in C^1[0,1]$. If $nb_n^2 \to \infty$, then*

$$\gamma_n(p_1,p_2) = \boldsymbol{P}_{H_1}\big(T_n \geq \widetilde{d}_n(\alpha)\big) \longrightarrow 1$$

as $n \to \infty$. Any pair $(p_1(x), p_2(x))$, $0 \leq p_i(x) \leq 1$, $p_i(x) \in C^1[0,1]$, $i = 1,2$, such that $p_1(x) \neq p_2(x)$ at at least one point x, $x \in [0,1]$. is an alternative of the hypothesis H_1.

Proof. Denote

$$\overline{T}_n = \frac{1}{2}\, nb_n \int\limits_{\Omega_n} \big(\overline{p}_{1n}(x) - \overline{p}_{2n}(x)\big)^2\, dx,$$

$$\overline{p}_{in}(x) = p_{in}(x) - \boldsymbol{E}p_{in}(x), \quad i = 1,2.$$

By analogy with (1), (2) and (24), we can readily show that the following is true for the hypothesis H_1

$$\begin{gathered}
b_n^{-1}\sigma_n^2 \longrightarrow \sigma^2(p_1,p_2) = 2\int\limits_0^1 d^2(x)\, dx \int\limits_{|x|\leq 2\tau} K_2^2(x)\, dx, \\
\widetilde{\sigma}_n^2 \xrightarrow{\boldsymbol{P}} \sigma^2(p_1,p_2), \quad \widetilde{\Delta}_n \xrightarrow{\boldsymbol{P}} \Delta(p_1,p_2), \quad \boldsymbol{E}\overline{T}_n \longrightarrow \Delta(p_1,p_2), \\
\Delta(p_1,p_2) = \int\limits_0^1 d(x)\, dx \int\limits_{|x|\leq\tau} K^2(x)\, dx, \\
d(x) = \frac{1}{2}\sum_{k=1}^2 p_k(x)(1-p_k(x)).
\end{gathered} \tag{26}$$

Further, in view of Lemma 2 and the fact that $\left[\frac{x-1}{b_n}, \frac{x}{b_n}\right] \supset [-\tau,\tau]$, $x \in \Omega_n(\tau)$ we obtain

$$\begin{aligned}
&\int\limits_{\Omega_n} \big(\boldsymbol{E}p_{1n}(x) - \boldsymbol{E}p_{2n}(x)\big)^2\, dx \\
&= \int\limits_{\Omega_n} \bigg(\int\limits_{-\tau}^{\tau} K(t)\big(p_1(x - b_n(u)) - p_2(x - b_n(u))\big)^2\, du\bigg)\, dx + O\Big(\frac{1}{nb_n}\Big).
\end{aligned}$$

According to the condition $p_1(x), p_2(x) \in C^1[0,1]$, we get

$$\int\limits_{\Omega_n} \big(\boldsymbol{E}p_{1n}(x) - \boldsymbol{E}p_{2n}(x)\big)^2\, dx = \int\limits_0^1 \big(p_1(x) - p_2(x)\big)^2\, dx + O(b_n) + O\Big(\frac{1}{nb_n}\Big). \tag{27}$$

By using (26) and (27), after simple transformations, we find

$$\begin{aligned}
\gamma_n(p_1,p_2) = \boldsymbol{P}_{H_1}\bigg[&\frac{\overline{T}_n - \boldsymbol{E}\overline{T}_n}{\sigma_n} \\
&\geq -nb_n^{1/2}\bigg(\int\limits_0^1 \big(p_1(x) - p_2(x)\big)^2\, dx + o_p(1)\bigg)\bigg].
\end{aligned} \tag{28}$$

Finally, since

$$(\overline{T}_n - \boldsymbol{E}\overline{T}_n)\sigma_n^{-1} \xrightarrow{d} N(0,1)$$

(the proof of this statement is similar to the proof of (22)) and $nb_n^{1/2} \to \infty$, it follows from (28) that $\gamma_n(p_1, p_2) \to 1$ as $n \to \infty$. $\square$

References

1. M. Aerts, N. Veraverbeke, Bootstrapping a nonparametric polytomous regression model. *Math. Methods Statist.* **4** (1995), no. 2, 189–200.
2. J. B. Copas, Plotting p against x. *Appl. Statist.* **32** (1983), no. 2, 25–31.
3. S. Efromovich, *Nonparametric Curve Estimation.* Methods, theory, and applications. Springer Series in Statistics. Springer-Verlag, New York, 1999.
4. R. Sh. Lipcer, A. N. Shirjaev, A functional central limit theorem for semimartingales. (Russian) *Teor. Veroyatnost. i Primenen.* **25** (1980), no. 4, 683–703.
5. H.-G. Müller, T. Schmitt, Kernel and probit estimates in quantal bioassay. *J. Amer. Statist. Assoc.* **83** (1988), no. 403, 750–759.
6. E. A. Nadaraya, On a regression estimate. (Russian) *Teor. Verojatnost. i Primenen.* **9** (1964), 157–159.
7. E. Nadaraya, P. Babilua, G. Sokhadze, Estimation of a distribution function by an indirect sample. *Ukr. Math. J.* **62** (2010), no. 12, 1642–1658.
8. H. Okumura, K. Naito, Weighted kernel estimators in nonparametric binomial regression. The International Conference on Recent Trends and Directions in Nonparametric Statistics. *J. Nonparametr. Stat.* **16** (2004), no. 1-2, 39–62.
9. G. S. Watson, Smooth regression analysis. *Sankhya Ser. A* **26** (1964), 359–372.

(Received 28.10.2019)

Department of Mathematics, Faculty of Exact and Natural Sciences, I. Javakhishvili Tbilisi State University, 3 University Str., Tbilisi 0143, Georgia

E-mail address: petre.babilua@tsu.ge

E-mail address: elizbar.nadaraya@tsu.ge

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 147–148

MARAO, ABOUT HOPF FIBRATIONS

GURAM BERISHVILI

Abstract. A marao is a cover of a vector space by a set of equidimensional subspaces with pairwise trivial intersections. Such structures give rise to fibrations of particular kind. Naturally occurring examples are described. In particular, it is explained how the classical Hopf fibrations can be uniformly obtained from maraos.

The stratification of the spheres is known, which Hopf noticed and published in 1931. It is a foliation of a three-dimensional sphere by circles, and this set of circles carries natural structure of a manifold diffeomorphic to a sphere of two dimensions. This situation is easy to describe, if we see a three-dimensional sphere as the set of rays of four-dimensional linear space. I must also draw attention to the analogy: a quotient space of linear space is the decomposition of the space into a set of affine subspaces, and Marao is the decomposition of the space into a set of linear subspaces.

Definition.

A set of linear subspaces of a linear space S is a marao if

– the intersection of each pair of two subspaces is zero

– the union of all subspaces is equal to the basic space S.

Notation: $M^k(S)$ is a marao of k-dimensional subspaces in the linear space S.

An example. The set of all one-dimensional subspaces of the linear space S is called projective space. This set meets the requirements of the definition and therefore it is the Marao $M^1(S)$.

The main example. Let $W \supset V$ be an extension of fields. Each W-linear space is also V-linear. The set of all W-linear subspaces of dimension one in a W-linear space S is a projective space. This set as the set of V-linear subspaces of the space S remains again a marao, but of larger dimension, $M^1(S)$ as seen from W and $M^k(S)$ as seen from V, $k = \dim_V W$.

From this example, we can show marao as a generalization of field extension.

The dimension of subspaces of the Marao does not exceed half of the dimension of the main space. The only exception is when Marao has a single element, Marao trivial, the only element of which is itself the linear space S. Marao M with only one subspace $S \in M$. Marao with dimension of subspaces half of the dimension of the main space will be called middle marao.

Let us have a marao $M^k(S)$. Let's consider the map from the set of all non-zero vectors in S to the marao $m : S^* \to M^k(S)$, $x \mapsto m(x) \ni x$, $x \in m(x) \in M^k(S)$. This map is part of the standard fibration over the Grassmanian, when viewing $M^k(S)$ as a subset of the Grassmanian of all k-dimensional linear subspaces of S. $S^* \to M$ and $S \to S/p$ are two orthogonal fibrations, for any $p \in M$: the fiber of one is a section in the other and vice versa.

Suppose that $M^k(S)$ is a marao in a linear space S of dimension n and the dimension of the members of the marao is k. Choose a point p of the marao and a complementary subspace A for p in S, of dimension $n - k$ and divide M into two subspaces M^1 and M^2. M^1 is the set of points of the marao which have zero intersection with A, while M^2 is the set of points having nonzero intersection with A, that is, elements of the marao which are completely in A or partially intersect A. Since a point of M^1 is completely (except zero, of course) outside A, it is represented as a graph of a linear map from p to A, hence M^1 can be identified with a subset of the space of all linear maps from p to A, $M^1 \subset \mathrm{Lin}(p, A)$. Consider the map $A^* \to M^k(S)$, $x \mapsto m(x) \in M^k(S)$. So the marao $M^k(S)$ is the union of the image of A^* under this map and of the subspace M^1 of the linear space $\mathrm{Lin}(p, A)$.

2020 *Mathematics Subject Classification.* 55R25, 57R22.

Key words and phrases. Marao; Projective space; Hopf fibration.

The above simplifies for middle maraos since in that case for a complementary subspace can be used any point q of the marao different from p, and in this case the map $(q = A)^* \to M^k(S)$ sends every vector to the same point q. Hence in this case the marao is obtained from $M^1 \subset \mathrm{Lin}(p, A)$ by adding the single point q. Moreover, in this case fixing any vector e not in q, the map $x \to m(e+x)$ gives a bijection between the linear space q and the set $M^k(S)\backslash\{q\}$ of all elements of the marao except q. Thus, a middle marao is obtained from the linear space q by adding a single point.

To a vector u from S/p can be assigned a subspace of the marao: to the vector x of u we associate its containing element $m(x)$ of the marao, and the subset of all $m(x)$ for $x \in u$ is denoted by $m(u) \subset M$, $x \in m(x) \in m(u) \subset M^k(S)$.

There are many structures on the Grassmannian: the natural linear bundle, the fiber over the point p being itself p as linear space, and the tangent space of the Grassmanian at its point p is naturally isomorphic to $\mathrm{Lin}(p, S/p)$. The map $m : S^* \to M$ induces linear map on tangent spaces $S \to T_pM \subset \mathrm{Lin}(p, S/p)$ with the kernel p. The tangent space T_pM is therefore isomorphic to the quotient space S/p.

Hopf fibration Suppose given a linear space S and in S a Marao $M^n(S)$ (dimension of the subspaces n), and given as well a marao in each of its elements C (dimension of the subspaces k). We have a total space $M^k(S)$ of the fibration (the union of the small Maraos $M^k(C)$; it is a marao in S with dimension of the subspaces k) and the base Marao $M^n(S)$. Over each point p the fiber is equal to the Marao $M^k(p)$. Such a fibration can be named as Hopf fibration since the famous Hopf fibrations are main examples.

$$(\mathbb{C}^8 = \mathbb{R}^{16})^* \to S^{15} \to \mathbb{R}P^{15} = M^1(\mathbb{R}^{16}) \to \mathbb{C}P^7 = M^2(\mathbb{R}^{16}) \to M^4(\mathbb{R}^{16}) \to M^8\mathbb{R}^{16} = S^8$$

fibers: ray, two opposite directional rays or the sphere S^0, $M^1(\mathbb{R}^2)$ or the sphere S^1, $M^2(\mathbb{R}^4)$ or the sphere S^2, $M^4(\mathbb{R}^8)$ or the sphere S^4, from the second to the end the fiber is rays of the 8-dimensional linear space or the sphere S^7.

$$(\mathbb{C}^4 = \mathbb{R}^8)^* \to S^7 \to \mathbb{R}P^7 = M^1(\mathbb{R}^8) \to M^2\mathbb{R}^8 = (\mathbb{C}P^3) \to M^4(\mathbb{R}^8) = S^4$$

fibers: ray, two opposite directional rays or the sphere S^0, M^1_2 or the sphere S^1, M^2_4 or the sphere S^2, from the second to the end the fiber is rays of the 4-dimensional linear space or the sphere S^3.

$$(\mathbb{C}^2 = \mathbb{R}^4)^* \to S^3 \to \mathbb{R}P^3 = M^1_4 \to \mathbb{C}P^1 = M^2_4 = S^2$$

fibers: ray, two opposite directional rays or the sphere S^0, M^1_2 or the sphere S^1, from the second to the end the fiber is rays of the 4-dimensional linear space or the sphere S^3.

$$(\mathbb{C} = \mathbb{R}^2)^* \to S^1 \to \mathbb{R}P^1 = M^1_2 = S^1$$

fibers: ray, two opposite directional rays or the sphere S^0, two points or sphere S^0.

References

1. Heinz Hopf, Über die Abbildungen der dreidimensionalen Sphäre auf die Kugelfiche. (German) *Math. Ann.* **104** (1931), no. 1, 637–665.

(Received 04.03.2020)

Georgian Technical University, 77 Kostava Str., Tbilisi 0175, Georgia
E-mail address: guram@berishvili.ge

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 149–154

CURIOSITIES REGARDING WAITING TIMES IN PÓLYA'S URN MODEL

NORBERT HENZE[1] AND MARK HOLMES[2]

Abstract. Consider an urn, initially containing b black and w white balls. Select a ball at random and observe its colour. If it is black, stop. Otherwise, return the white ball together with another white ball to the urn. Continue selecting at random, each time adding a white ball, until a black ball is selected. Let $T_{b,w}$ denote the number of draws until this happens. Surprisingly, the expectation of $T_{b,w}$ is infinite for the "fair" initial scenario $b = w = 1$, but finite if $b = 2$ and $w = 10^9$. In fact, $\mathbb{E}[T_{b,w}]$ is finite if and only if $b \geq 2$, and the variance of $T_{b,w}$ is finite if and only if $b \geq 3$, regardless of the number w of white balls. These observations extend to higher moments.

1. Introduction

The classical Pólya–Eggenburger urn is an elegant model in probability theory that is often presented in a first course on martingales (typically in a graduate probability theory course). In its simplest case, the model can be described as follows. Starting with b black and w white balls in an urn, choose a ball uniformly at random from the urn, observe the colour, return the chosen ball to the urn together with another ball of the same colour, then repeat. The number B_n (say) of times a black ball is drawn after n drawings has the well-known Pólya distribution

$$\mathbb{P}(B_n = k) = \binom{n}{k} \frac{\prod_{i=0}^{k-1}(b+i)\prod_{j=0}^{n-k}(w+j)}{\prod_{\ell=0}^{n-1}(b+w+\ell)}, \qquad k = 0, \ldots, n, \tag{1}$$

where an empty product is defined to be one, see, e.g., [4, p. 177]. It is easy to see that the proportion $X_n = (b + B_n)/(b + w + n)$ of black balls at time n is a bounded martingale (with respect to the natural filtration), with $B_0 = b/(b + w)$, and thus X_n converges almost surely to a random variable X. Here, X has a beta $\beta(b, w)$ distribution, see, e.g., [7, Theorem 2.1]. In the special case $b = w = 1$, equation (1) reduces to the discrete uniform distribution $\mathbb{P}(B_n = k) = 1/(n+1)$, and the limit X has a standard uniform distribution.

For later purposes, it will be convenient to regard the distribution of B_n as a special case of a Beta-binomial distribution, see, e.g., [5, p. 242]. The latter distribution originates as follows: Let P have a Beta $\beta(u, v)$-distribution, where $u, v > 0$. Suppose that, conditionally on $P = p$, the random variable M has a binomial distribution $\mathrm{Bin}(n, p)$. Then, for $k \in \{0, 1, \ldots, n\}$, we have

$$\mathbb{P}(M = k) = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{1}{\mathrm{B}(u,v)} p^{u-1}(1-p)^{v-1}\,\mathrm{d}p \tag{2}$$

$$= \binom{n}{k} \frac{\mathrm{B}(u+k, v+n-k)}{\mathrm{B}(u,v)}, \tag{3}$$

where $\mathrm{B}(\cdot,\cdot)$ is the Beta function. The distribution of M is called the Beta-binomial distribution with parameters n, u and v. By using the relation $\mathrm{B}(u, v) = \Gamma(u)\Gamma(v)/\Gamma(u + v)$, where $\Gamma(\cdot)$ is the Gamma function, we see that the distribution of B_n is obtained from (3) by putting $u = b$ and $w = v$.

Inverse Pólya distributions originate if one asks for the number of drawings needed to observe a specified number of black balls under the above or more general replacement schedules, see, e.g., [4, p. 192]. Paper [3] considers waiting times for the first occurrence of a specified pattern in Pólya's urn scheme. A special case is the waiting time until the first occurrence of a black ball, which we will

2020 *Mathematics Subject Classification.* Primary 60E99, Secondary 00A08.
Key words and phrases. Pólya's urn model; Waiting time; Inverse Pólya distribution.

focus on in this note. For the recent work on the inverse Pólya distributions, see, e.g., [1,2], and [6]. In what follows, we consider some curiosities concerning the (random) time until we first draw a black ball, denoted by $T_{w,b}$, that evidently have not been highlighted before.

2. One Black Ball

We first consider the standard "fair" case where the urn contains one black and one white ball at the outset. We then have

$$\mathbb{P}(T_{1,1} > n) = \frac{1}{2} \cdot \frac{2}{3} \cdot \cdots \cdot \frac{n-1}{n} \cdot \frac{n}{n+1} = \frac{1}{n+1}$$

and thus $\mathbb{P}(T_{1,1} < \infty) = 1$. Hence, the black ball will be drawn with probability one in finite time. However, since $\sum_{n=0}^{\infty} \mathbb{P}(T_{1,1} > n) = \infty$, *the expectation of $T_{1,1}$ is infinite.*

In view of $\mathbb{P}(T_{1,1} = j) = \mathbb{P}(T_{1,1} > j-1) - \mathbb{P}(T_{1,1} > j) = 1/(j(j+1))$, notice that the conditional expectation of $T_{1,1}$, given $T_{1,1} \le k$, is

$$\mathbb{E}[T_{1,1}|T_{1,1} \le k] = \frac{1}{\mathbb{P}(T_{1,1} \le k)} \sum_{j=1}^{k} j\, \mathbb{P}(T_{1,1} = j) = \frac{(k+1)}{k} \sum_{j=1}^{k} \frac{1}{j+1}.$$

Using $\sum_{j=1}^{n} \frac{1}{j} = \log n + \gamma + o(1)$, where $\gamma = 0.57721\ldots$ is the Euler–Mascheroni constant, it follows that

$$\mathbb{E}[T_{1,1}|T_{1,1} \le k] = \log k + \gamma - 1 + o(1), \qquad \text{as } k \to \infty.$$

In other words, given that you have selected a black ball by time k, on average you first picked one at a relatively early time of $\log(k)$. This is intuitively reasonable because it is much easier to choose a black ball for the first time at an early time, before white balls have been reinforced too much. Indeed, for large k, we find that $\mathbb{P}(T_{1,1} > k/2|T_{1,1} \le k)$ is of order $1/k$.

We incidentally note that the probability that $T_{1,1}$ takes an odd value equals $\log 2$, since

$$\begin{aligned}\sum_{\ell=0}^{\infty} \mathbb{P}(T_{1,1} = 2\ell+1) &= \sum_{\ell=0}^{\infty} \frac{1}{(2\ell+1)(2\ell+2)} = \sum_{\ell=0}^{\infty} \left(\frac{1}{2\ell+1} - \frac{1}{2\ell+2} \right) \\ &= \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j}.\end{aligned}$$

Continue to set $b = 1$, but now allow w to be arbitrarily large. Since

$$\mathbb{P}(T_{1,w} > n) = \frac{w}{w+1} \cdot \frac{w+1}{w+2} \cdot \cdots \cdot \frac{w+n-2}{w+n-1} \cdot \frac{w+n-1}{w+n} = \frac{w}{w+n},$$

it follows that $\mathbb{P}(T_{1,w} < \infty) = 1$, regardless of the number of white balls. If, for example, $w = 10^9$, drawing the only black ball seems to be like finding a needle in a haystack, but you have time beyond all limits, and the situation of having one black and 10^9 white balls in the urn could have happended in the course of the stochastic process involving over time under the initial scenario $b = w = 1$ after $10^9 - 1$ draws.

3. A Second Black Ball Works Wonders

Suppose now that at the beginning there are $b = 2$ black and w white balls in the urn. We now have

$$\mathbb{P}(T_{2,w} > n) = \frac{w}{w+2} \cdot \frac{w+1}{w+3} \cdot \frac{w+2}{w+4} \cdot \cdots \cdot \frac{w+n-1}{w+n+1} = \frac{w(w+1)}{(w+n)(w+n+1)}.$$

Since $\sum_{n=1}^{\infty} \mathbb{P}(T_{2,w} > n) < \infty$, we do not only have $\mathbb{P}(T_{2,w} < \infty) = 1$, but, in addition, *the expectation of $T_{2,w}$ is finite, irrespective of the number of white balls.* More specifically, we have

$$\mathbb{E}[T_{2,w}] = \sum_{k=0}^{\infty} \mathbb{P}(T_{2,w} > k) = w(w+1) \sum_{k=0}^{\infty} \frac{1}{(w+k)(w+k+1)} = w+1.$$

Here, the last equality follows because the series is telescoping.

Remark 3.1. Starting from $b = 1, w = 1$, we may continue observing Pólya's urn after $T_{1,1}$ until the time $T_{1,1}^{(2)}$ at which we draw a second black ball. At the time $T_{1,1}$ that we first draw a black ball, we return it and add another, so there are then 2 black balls and $T_{1,1}$ white balls. Since $\mathbb{E}[T_{1,1}^{(2)} - T_{1,1} \,|\, T_{1,1} = w] = \mathbb{E}[T_{2,w}] = w+1$, we know that this expectation is finite for every w. We can interpret this as $\mathbb{E}[T_{1,1}^{(2)} - T_{1,1} \,|\, T_{1,1}] = T_{1,1}+1$, or "given the value of $T_{1,1}$, the expected additional time required to draw a second black ball is finite" (a.s.). Nevertheless, $\mathbb{E}[T_{1,1}^{(2)} - T_{1,1}] = \mathbb{E}[T_{1,1} + 1] = \infty$.

4. The General Case

We now assume that the initial configuration is b black and w white balls. The event that each of the first n draws yields a white ball has probability

$$\begin{aligned}\mathbb{P}(T_{b,w} > n) &= \prod_{i=0}^{n-1} \frac{w+i}{b+w+i} \\ &= \frac{(b+w-1)!}{(w-1)!} \cdot \frac{(w-1+n)!}{(b+w-1+n)!}, \qquad n \geq 1.\end{aligned}$$

The first ratio does not depend on n, and the second is equal to

$$\frac{1}{(w+n) \cdot \dots \cdot (b+w-1+n)}. \tag{4}$$

It immediately follows that $\mathbb{P}(T_{b,w} < \infty) = 1$, but we can infer more from (4). To this end, notice that this expression is bounded from below by $(b+w+n)^{-b}$ and from above by n^{-b}, which, for each integer r, shows that

$$\begin{aligned}\mathbb{E}\big[T_{b,w}^r\big] &= \sum_{n=1}^{\infty} n^r \mathbb{P}(T_{b,w} = n) \\ &= \sum_{n=1}^{\infty} n^r \frac{(b+w-1)!}{(w-1)!} \frac{(w+n-2)!}{(b+w+n-2)!} \frac{b}{b+w+n-1} \\ &= \sum_{n=1}^{\infty} n^r O(n^{-(b+1)}).\end{aligned}$$

Hence, $\mathbb{E}\big[T_{b,w}^r\big] < \infty$ if and only if $b > r$. Surprisingly, this moment condition does not depend on the number w of white balls. In particular, the variance of $T_{b,w}$ exists if and only if there are at least 3 black balls in the urn at the beginning. In the case $b = 3$, straightforward calculations involving telescoping series yield $\mathbb{E}[T_{3,w}] = (w+2)/2$, and, using the fact that $\mathbb{E}[L^2] = \sum_{n=0}^{\infty}(2n+1)\mathbb{P}(L > n)$ for a nonnegative integer-valued random variable L, we have $\mathbb{E}[T_{3,w}^2] = (w+2)(2w+1)/2$, and thus the variance is $\mathbb{V}(T_{3,w}) = 3w(w+2)/4$.

Remark 4.1. In [8], one finds the general formula

$$\mathbb{E}[T_{b,w}] = \frac{b+w-1}{b-1} \tag{5}$$

if $b \geq 2$, which was obtained from a hypergeometric series. As remarked in [9], (5) follows readily from (2), since, conditionally on $P = p$, drawings are according to an independent and identically distributed Bernoulli sequence with probability of success given by p, where success means drawing a black ball. Since, conditionally on $P = p$, the distribution of $T_{b,w}$ is geometric, we have $\mathbb{E}[T_{b,w}|P = p] = 1/p$ and thus

$$\begin{aligned}\mathbb{E}[T_{b,w}] &= \int_0^1 \mathbb{E}[T_{b,w}|P = p] \, \frac{1}{\mathrm{B}(b,w)} \, p^{b-1}(1-p)^{w-1} \, \mathrm{d}p = \frac{\mathrm{B}(b-1,w)}{\mathrm{B}(b,w)} \\ &= \frac{b+w-1}{b-1}.\end{aligned}$$

From (2) and the fact that $\mathbb{V}(T_{b,w}) = \mathbb{E}[\mathbb{V}(T_{b,w}|P)] + \mathbb{V}(\mathbb{E}[T_{b,w}|P])$, we can also obtain a general formula for the variance of $T_{b,w}$ if $b \geq 3$. Since the conditional variance of $T_{b,w}$, given $P = p$, is the variance of a geometric distribution with parameter p and thus equal to $(1-p)/p^2$, a straightforward algebra gives

$$\mathbb{E}[\mathbb{V}(T_{b,w}|P)] = \int_0^1 \frac{1-p}{p^2} \frac{1}{\mathrm{B}(b,w)} p^{b-1}(1-p)^{w-1}\,\mathrm{d}p = \frac{w(b+w-1)}{(b-1)(b-2)}.$$

Furthermore, $\mathbb{E}[T_{b,w}|P] = 1/P$, and thus some algebra yields

$$\mathbb{V}(\mathbb{E}[T_{b,w}|P]) = \frac{w(b+w-1)}{(b-1)^2(b-2)}.$$

Summing up, we obtain

$$\mathbb{V}(T_{b,w}) = \frac{bw(b+w-1)}{(b-1)^2(b-2)}.$$

Notice that, in view of $\mathbb{E}[T_{b,w}^\ell] = \mathbb{E}\big[\mathbb{E}[T_{b,w}^\ell|P]\big]$, one can fairly easily even obtain closed-form expressions for higher moments of $T_{b,w}$.

5. A General Replacement Scheme

Suppose now that if a white ball shows up at time k, we return this ball and additionally a_k white balls, where $a_k \geq 1$. Notice that this flexible model includes the special case $a_k = 1$ that has been considered so far, but also the case that a constant number larger than one of white balls is returned to the urn together with the chosen ball. The following result gives a necessary and sufficient condition on the sequence (a_k) for the probability that a black ball shows up at a finite time.

Lemma 5.1. *Let $s_k = a_1 + \cdots + a_k$, $k \geq 1$. We then have*

$$\mathbb{P}(T_{b,w} < \infty) = 1 \iff \sum_{j=1}^{\infty} \frac{1}{s_j} = \infty.$$

Proof. Putting $s_0 = 0$, we have

$$\mathbb{P}(T_{b,w} > n) = \prod_{j=0}^{n-1} \frac{w + s_j}{b + w + s_j}.$$

Using the inequalities $1 - 1/t \leq \log t \leq t - 1$, $t > 0$, straightforward calculations yield

$$-b \sum_{j=0}^{n-1} \frac{1}{w + s_j} \leq \log \mathbb{P}(T_{b,w} > n) \leq -b \sum_{j=0}^{n-1} \frac{1}{b + w + s_j}.$$

Hence $\log \mathbb{P}(T_{b,w} > n) \to -\infty$ as $n \to \infty$ if and only if the series $\sum_{j=0}^{\infty} 1/s_j$ diverges, and the assertion follows. $\square$

From this result it follows that $\mathbb{P}(T_{b,w} < \infty) = 1$ even if $b = 1$, w is arbitrarily large, and a fixed huge number of additional white balls is added to the urn after each draw of a white ball, but not if at the kth time we select a white ball we return it and add k extra white balls, for example.

In the case where we add a constant c additional number of white balls to the urn whenever we select a white ball, we can also consider the expected time to select a black ball.

Lemma 5.2. *In the case where we start with w white balls and b black balls in the urn, and add $c \geq 1$ additional white balls whenever white is selected from the urn, we find that $\mathbb{E}[T_{b,w}] < \infty$ if and only if $b > c$.*

Proof. In this context we can write

$$\begin{aligned}\mathbb{P}(T_{b,w} > n) &= \prod_{j=0}^{n-1} \frac{\frac{w}{c}+j}{\frac{b}{c}+\frac{w}{c}+j} \\ &= \frac{\frac{w}{c}}{\frac{b}{c}+\frac{w}{c}} \times \frac{\frac{w}{c}+1}{\frac{b}{c}+\frac{w}{c}+1} \times \cdots \times \frac{\frac{w}{c}+n-2}{\frac{b}{c}+\frac{w}{c}+n-2} \times \frac{\frac{w}{c}+n-1}{\frac{b}{c}+\frac{w}{c}+n-1}.\end{aligned}$$

If $b/c \leq 1$, then the numerator of the $j+1$st term in the product is greater than or equal to the denominator of the jth term and so this product is at least

$$\frac{\frac{w}{c}}{\frac{w}{c}+\frac{b}{c}+n-1},$$

which is not summable in n, so the expectation of $T_{b,n}$ is infinite.

If $b/c \geq 2$, then the numerator of the $j+2$nd term in the product is no larger than the denominator of the jth term, so for some constant a we have $\mathbb{P}(T_{b,w} > n) \leq an^{-2}$ for all n sufficiently large. This is summable in n, so the expectation is finite when $b/c \geq 2$.

The case $b/c \in (1,2)$ can be handled by a slightly more elaborate (but standard) approach, which we now quickly present. We can write

$$\mathbb{P}(T_{b,w} > n) = \prod_{j=0}^{n-1} \left(1 - \frac{\frac{b}{c}}{\frac{b}{c}+\frac{w}{c}+j}\right) \leq \exp\left\{ -\frac{b}{c} \sum_{j=0}^{n-1} \frac{1}{\frac{b}{c}+\frac{w}{c}+j} \right\},$$

where we have used $1-x \leq e^{-x}$ and that the product of exponentials is the exponential of a sum. For $n \geq 1$, the sum is at least $\int_0^{n-1} \frac{1}{d+x} \mathrm{d}x = \log(d+n-1) - \log(d)$, where $d = (b+w)/c > 0$. Thus for $n \geq 1$,

$$\mathbb{P}(T_{b,w} > n) \leq d^{b/c} \exp\left\{ \log\left((n-1+d)^{-b/c} \right) \right\} = \frac{d^{b/c}}{(n-1+d)^{b/c}}.$$

Since $b/c > 1$, this is summable in n. $\square$

Acknowledgement

The work of MH is supported by Future Fellowship FT160100166 from the Australian Research Council. The authors thank a referee whose comments helped to improve the paper. The authors thank the organisers of the International Conference on Probability Theory and Statistics in Tbilisi, 2019, for bringing them together. Finally, the authors thank and congratulate Estate Khmaladze for many years of outstanding contributions to statistics and probability.

References

1. Ch. A. Charalambides, A q-Pólya urn model and the q-Pólya and inverse q-Pólya distributions. *J. Statist. Plann. Inference* **142** (2012), no. 1, 276–288.
2. T. Garg, P. N. Agrawal, A. Kajla, Jain-Durrmeyer operators involving inverse Pólya-Eggenberger distribution. *Proc. Nat. Acad. Sci. India Sect. A* **89** (2019), no. 3, 547–557.
3. K. Inoue, S. Aki, Generalized waiting time problems associated with pattern in Pólya's urn scheme. *Ann. Inst. Statist. Math.* **54** (2002), no. 3, 681–688.
4. N. L. Johnson, S. Kotz, *Urn Models and their Application.* An approach to modern discrete probability theory. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-London-Sydney, 1977.
5. N. L. Johnson, S. Kotz, A. Kemp, *Univariate Discrete Distributions.* Second edition. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1992.
6. F. Makri, A. N. Philippou, Z. M. Psillakis, Pólya, inverse Polya, and circular Polya distributions of order k for l-overlapping success runs. *Comm. Statist. Theory Methods* **36** (2007), no. 1-4, 657–668.
7. R. Pemantle, A survey of random processes with reinforcement. *Probab. Surv.* **4** (2007), 1–79.
8. https://math.stackexchange.com/questions/2560846/polya-urn-mean-number-of-draws-to-get-a-specific-sequence-of-colors

9. https://mathoverflow.net/questions/289247/polya-urn-mean-number-of-draws-to-get-a-specific-sequence-of-colors

(Received 28.02.2020)

[1]Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Englerstr. 2, D-76133 Karlsruhe
E-mail address: Norbert.Henze@kit.edu

[2]University of Melbourne, School of Mathematics and Statistics, Peter Hall Building, Vic 3010, Australia
E-mail address: holmes.m@unimelb.edu.au

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 155–173

PROJECTION APPROACH TO DISTRIBUTION-FREE TESTING FOR POINT PROCESSES. REGULAR MODELS

ESTATE V. KHMALADZE

Abstract. We create the notion of equivalence between different martingale models for point processes. This allows to map one model into another model in the same equivalence class. Therefore the distribution of test statistics for goodness of fit testing needs to be calculated in only once, for "standard" model, in each equivalence class. The equivalence classes are surprisingly broad, and thus the economy on computational work is considerable.Namely, any such class includes a non-time homogeneous Poisson model. Therefore it is sufficient to know the distribution of test statistics only for Poisson models.

The situation, therefore, becomes comparable to testing simple hypothesis about a continuous distribution function for a sample of i.i.d. random variables with continuous distribution F, when it is sufficient to consider F, uniform on $[0, 1]$. However, for point processes we consider here parametric cases, and the nature of equivalence is entirely different.

1. In Place of Introduction

This text was mainly written as a basic background material for the project which I was working on with Dr. S. Umut Can and Prof. R. Laeven from the University of Amsterdam. The aim of the project is to establish equivalence between testing parametric models for point processes with different forms of random intensities. Eventually, we intend to show that a huge majority of testing such models is equivalent to that of the non-time-homogeneous Poisson process which involves estimated parameters.

The text is not yet the final version, it is even not completely finished, but as it is, it may be useful for many readers. It is the first general and unified text with the material, which can be either found in various papers, or is new.

Umut Can greatly helped in preparation of the text and Roger Laeven made a number of useful remarks and I am grateful to both.

2. Basic Asymptotic Set-up

The method we want to develop for the testing problems for intensities of point processes can be first explained by drawing parallels between point processes and empirical processes, as the method for the latter has already been developed (see [6–8]).

Given a sample, i.e., a collection of independent and identically distributed (i.i.d.) positive random variables $X_1, \dots, X_n$, let us first consider the so-called *binomial process*

$$Z_n(t) = \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le t\}} = \sum_{i=1}^{n} \mathbb{1}_{\{X_{(i:n)} \le t\}}, \quad t \ge 0. \tag{1}$$

Here, $X_{(i:n)}$ denotes the i^{th} order statistic of the sample $X_1, \dots, X_n$, with $X_{(1:n)} = \min\{X_1, \dots, X_n\}$ and $X_{(n:n)} = \max\{X_1, \dots, X_n\}$. Also, $\mathbb{1}_E$ denotes the indicator function of the event E, so for example,

$$\mathbb{1}_{\{X_i \le t\}} = \begin{cases} 1 & \text{if } X_i \le t \\ 0 & \text{otherwise} \end{cases}, \quad t \ge 0.$$

2020 *Mathematics Subject Classification.* 62C07, 62E20, 62F03, 60G55.

Key words and phrases. Martingale models for point processes; Models with estimated parameters; Asymptotic methods; Unitary operators.

For a given X_i, the indicator function $\mathbb{1}_{\{X_i \le t\}}$ is a step function of t, and since $X_1, \dots, X_n$ are i.i.d. random variables, $\mathbb{1}_{\{X_1 \le t\}}, \dots, \mathbb{1}_{\{X_n \le t\}}$ are i.i.d. stochastic processes in t. If we fix the value of $t > 0$, then $\mathbb{1}_{\{X_1 \le t\}}, \dots, \mathbb{1}_{\{X_n \le t\}}$ become independent Bernoulli random variables with

$$P[\mathbb{1}_{\{X_i \le t\}} = 1] = P[X_i \le t] = F(t),$$

where F denotes the common distribution function of the X_i's. It now follows from the first equality in (1) that $Z_n(t) \sim \mathrm{Binom}(n, F(t))$ and, in particular, $\mathbb{E}[Z_n(t)] = nF(t)$. It also follows from the Central Limit Theorem that for any $t > 0$,

$$v_n(t) := \frac{1}{\sqrt{n}}[Z_n(t) - nF(t)] \tag{2}$$

is asymptotically Gaussian as $n \to \infty$. In fact, we know from the Functional Limit Theorem that not just $v_n(t)$ for any given $t > 0$ is asymptotically Gaussian, but the stochastic process $\{v_n(t) : t \ge 0\}$ is asymptotically Gaussian as well, in the sense that it converges weakly to a Gaussian process v. The process v_n is called the empirical process associated with the sample $X_1, \dots, X_n$, and the limiting Gaussian process v is called the F-Brownian bridge. Occasionally, it will be convenient to use the notation $F_n(t) = Z_n(t)/n$ for an empirical distribution function and write empirical process v_n in the equivalent form

$$v_n(t) := \sqrt{n}[F_n(t) - F(t)].$$

For these and many more nice facts about empirical processes we refer the readers to the monograph [10]. Some of these facts may not be, however, very visible from the second definition in (1). Indeed, the random variables $X_{(1:n)}, \dots, X_{(n:n)}$ are neither independent nor identically distributed. Although $\mathbb{1}_{\{X_{(1:n)} \le t\}}, \dots, \mathbb{1}_{\{X_{(n:n)} \le t\}}$ are still the Bernoulli random variables for any fixed $t \ge 0$, they are now very much dependent, and the distribution functions

$$F_{(i:n)}(t) := P[X_{(i:n)} \le t]$$

are very different for different i. The properly centered form of $Z_n(t)$ taken from the second definition in (1) is, therefore,

$$Z_n(t) - \sum_{i=1}^{n} F_{(i:n)}(t), \quad t \ge 0, \tag{3}$$

and it is almost an accident that

$$\sum_{i=1}^{n} F_{(i:n)}(t) = nF(t).$$

The second definition in (1) has, however, the advantage that it represents $Z_n(t)$ as a point process with order statistics corresponding to arrival times: $X_{(i:n)}$ can be interpreted as the arrival time of the i^{th} event.

As the martingale theory of point processes is well-developed and widely known, almost nobody would center point processes by their unconditional expected values as in (3). What is done instead is the *conditional* centering of increments of $Z_n(t)$ given the past history of this process:

$$\mathrm{d}Z_n(t) - \mathbb{E}[\mathrm{d}Z_n(t) \,|\, Z_n(s), 0 \le s \le t] =: \mathrm{d}M_n(t). \tag{4}$$

The resulting process $\{M_n(t) : t \ge 0\}$ is a martingale and the equality (4) itself is called the Doob-Meyer decomposition of $Z_n(t)$, which we now view as a submartingale.

Let us now define

$$\lambda_n(t) = \mathbb{E}[\mathrm{d}Z_n(t) \,|\, Z_n(s), 0 \le s \le t]/\mathrm{d}t, \quad t \ge 0,$$

which is called the *intensity* of the point process $Z_n(t)$, and

$$w_n(t) = \frac{1}{\sqrt{n}} M_n(t) = \frac{1}{\sqrt{n}} \left[Z_n(t) - \int_0^t \lambda_n(s)\,\mathrm{d}s \right],$$

which is also a martingale in t. Thus from $Z_n(t)$ we have produced, using different methods of centering, two different processes, namely, the empirical process $v_n(t)$ and the process $w_n(t)$, which we will refer to as an *innovation martingale* of the process $Z_n(t)$. Yet, we will see below that there is a very important similarity between the asymptotic behavior of v_n and w_n in the practically important case when the underlying distribution function F depends on some finite-dimensional parameter θ, and when the random intensity λ_n also depends on such a parameter.

In the context of goodness of fit testing, when the null hypothesis does not completely specify the distribution function F, but only states that it belongs to a parametric family $\{F_\theta : \theta \in \Theta\}$, with $\Theta \subset \mathbb{R}^m$, we call this hypothesis a parametric hypothesis. The same term is used if we hypothesize that the intensity of Z_n belongs to a parametric family of intensities $\{\lambda_{n,\theta} : \theta \in \Theta\}$. In the case of a parametric hypothesis, we will need to estimate the parameter θ and then to make a judgment on whether the hypothesis is true or not, by observing the behavior of the processes

$$\frac{1}{\sqrt{n}}[Z_n(t) - nF_{\widehat{\theta}}(t)] = v_{n,\widehat{\theta}}(t) = \widehat{v}_n(t)$$

and

$$\frac{1}{\sqrt{n}}\left[Z_n(t) - \int_0^t \lambda_{n,\widehat{\theta}}(s)\,\mathrm{d}s\right] = w_{n,\widehat{\theta}}(t) = \widehat{w}_n(t),$$

respectively. The 'similarity' that was alluded to above consists in the fact that $\widehat{v}_n$ is asymptotically a projection of v_n, and $\widehat{w}_n$ is asymptotically a projection of w_n; substituting the estimate $\widehat{\theta}$ in place of the true parameter θ is asymptotically equivalent to projecting the initial process. Thus if we have a method that exploits this geometric fact in the case a parametric hypothesis about distribution functions, it should be possible to develop a similar method in the situation with point processes.

Let us now review why we have a projection in the case of a parametric hypothesis about F. Suppose that $\{F_\theta : \theta \in \Theta\}$ is a *regular* parametric family of distributions in the following sense:

(a_1) the space Θ of feasible parameter values is an open subset of the Euclidean space $\mathbb{R}^m$;

(a_2) the vector of the derivatives

$$\frac{\partial}{\partial\theta}\ln f_\theta(x) = [\dot{f}/f]_\theta(x)$$

is square-integrable, i.e., the Fisher information matrix

$$R_\theta = \int [\dot{f}/f]_\theta(x)\,[\dot{f}/f]_\theta^{\mathsf{T}}(x)\,f_\theta(x)\,\mathrm{d}x$$

is finite and non-degenerate for every $\theta \in \Theta$,

(a_3) for any $\theta \in \Theta$;

$$\int [\dot{f}/f]_\theta(x)\,f_\theta(x)\,\mathrm{d}x = 0.$$

The openness of Θ is useful because then every θ has a neighborhood in Θ and we can differentiate at θ without worrying about boundary effects. Conditions (a_2) and (a_3) are ubiquitous in all asymptotic statistics with regular parametric families.

To describe the difference between $\widehat{v}_n$ and v_n we first need an asymptotic representation of the maximum likelihood estimator (MLE) $\widehat{\theta}$, or rather, of $\sqrt{n}(\widehat{\theta} - \theta)$. The MLE is the (correctly chosen) root of the maximum likelihood equation

$$\sum_{i=1}^n [\dot{f}/f]_{\widehat{\theta}}(X_i) = 0. \tag{5}$$

Using the regularity condition (a_3), we can rewrite (5) as

$$\int [\dot{f}/f]_{\widehat{\theta}}(x)\,[\mathrm{d}Z_n(x) - n\mathrm{d}F_{\widehat{\theta}}(x)] = 0,$$

that is,

$$\int [\dot{f}/f]_{\widehat{\theta}}(x)\,\mathrm{d}v_{n,\widehat{\theta}}(x) = 0. \tag{6}$$

Replacing the left-hand side of (6) by the Taylor expansion around θ, we obtain

$$0 = \int [\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x) + \int \frac{\partial}{\partial\theta}[\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x)(\widehat\theta - \theta)$$
$$-\sqrt{n}\int [\dot f/f]_\theta(x)\,\dot f_\theta(x)^{\mathsf T}\mathrm{d}x(\widehat\theta - \theta) + o_P(1). \tag{7}$$

Here, the assumption that the residual term is indeed $o_P(1)$ is, actually, another regularity assumption, (a_4), on the family $\{F_\theta : \theta \in \Theta\}$. Note that we can write the second term in the right-hand side of (7) as

$$\frac{1}{\sqrt{n}}\int \frac{\partial}{\partial\theta}[\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x)\sqrt{n}(\widehat\theta - \theta)$$
$$= \int \frac{\partial}{\partial\theta}[\dot f/f]_\theta(x)\,\mathrm{d}\,[F_n(x) - F_\theta(x)]\,\sqrt{n}(\widehat\theta - \theta),$$

which is asymptotically negligible as long as the matrix

$$\frac{\partial}{\partial\theta}\left[\dot f/f\right]_\theta(x) = \frac{\partial^2}{\partial\theta^2}\ln f_\theta(x)$$

is integrable with respect to F_θ – this follows from the Law of Large Numbers. Using the regularity assumption (a_2) for the third term on the right-hand side of (7), we obtain

$$0 = \int [\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x) - R_\theta\sqrt{n}(\widehat\theta - \theta) + o_P(1),$$

or equivalently,

$$\sqrt{n}(\widehat\theta - \theta) = R_\theta^{-1}\int [\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x) + o_P(1), \tag{8}$$

which is the asymptotic MLE representation we wanted.

Now, let us apply the Taylor expansion again and write

$$v_{n,\widehat\theta}(t) = v_{n,\theta}(t) - \left[\frac{\partial}{\partial\theta}F_\theta(t)\right]^{\mathsf T}\sqrt{n}(\widehat\theta - \theta) + o_P(1),$$

or, by virtue of (8),

$$v_{n,\widehat\theta}(t) = v_{n,\theta}(t) - \int_0^t [\dot f/f]_\theta^{\mathsf T}(s) f_\theta(s)\,\mathrm{d}s\, R_\theta^{-1}\int [\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x) + o_P(1). \tag{9}$$

The main part on the right-hand side of (9) is a linear transformation of $v_{n,\theta}$; moreover, as the proposition below shows, it is a projection.

Proposition 2.1. *The linear operator* Π *defined by*

$$\Pi\gamma(t) = \gamma(t) - \int_0^t [\dot f/f]_\theta^{\mathsf T}(s)\,\mathrm{d}F_\theta(s)\,R_\theta^{-1}\int [\dot f/f]_\theta(x)\,\mathrm{d}\gamma(x)$$

is an orthogonal projector, i.e., it satisfies the conditions

(i) $\Pi\Pi\gamma(t) = \Pi\gamma(t)$,

(ii) $\Pi\gamma(t) \equiv 0 \Leftrightarrow \dfrac{\mathrm{d}\gamma}{\mathrm{d}F}(t) = c^{\mathsf T}[\dot f/f]_\theta(t)$ *for some* $c \in \mathbb{R}^m$,

(iii) $\displaystyle\int [\dot f/f]_\theta(s)\,\mathrm{d}\Pi\gamma(s) = 0.$

This fact has several useful consequences which we will discuss later. Right now we would like to establish the analogous result for point processes.

Given a point process $\{N_n(t) : t \geq 0\}$, let

$$\lambda_{n,\theta}(t)\mathrm{d}t = \mathbb{E}[\mathrm{d}N_n(t)\,|\,N_n(s), 0 \leq s \leq t]$$

denote the intensity as above, and let $t_0 = 0 < t_1 < \cdots < t_k = T$ be a partition of the interval $[0, T)$, where we are considering our point process. If the partition is sufficiently fine, the likelihood of the vector $(N_n(t_1), \ldots, N_n(t_k))^\mathsf{T}$ is given as a product of Bernoulli's likelihoods:

$$\prod_{j=1}^{k} \left[\lambda_{n,\theta}(t_j)\Delta t_j\right]^{\Delta N_n(t_j)} \left[1 - \lambda_{n,\theta}(t_j)\Delta t_j\right]^{1-\Delta N_n(t_j)}$$
$$= \exp\left\{ \sum_{j=1}^{k} \Delta N_n(t_j) \ln\left[\lambda_{n,\theta}(t_j)\Delta t_j\right] + \sum_{j=1}^{k} \left(1 - \Delta N_n(t_j)\right) \ln\left[1 - \lambda_{n,\theta}(t_j)\Delta t_j\right] \right\}. \tag{10}$$

The likelihood of the same vector under the assumption that N_n is a Poisson process with constant intensity λ has a similar form, with $\lambda_{n,\theta}$ replaced by λ. This Poisson likelihood is only a reference likelihood and we could have used many other measures in order to create likelihood ratios. If we take the limit of the likelihood in (10) as $k \to \infty$ and $\max_j\{\Delta t_j : 1 \le j \le k\} \to 0$, we will obtain zero, but the likelihood ratio below will have a non-trivial limit:

$$\frac{\prod_{j=1}^{k} \left[\lambda_{n,\theta}(t_j)\Delta t_j\right]^{\Delta N_n(t_j)} \left[1 - \lambda_{n,\theta}(t_j)\Delta t_j\right]^{1-\Delta N_n(t_j)}}{\prod_{j=1}^{k} \left[\lambda\Delta t_j\right]^{\Delta N_n(t_j)} \left[1 - \lambda\Delta t_j\right]^{1-\Delta N_n(t_j)}}$$
$$= \exp\left\{ \sum_{j=1}^{k} \Delta N_n(t_j) \ln \frac{\lambda_{n,\theta}(t_j)}{\lambda} + \sum_{j=1}^{k} \left(1 - \Delta N_n(t_j)\right) \ln \frac{1 - \lambda_{n,\theta}(t_j)\Delta t_j}{1 - \lambda\Delta t_j} \right\}$$
$$\to \exp\left\{ \int_0^T \ln \frac{\lambda_{n,\theta}(t)}{\lambda}\, \mathrm{d}N_n(t) - \int_0^T [\lambda_{n,\theta}(t) - \lambda]\, \mathrm{d}t \right\}.$$

Differentiating this log-likelihood ratio with respect to θ and setting the result equal to zero, we obtain the maximum likelihood equation

$$\int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(t)\, \mathrm{d}N_n(t) - \int_0^T \dot{\lambda}_{n,\theta}(t)\, \mathrm{d}t = 0,$$

which can be rewritten as

$$\int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(t) \left[\mathrm{d}N_n(t) - \lambda_{n,\theta}(t)\mathrm{d}t\right] = 0,$$

or

$$\int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(t)\, \mathrm{d}w_{n,\theta}(t) = 0.$$

Now we need regularity assumptions on $\lambda_{n,\theta}$ as a function of θ, namely:

- (b_1) differentiation with respect to θ, integration with respect to $\mathrm{d}N_n(t)$, and $\mathrm{d}t$ can be interchanged,
- (b_2) the ratio $[\dot{\lambda}/\lambda]_{n,\theta}(t)$ is well-defined on $\{t : \lambda_{n,\theta}(t) > 0\}$, and can be defined as a constant on $\{t : \lambda_{n,\theta}(t) = 0\}$,
- (b_3) we have

$$\frac{1}{\sqrt{n}} \int_0^T \frac{\partial}{\partial\theta} [\dot{\lambda}/\lambda]_{n,\theta}(t)\, \mathrm{d}w_{n,\theta}(t) = \int_0^T \frac{\partial}{\partial\theta} [\dot{\lambda}/\lambda]_{n,\theta}(t) \left[\frac{\mathrm{d}N_n(t)}{n} - \frac{\lambda_{n,\theta}(t)}{n}\mathrm{d}t\right] = o_P(1),$$

which is a form of Law of Large Numbers for the process N_n and the vector function $[\dot{\lambda}/\lambda]_{n,\theta}(t)$.

To obtain a suitable asymptotic expansion for the MLE $\widehat{\theta}$, we use the Taylor expansion once again, and rewrite the maximum likelihood equation

$$\int_0^T [\dot\lambda/\lambda]_{n,\widehat{\theta}}(t)\,\mathrm{d}w_{n,\widehat{\theta}}(t) = 0$$

as follows:

$$0 = \int_0^T [\dot\lambda/\lambda]_{n,\theta}(t)\,\mathrm{d}w_{n,\theta}(t) + \frac{1}{\sqrt{n}}\int_0^T \frac{\partial}{\partial\theta}[\dot\lambda/\lambda]_{n,\theta}(t)\,\mathrm{d}w_{n,\theta}(t)\sqrt{n}(\widehat{\theta}-\theta)$$
$$- \frac{1}{n}\int_0^T [\dot\lambda/\lambda]_{n,\theta}(t)[\dot\lambda/\lambda]^{\mathsf{T}}_{n,\theta}(t)\lambda_{n,\theta}(t)\,\mathrm{d}t\sqrt{n}(\widehat{\theta}-\theta) + o_P(1).$$

Here, the last $o_P(1)$ is our regularity assumption (b_4) and we will also use

(b_5) the random matrix

$$R_{n,\theta} = \frac{1}{n}\int_0^T [\dot\lambda/\lambda]_{n,\theta}(t)\,[\dot\lambda/\lambda]^{\mathsf{T}}_{n,\theta}(t)\,\lambda_{n,\theta}(t)\,\mathrm{d}t$$

is well-defined and non-degenerate for all $\theta \in \Theta$ and all n sufficiently large. Moreover, there is a non-degenerate matrix R_θ such that $R_{n,\theta} \to R_\theta$ as $n \to \infty$.

This leads to the desired asymptotic representation

$$\sqrt{n}(\widehat{\theta}-\theta) = R^{-1}_{n,\theta}\int_0^T [\dot\lambda/\lambda]_{n,\theta}(t)\,\mathrm{d}w_{n,\theta}(t) + o_P(1), \tag{11}$$

which is an analog of (8) for point processes.

Now we turn to the difference between $w_{n,\theta}$ and $w_{n,\widehat{\theta}}$. Using Taylor's expansion again, we obtain

$$w_{n,\widehat{\theta}}(t) = w_{n,\theta}(t) - \frac{1}{n}\int_0^t [\dot\lambda/\lambda]^{\mathsf{T}}_{n,\theta}(s)\lambda_{n,\theta}(s)\,\mathrm{d}s\sqrt{n}(\widehat{\theta}-\theta) + o_P(1),$$

or, by virtue of (11),

$$w_{n,\widehat{\theta}}(t) = w_{n,\theta}(t) - \frac{1}{n}\int_0^t [\dot\lambda/\lambda]^{\mathsf{T}}_{n,\theta}(s)\lambda_{n,\theta}(s)\,\mathrm{d}s\,R^{-1}_{n,\theta}\int_0^T [\dot\lambda/\lambda]_{n,\theta}(t)\,\mathrm{d}w_{n,\theta}(t) + o_P(1),$$

an expression analogous to (9). The main part on the right-hand side is a linear transformation of $w_{n,\theta}$. Moreover, defining

$$\Lambda_{n,\theta}(t) = \int_0^t \lambda_{n,\theta}(s)\,\mathrm{d}s,$$

we have the following analog of Proposition 2.1.

Proposition 2.2. *The linear operator* Π_n *defined by*

$$\Pi_n\gamma(t) = \gamma(t) - \frac{1}{n}\int_0^t [\dot\lambda/\lambda]^{\mathsf{T}}_{n,\theta}(s)\,\mathrm{d}\Lambda_{n,\theta}(s)\,R^{-1}_{n,\theta}\int_0^T [\dot\lambda/\lambda]_{n,\theta}(t)\,\mathrm{d}\gamma(t)$$

is an orthogonal projector, i.e., it satisfies the conditions

(i) $\Pi_n\Pi_n\gamma(t) = \Pi_n\gamma(t)$,

(ii) $\Pi_n\gamma(t) \equiv 0 \Leftrightarrow \dfrac{\mathrm{d}\gamma}{\mathrm{d}\Lambda_{n,\theta}}(t) = c^{\mathsf{T}}[\dot\lambda/\lambda]_{n,\theta}(t)$ *for some* $c \in \mathbb{R}^m$,

(iii) $\int_0^T [\dot\lambda/\lambda]_{n,\theta}(s)\,\mathrm{d}\Pi_n\gamma(s) = 0.$

Therefore, substitution of ML estimation in place of the 'true' value of the parameter again is asymptotically equivalent to taking a projection of the martingale $w_{n,\theta}$. Heuristically speaking, this makes the process $w_{n,\hat\theta}$ stochastically 'smaller', less volatile, less 'noisy', and makes the tests based on $w_{n,\hat\theta}$ better, more powerful, as it is the case for the process $v_{n,\hat\theta}$ (see, e.g., [4]).

3. Function Parametric Versions and Unitary Operators

We realised that the empirical process $v_{n,\widehat\theta}$ with estimated parameter $\widehat\theta$ is essentially a projection of the corresponding empirical process $v_{n,\theta}$ (see (9)). However, for different parametric families we have different score functions $\dot f/f$, and therefore different projections. Even within the same parametric family, different values of the parameter θ again lead to different $\dot f/f$, and $v_{n,\widehat\theta}$ will have different limit behavior. Consequently, the limiting distribution of any given test statistic $T(v_{n,\widehat\theta})$ will be different in any new testing problem.

In the goodness of fit problems, the test statistics T, as functionals of $v_{n,\widehat\theta}$, are non-linear and their limiting distributions are difficult to calculate, so that numerical methods have to be used. The theory becomes fragmented. Our eventual goal is to unify the theory again. We will see that what looks like many similar but different problems actually is one single problem, which requires the calculation of limiting distributions of test statistics $T(v_{n,\widehat\theta})$, for many "similar" $v_{n,\widehat\theta}$, only once. The same is true for testing parametric hypotheses about (random) intensities of point processes.

The main idea behind the methods we are going to employ consists in building a unitary operator, or rotation, of one testing problem into another, thus creating surprisingly broad families of equivalent testing problems. However, it may look awkward to "rotate" empirical processes. We introduce now a form of empirical processes which will create a natural setting to apply unitary operators.

Let v_F denote an F-Brownian bridge, i.e., a Gaussian process with mean zero and covariance function

$$\mathbb{E}[v_F(t)v_F(t')] = \min\{F(t), F(t')\} - F(t)F(t').$$

This process is the limit in distribution of the empirical process v_n in (2). It is convenient to recall that if w_F is an F-Brownian motion, i.e., a Gaussian process with mean zero and covariance

$$\mathbb{E}[w_F(t)w_F(t')] = \min\{F(t), F(t')\},$$

then one well-known connection between w_F and v_F is

$$v_F(t) \stackrel{d}{=} w_F(t) - F(t)w_F(\infty), \tag{12}$$

and if we agree to choose F supported on the unit interval $[0,1]$ so that $F(0)=0$ and $F(1)=1$, then we can write $w_F(1)$ instead of $w_F(\infty)$.

If $\{F_\theta : \theta \in \Theta\}$ is a regular parametric family, then from (9) it is possible to derive that the Gaussian process

$$\widehat v_F(t) = v_F(t) - \int_0^t [\dot f/f]_\theta^{\mathsf T}(s)\,\mathrm{d}F_\theta(s) R_\theta^{-1} \int [\dot f/f]_\theta(x)\,\mathrm{d}v_F(x)$$

is the limit in distribution of the parametric empirical process $v_{n,\widehat\theta}$ with θ denoting the true parameter value. Now let us rewrite $\widehat v_F$ in what is called function parametric form.

Suppose, as before, that θ is an m-dimensional parameter. Then $[\dot f/f]_\theta(\cdot)$ is an m-dimensional vector function with linearly independent components. Let us introduce the notation

$$q_\theta(\cdot) = R_\theta^{-1/2}[\dot f/f]_\theta(\cdot)$$

for the ortho-normalised form of the score function $[\dot{f}/f]_\theta$. Indeed, q_θ is a vector function with orthogonal and normalized components in the space $L_2(F)$, since

$$\int q_\theta\, q_\theta^\mathsf{T}\, \mathrm{d}F_\theta = R_\theta^{-1/2} \int [\dot{f}/f]_\theta [\dot{f}/f]_\theta^\mathsf{T}\, \mathrm{d}F_\theta\, R_\theta^{-1/2} = I.$$

Below, we will drop the subscript θ in F_θ when F_θ is used as a subscript.

Now, given a function $\varphi \in L_2(F_\theta)$, let us introduce what is called *function parametric* version of our processes. Consider the integral

$$\widehat{v}_F(\varphi) := \int \varphi(x)\, \mathrm{d}\widehat{v}_F(x) = \int \varphi(x)\, \mathrm{d}v_F(x) - \int \varphi(x) q_\theta^\mathsf{T}(x)\, \mathrm{d}F_\theta(x) \int q_\theta(y)\, \mathrm{d}v_F(y). \tag{13}$$

This is a Wiener stochastic integral, well-defined on $L_2(F_\theta)$. It is clear that

$$v_F(\varphi) = \int \varphi(x)\, \mathrm{d}v_F(x)$$

is linear in φ, that is, if $\varphi_1, \varphi_2 \in L_2(F_\theta)$ and $\alpha_1, \alpha_2 \in \mathbb{R}$, then

$$v_F(\alpha_1\varphi_1 + \alpha_2\varphi_2) = \alpha_1 v_F(\varphi_1) + \alpha_2 v_F(\varphi_2).$$

This implies that $\widehat{v}_F(\varphi)$ is also linear in φ, and (13) can be rewritten as

$$\widehat{v}_F(\varphi) = v_F(\varphi) - \langle \varphi, q_\theta\rangle_F^\mathsf{T}\, v_F(q_\theta) = v_F\big(\varphi - \langle \varphi, q_\theta\rangle_F^\mathsf{T}\, q_\theta\big), \tag{14}$$

where $\langle \varphi, q_\theta\rangle_F$ denotes the vector of inner products in $L_2(F)$ of φ and the components of q_θ:

$$\langle \varphi, q_\theta\rangle_F := \int \varphi(x) q_\theta(x)\, \mathrm{d}F_\theta(x).$$

Thus we have the following reformulation of Proposition 2.1. To formulate its (ii) part, we extend the m-dimensional vector of score functions q_θ to the $(m+1)$-dimensional vector having q_0 as the first coordinate:

$$q = \begin{pmatrix} q_0 \\ q_\theta \end{pmatrix} = \begin{pmatrix} q_0 \\ q_1 \\ \vdots \\ q_m \end{pmatrix}.$$

Here q_0 is the function, which is constant and equals 1 for all x. Note that the extended q will still be a vector with orthonormal coordinates, because our regularity assumption (a_3) implies

$$\int [\dot{f}/f]_\theta(x)\, \mathrm{d}F_\theta(x) = \langle [\dot{f}/f]_\theta, q_0\rangle_F = 0, \quad \text{or} \quad \langle q_\theta, q_0\rangle_F = 0,$$

Proposition 3.1 ([4]). *For the limiting processes of $v_{n,\theta}$ and $v_{n,\hat\theta}$ we have*

(i)

$$\widehat{v}_F(\varphi) = v_F\big(\varphi - \langle \varphi, q_\theta\rangle_F^\mathsf{T}\, q_\theta\big),$$

which represents $\widehat{v}_F$ as a projection of function parametric Brownian bridge v_F, and

(ii)

$$\widehat{v}_F(\varphi) = w_F\big(\varphi - \langle \varphi, q\rangle_F^\mathsf{T}\, q\big), \tag{15}$$

which represents $\widehat{v}_F$ as a projection of function parametric Brownian motion w_F.

Proof. To see that (i) is true, note that the form of the argument of v_F in (i) follows from (14), and since q_θ is orthonormal, this is the orthogonal projection of φ, parallel to q_θ,

$$\pi\varphi = \varphi - \langle \varphi, q_\theta\rangle_F^\mathsf{T}\, q_\theta.$$

To see that (ii) is true, we use the projection structure behind the function parametric form of Brownian bridge $v_F(\varphi)$, $\varphi \in L_2(F)$, as well. According to (12),

$$\begin{aligned} v_F(\varphi) = \int \varphi(x)\, \mathrm{d}v_F(x) &= \int \varphi(x)\, \mathrm{d}w_F(x) - \int \varphi(x)\, \mathrm{d}F_\theta(x) w_F(\infty) \\ &= w_F(\varphi) - \langle \varphi, q_0\rangle_F\, w_F(q_0), \end{aligned} \tag{16}$$

where we recall that q_0 denotes the function, identically equal to 1, $q_0(\cdot) \equiv 1$. Again, since $w_F(\varphi)$ is linear in φ, we can rewrite the last expression as

$$v_F(\varphi) = w_F\big(\varphi - \langle\varphi, q_0\rangle_F\, q_0\big), \tag{17}$$

and the argument of w_F here is the orthogonal projection of φ, parallel to q_0. Now we substitute (16) into (14). This will represent $\widehat{v}_F$ as a projection of w_F:

$$\widehat{v}_F(\varphi) = w_F(\varphi) - \langle\varphi, q_0\rangle_F\, w_F(q_0) - \langle\varphi, q_\theta\rangle_F^{\mathsf{T}}\, w_F(q_\theta). \tag{18}$$

We replaced the term $v_F(q_\theta)$ in (14) by the term $w_F(q_\theta)$, and we can indeed do this: as it follows from (17),

$$v_F(q_\theta) = w_F(q_\theta - \langle q_\theta, q_0\rangle_F\, q_0),$$

while $\langle q_\theta, q_0\rangle_F = 0$, and therefore $v_F(q_\theta) = w_F(q_\theta)$. □

Now let us consider two different regular parametric families $\{F_\theta : \theta \in \Theta\}$ and $\{G_\theta : \theta \in \Theta\}$, with two different score functions q and r (extended and orthonormal, as above). We assume, however, that the vector functions q and r are of equal dimensions. In this notation we use the same letters θ and Θ, but we do not mean to say that these are in any sense the "same" parameters, say shift and scale parameters in both cases. They may be parameters of entirely different nature in these two different families. They only should be of the same dimension and they should lead to linearly independent, and therefore eventually orthonormal, score functions.

Consider two limiting Gaussian processes

$$\widehat{v}_F(\varphi) = w_F\Big(\varphi - \sum_{i=0}^{m}\langle\varphi, q_i\rangle_F\, q_i\Big), \quad \varphi \in L_2(F_\theta), \quad \langle q_i, q_j\rangle_F = \delta_{ij},$$

$$and \widehat{v}_G(\psi) = w_G\Big(\psi - \sum_{i=0}^{m}\langle\psi, r_i\rangle_G\, r_i\Big), \quad \psi \in L_2(G_\theta), \quad \langle r_i, r_j\rangle_G = \delta_{ij}.$$

What we will show now is that, under the additional assumption of equivalence (mutual absolute continuity) between the distributions F_θ and G_θ, we can map $\widehat{v}_F$ to $\widehat{v}_G$ in a one-to-one way, and the mapping has a practically convenient form. More specifically, we will construct a unitary operator $K = K_{q,r}$ mapping $L_2(G_\theta)$ onto $L_2(F_\theta)$, so that

$$\widehat{v}_F(K\psi) \stackrel{d}{=} \widehat{v}_G(\psi), \quad \psi \in L_2(G_\theta).$$

Because this K is a unitary operator, we will also have

$$\widehat{v}_G(K^{-1}\varphi) \stackrel{d}{=} \widehat{v}_F(\varphi), \quad \varphi \in L_2(F_\theta).$$

Allowing ourselves some freedom of speech, we will say that $\widehat{v}_F$ is "rotated" into $\widehat{v}_G$.

For the sake of better transparency, let us construct K in a sequence of three problems. In the first, or "zero problem," let us map w_F into w_G isometrically. Here, the dependence on the parameter θ will play no role, and it can be skipped from the notations. Consider the square root of the density of G with respect to F:

$$\ell(x) = \left(\frac{\mathrm{d}G(x)}{\mathrm{d}F(x)}\right)^{1/2}.$$

Since F and G are equivalent measures, we have

$$\ell \in L_2(F), \text{ with } \|\ell\|_F^2 = \int \ell^2(x)\,\mathrm{d}F(x) = 1,$$

$$1/\ell \in L_2(G), \text{ with } \|1/\ell\|_G^2 = \int \frac{1}{\ell^2(x)}\,\mathrm{d}G(x) = 1.$$

Let $\ell\psi(\cdot) = \ell(\cdot)\psi(\cdot)$ denote the operator of multiplication by the function ℓ, acting on functions $\psi \in L_2(G)$.

Lemma 3.2. *The operator ℓ is an isometry from $L_2(G)$ to $L_2(F)$, and we have $w_F(\ell\psi) = w_G(\psi)$.*

Proof. Indeed,

$$\int \big[l(x)\psi(x)\big]^2 \,\mathrm{d}F(x) = \int \psi^2(x)\,\mathrm{d}G(x),$$

and therefore,

$$\mathbb{E}w_F^2(\ell\psi) = \|\ell\psi\|_F^2 = \|\psi\|_G^2 = \mathbb{E}w_G^2(\psi).$$

□

The next problem is to rotate the Brownian bridge v_F into the Brownian bridge v_G. Now we have one-dimensional functions $q_0(\cdot) = 1$ and $r_0(\cdot) = 1$. Note that the latter function is identically equal to 1 in $L_2(G)$, but its image under the operator ℓ will not be identically equal to 1 in $L_2(F)$. We know that

$$v_G(\psi) = \begin{cases} w_G(\psi) & \text{if } \psi \perp r_0 \\ 0 & \text{if } \psi = r_0 \end{cases} \quad \text{and} \quad v_F(\varphi) = \begin{cases} w_F(\varphi) & \text{if } \varphi \perp q_0 \\ 0 & \text{if } \varphi = q_0 \end{cases}.$$

Therefore, in order to rotate v_F into v_G we need a unitary operator from $L_2(G)$ to $L_2(F)$ which will map the linear subspace $\mathcal{L}_G(r_0) = \{cr_0(\cdot) : c \in \mathbb{R}\}$ into the linear subspace $\mathcal{L}_F(q_0)$, and which, therefore, will map the orthogonal complement of $\mathcal{L}_G(r_0)$ in $L_2(G)$ (denoted by $\mathcal{L}_{G,\perp}(r_0)$) into the orthogonal complement of $\mathcal{L}_F(q_0)$ in $L_2(F)$ (denoted by $\mathcal{L}_{F,\perp}(q_0)$). In order to do this, consider first the operator $K_{a,b}$ mapping $L_2(F)$ to $L_2(F)$ via

$$K_{a,b}(\cdot) = I - 2\frac{\langle a-b, \cdot\rangle_F}{\|a-b\|_F^2}(a-b), \tag{19}$$

where $a, b \in L_2(F)$ are two fixed functions of unit norm, and I is the identity operator. It is easy to check that this operator has the following properties:

1. $K_{a,b}$ is unitary, i.e., $\|K_{a,b}\varphi\|_F = \|\varphi\|_F$,
2. $K_{a,b} = K_{a,b}^{-1}$, i.e., $K_{a,b}K_{a,b} = I$,
3. $K_{a,b}$ is self-adjoint, i.e., $\langle K_{a,b}\varphi, \gamma\rangle_F = \langle \varphi, K_{a,b}\gamma\rangle_F$,
4. $K_{a,b}\, a = b$ and $K_{a,b}\, b = a$.

Now let us choose $a = q_0$ and $b = \ell r_0$, and consider the process $v_F(K_{q_0,\ell r_0}\ell\psi)$ for $\psi \in L_2(G)$. We claim that this process has the same distribution as $v_G(\psi)$. Together with the statement on Brownian motions, above, we obtain

Proposition 3.3. *If distributions F and G are equivalent, then*

$$w_F(\ell\psi) = w_G(\psi) \quad \textit{and} \quad v_F(K_{q_0,\ell r_0}\ell\psi) \stackrel{d}{=} v_G(\psi).$$

Proof. Indeed, if $\psi = r_0$, then $K_{q_0,\ell r_0}\ell\psi = q_0$, and so

$$v_F(K_{q_0,\ell r_0}\ell r_0) = v_F(q_0) = 0 = v_G(r_0).$$

On the other hand, if $\psi \perp r_0$, then $\ell\psi \perp \ell r_0$, and therefore $K_{q_0,\ell r_0}\ell\psi \perp q_0$. From the equality $v_F(\varphi) = w_F(\varphi)$ when $\varphi \perp q_0$, it follows that

$$v_F(K_{q_0,\ell r_0}\ell\psi) = w_F(K_{q_0,\ell r_0}\ell\psi),$$

and the variance of the right-hand side for any such ψ is

$$\mathbb{E}w_F^2(K_{q_0,\ell r_0}\ell\psi) = \langle K_{q_0,\ell r_0}\ell\psi, K_{q_0,\ell r_0}\ell\psi\rangle_F = \langle \ell\psi, \ell\psi\rangle_F = \langle\psi,\psi\rangle_G = \mathbb{E}w_G^2(\psi).$$

Therefore, indeed, v_F was "rotated" into v_G. □

We are now ready to tackle the third problem in our sequence, the rotation of $\widehat{v}_F$ into $\widehat{v}_G$. Let us first consider the case of regular parametric families with one-dimensional parameter, leading to two-dimensional extended score functions $(q_0, q_1)^\mathsf{T}$ for one family and $(r_0, r_1)^\mathsf{T}$ for the other. Now we have

$$\widehat{v}_F(\varphi) = w_F\big(\varphi - \langle\varphi, q_0\rangle_F q_0 - \langle\varphi, q_1\rangle_F q_1\big), \quad \varphi \in L_2(F_\theta),$$
$$\widehat{v}_G(\psi) = w_G\big(\psi - \langle\psi, r_0\rangle_G r_0 - \langle\psi, r_1\rangle_G r_1\big), \quad \psi \in L_2(G_\theta).$$

Consider the operator $K_{q_0,\ell r_0}$ used for the previous problem above and apply it to ℓr_1, thus creating

$$\widetilde{\ell r_1} := K_{q_0,\ell r_0}\ell r_1.$$

The operator $K_{q_0,\ell r_0}$ correctly rotates the function ℓr_0 into q_0, but it does not necessarily rotate ℓr_1 into q_1, but only into $\widetilde{\ell r_1}$. Since it is a unitary operator, it preserves angles, and therefore $\widetilde{\ell r_1} \perp q_0$. Now we can rotate $\widetilde{\ell r_1}$ further into q_1 using the operator $K_{q_1,\widetilde{\ell r_1}}$. Note that this operator leaves all functions orthogonal to q_1 and $\widetilde{\ell r_1}$ unchanged, so it will leave q_0 unchanged. Now consider the operator $K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0}$. We have

Proposition 3.4. *If distributions F_θ and G_θ are equivalent and if $q_0, q_1 \in L_2(F_\theta)$ are orthonormal as well as $r_0, r_1 \in L_2(G_\theta)$, then*

$$\widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi) \stackrel{d}{=} \widehat{v}_G(\psi), \quad \psi \in L_2(G_\theta).$$

Proof. Indeed, if $\psi = r_0$, then

$$\widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell r_0) = \widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} q_0) = \widehat{v}_F(q_0) = 0 = \widehat{v}_G(r_0),$$

and similarly, if $\psi = r_1$,

$$\widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell r_1) = \widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} \widetilde{\ell r_1}) = \widehat{v}_F(q_1) = 0 = \widehat{v}_G(r_1).$$

Moreover, $K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0}$ is a product of unitary operators, hence it is itself a unitary operator. As we have just seen, it maps ℓr_0 into q_0 and ℓr_1 into q_1. Therefore, it will map $\ell\psi$, for any $\psi \perp r_0, r_1$, into a function, orthogonal to q_0 and q_1. It follows that for any such ψ,

$$\widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi) = w_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi),$$

and the variance of the right-hand side is

$$\langle K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi, K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi \rangle_F = \langle \ell\psi, \ell\psi \rangle_F = \langle \psi, \psi \rangle_G.$$

This means that for $\psi \perp r_0, r_1$,

$$\widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi) \stackrel{d}{=} w_G(\psi) = \widehat{v}_G(\psi),$$

as claimed. □

Finally, for parametric families with an m-dimensional parameter, we use induction. Given $j \in \{0, 1, \ldots, m-1\}$, suppose we have a unitary operator $U_{q,\ell r}(j)$ that maps ℓr_i to q_i for $0 \le i \le j$. For example, we have constructed above $U_{q,\ell r}(0) = K_{q_0,\ell r_0}$ and $U_{q,\ell r}(1) = K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0}$. Now define the function

$$\widetilde{\ell r_{j+1}} := U_{q,\ell r}(j) \ell r_{j+1}$$

and introduce

$$U_{q,\ell r}(j+1) = K_{q_{j+1},\widetilde{\ell r_{j+1}}} U_{q,\ell r}(j).$$

Then $U_{q,\ell r}(j+1)$ is a unitary operator that maps ℓr_i to q_i for $0 \le i \le j+1$. Continuing in this fashion, we see that $U_{q,\ell r}(m)$ is a unitary operator that maps ℓr_i to q_i for $0 \le i \le m$. Therefore we obtain our final statement.

Proposition 3.5 ([7])**.** *If distributions F_θ and G_θ are equivalent and if q and r are orthonormal systems of $m+1$ functions* (*as described above*) *from $L_2(F_\theta)$ and $L_2(G_\theta)$, respectively, then*

$$\widehat{v}_F(U_{q,\ell r}(m) \ell\psi) \stackrel{d}{=} \widehat{v}_G(\psi). \tag{20}$$

It can be proved by an argument analogous to the previous case.

4. The Case of Point Processes. Unitary Transformations Again

In this section we describe similarities in rotation between the situation with parametric families of distribution and parametric models for intensities of point process. Let us consider a sequence of point processes N_n with (random) intensity functions $\lambda_{n,\theta}$ and compensators

$$\Lambda_{n,\theta}(t) = \int_0^t \lambda_{n,\theta}(s)\,\mathrm{d}s.$$

One of the key facts for us is that, if $\Lambda_{n,\theta}(t)/n$ converges to a deterministic function, say $B(t)$, as $n \to \infty$, then the normalized martingale

$$w_{n,\theta} = \frac{1}{\sqrt{n}}[N_n(t) - \Lambda_{n,\theta}(t)]$$

converges to the Brownian motion (see, e.g., [2,3]), while the same process with the estimated parameter $w_{n,\widehat{\theta}}$, can be approximated by a projection of $w_{n,\theta}$:

$$w_{n,\widehat{\theta}}(t) = w_{n,\theta}(t) - \frac{1}{n}\int_0^t [\dot{\lambda}/\lambda]^{\mathsf{T}}_{n,\theta}(s)\lambda_{n,\theta}(s)\,\mathrm{d}s\,R^{-1}_{n,\theta}\int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(s)\,\mathrm{d}w_{n,\theta}(s) + o_P(1).$$

The key regularity assumptions ate such that

$$[\dot{\lambda}/\lambda]_{n,\theta}(t) \to \alpha(t), \quad \frac{1}{n}\lambda_{n,\theta}(t) \to \beta(t), \quad n \to \infty, \tag{21}$$

for some deterministic functions α and β. As a consequence, we expect that

$$w_{n,\theta} \xrightarrow{d} w_B, \quad w_{n,\widehat{\theta}} \xrightarrow{d} \widehat{w}_B,$$

with $B(t) = \int_0^t \beta(s)\,\mathrm{d}s$, and

$$\widehat{w}_B(t) = w_B(t) - \int_0^t \alpha^{\mathsf{T}}(s)\beta(s)\,\mathrm{d}s\,R_\theta^{-1}\int_0^T \alpha(s)\,\mathrm{d}w_B(s). \tag{22}$$

If we have another parametric model with the same regularity assumptions, then we will end up with another Brownian motion $w_{\widetilde{B}}(t)$, in time $\widetilde{B}$, and another projection $\widehat{w}_{\widetilde{B}}(t)$, parallel to a different score function $\widetilde{\alpha}$. If the parameters in the two cases are of the same dimension, then it again becomes possible to "rotate" $\widehat{w}_B(t)$ into $\widehat{w}_{\widetilde{B}}(t)$, and back if we wish. The form of the unitary operator needed for this task will be exactly the same as the one we have obtained for the case of i.i.d. samples.

There is, however, one difference that for the case of empirical processes the first coordinate of the score function always is the function $q_0(\cdot) = 1$, while this is not the case for the point processes: the first coordinate of the vector α may be any function, square-integrable with respect to the limiting "time" B.

As in the i.i.d. case, there is a question how to choose the "standard" problem in which to rotate the other problems. Indeed, one has here multiplicity of choices. As a simple choice, we suggest below to use Poisson processes with a variable intensity. At the first glance this looks somewhat strange, because then the function $\lambda_{n,\theta}$ will be a deterministic function from the very beginning and the regularity assumptions (21) will be easily satisfied. This is surprisingly simple, but, on the other hand, it is convenient.

Specifically, let us start with the space $L_2(\omega)$ of square-integrable functions on $[0,T], T \le \infty$, with a weight function ω and choose orthonormal functions $p_0, \dots, p_{m-1}$ from $L_2(\omega)$, i.e., such that

$$\int_0^T p_j(s)p_k(s)\omega(s)ds = \delta_{j,k}.$$

One example can be given by the Laguerre polynomials with $\omega(t) = e^{-t}$, $t \geq 0$. If we agree to consider a finite time horizon $T < \infty$, then it would be natural to use the constant weight function ω. Define now the intensity function

$$\mu_{n,\theta}(t) = n \exp\left(\sum_{j=0}^{m-1} \theta_j p_j(t)\right)\omega(t), \quad 0 \leq t \leq T.$$

Another possibility is to choose

$$\mu_{n,\theta}(t) = n \exp\left(\sum_{j=0}^{m-1} \theta_j p_j(t)\omega^{1/2}(t)\right), \quad 0 \leq t \leq T. \tag{23}$$

In this latter case one can choose as a true "target" distribution the distribution of the time-homogeneous Poisson process with intensity n. This distribution is a part of the parametric family above with the vector $\theta = 0$. As the target parametric family we choose distributions of Poisson processes with parameter $\theta = (\theta_0, \theta_1, \ldots, \theta_{m-1})^{\mathsf{T}}$, which takes values in a small open neighbourhood of 0. We need an open neighbourhood such that differentiation with respect to θ will not meet with difficulties, and it suffices to have this neighbourhood small. For this neighbourhood, at $\theta^* = (0, 0, \ldots, 0)^{\mathsf{T}}$ and $t \in [0, T]$ we have

$$[\dot{\mu}/\mu]_{n,\theta^*}(t) = \Big(p_j(t)\omega^{1/2}(t)\Big)_{j=0}^{m-1},$$

while

$$\frac{1}{n}\mu_{n,\theta^*}(t) = 1,$$

and one can easily see the consequence of the assumption of orthonormality of the functions $(p_j)_{j=0}^{m-1}$: the matrix

$$R_{\theta^*} = \left[\int_0^T p_j(t)p_k(t)\omega(t)dt\right]_{j,k=0}^{m-1} = I.$$

Therefore the coordinates of $[\dot{\mu}/\mu]_{n,\theta^*}$ are already ortho-normal.

Now we can show, in more or less explicit form, the rotation of $w_{n,\widehat{\theta}}(t)$ into the process, $\tilde{w}_{n,\widehat{\theta}}(t)$, which would arise from our Poisson model above. As in Section 3, it is notationally convenient to introduce orthonormal version of the vector-functions $[\dot{\lambda}/\lambda]_{n,\theta}(t)$ and of these functions for Poisson process. For the intensity $\lambda_{n,\theta}(t)$ we could have done it already after Proposition 2.1. Namely, denote

$$q_{n,\theta}(t) = R_{n,\theta}^{-1/2}[\dot{\lambda}/\lambda]_{n,\theta}(t).$$

This is the vector-function with orthonormal coordinates in $L_2(\Lambda_{n,\theta}/n)$:

$$\frac{1}{n}\int q_{n,\theta}(t)q_{n,\theta}^{\mathsf{T}}(t)\lambda_{n,\theta}(t)\mathrm{d}t = I.$$

The limits of $[\dot{\lambda}/\lambda]_{n,\theta}(t)$ and $\frac{1}{n}\lambda_{n,\theta}(t)$ in (21) suggest the limiting form of this vector-function:

$$q(t) = R_\theta^{-1/2}\alpha(t) = (q_0(t), q_1(t), \ldots, q_{m-1}(t))^{\mathsf{T}}$$

with the orthonormality property

$$\int q(t)q(t)^{\mathsf{T}}\beta(t)\mathrm{d}t = I.$$

For our Poisson process we have already the vector-function $[\dot{\mu}/\mu]_{n,\theta^*}$, whose coordinates are orthonormal on $[0, T]$. It does not change with n.

Now the procedure will look literary the same as the rotation of Brownian bridges v_F and v_G. Adopting (23) as the target parametric model, denote

$$M_{n,\theta}(t) = \int_0^t \mu_{n,\theta}(s)ds, \quad \text{so that} \quad M_{n,\theta^*}(t) = nt.$$

Now choose a function ℓ (the Hellinger function) as

$$\ell_{n,\theta}(t) = \left(\frac{\mathrm{d}M_{n,\theta^*}}{\mathrm{d}\Lambda_{n,\theta}}(t)\right)^{1/2} = \left(\frac{\mu_{n,\theta^*}}{\lambda_{n,\theta}}(t)\right)^{1/2}$$

or

$$\ell_{n,\theta}(t) = \left(\frac{n}{\lambda_{n,\theta}(t)}\right)^{1/2}.$$

Thus, if $\psi \in L_2(M_{n,\theta^*}/n)$, then $\ell\psi \in L_2(\Lambda_{n,\theta}/n)$.

In limiting form, this expression becomes

$$\ell_\theta(t) = \left(\frac{1}{\beta(t)}\right)^{1/2},$$

and if $\psi \in L_2(M)$, then $\ell\psi \in L_2(B)$, where, as above, $B(t) = \int_0^t \beta(s)\mathrm{d}s$. Thus, for the possibility to rotate to the Poisson model we need to require that ℓ be well defined, that is, $\lambda_{n,\theta}(t) > 0$ and $\beta(t) > 0$ for all $t > 0$.

In the expression of $K_{a,b}$ (see (19)), we first will go straight to the limiting expressions, that is, we will prepare for the case of large n. If it happens that the result of our rotation behaves close to what is expected in the Poisson case, then we will fond out that the values of n, which we have used in our simulations, are "large enough". We can use expressions for finite n and compare the outcomes later.

Let us use $a = q_0(t)$ and $b = \ell p_0(t) = \ell(t)$. This leads to the transformation

$$\hat{w}_B(K_{q_0,\ell}\ell\psi) \stackrel{d}{=} \hat{w}_M(\psi),$$

and again, if we choose $\psi = p_0$, then $K_{q_0,\ell}\ell p_0 = q_0$, and therefore

$$\hat{w}_B(K_{q_0,\ell}\ell p_0) = \hat{w}_M(p_0) = 0,$$

which is, certainly, correct.

As the next step, we create the function $K_{q_0,\ell}\ell p_1 = \widetilde{\ell p_1}$ and use it to construct our next operator, $K_{q_1,\widetilde{\ell p_1}}$. The product $K_{q_1,\widetilde{\ell p_1}} K_{q_0,\ell}$ will map ℓp_0 and ℓp_1 into q_0 and q_1, respectively. Now we have, again,

$$\hat{w}_B(K_{q_1,\widetilde{\ell p_1}} K_{q_0,\ell}\psi) \stackrel{d}{=} \hat{w}_M(\psi),$$

and if the parameter θ is two-dimensional, then this equality is the final result. For a general dimension m we proceed as in the previous section: for

$$U_{q,p}(0) = K_{q_0,\ell} \quad \text{and} \quad U_{q,p}(1) = K_{q_1,\widetilde{\ell p_1}} U_{q,p}(0)$$

we continue with

$$\widetilde{\ell p_j} = U_{q,p}(j-1)\ell p_j$$

and then define

$$U_{q,p}(j) = K_{q_1,\widetilde{\ell p_j}} U_{q,p}(j-1).$$

It is the final operator $U_{q,p}(m)$ which will be needed in the sequel: a unitary operator which will map $p_0, \dots, p_{m-1}$ into $q_0, \dots, q_{m-1}$, and, therefore, will map all functions, orthogonal to $p_0, \dots, p_{m-1}$, to functions, orthogonal to $q_0, \dots, q_{m-1}$.

The situation should be clearer described in terms of subspaces. Decompose $L_2(M)$ into the subspace $\mathcal{L}(p)$ spanned by $p_0, \dots, p_{m-1}$ and its orthogonal complement $\mathcal{L}_\perp(p)$. Similarly, decompose $L_2(B)$ into the subspace $\mathcal{L}(q)$ spanned by $q_0, \dots, q_{m-1}$ and its orthogonal complement $\mathcal{L}_\perp(q)$. Then considering multiplication by ℓ as an isometry from $L_2(M)$ to $L_2(B)$,

$$\psi \in L_2(M) \implies \ell\psi \in L_2(B), \quad \|\psi\| = \|\ell\psi\|,$$

then this operator will map $\mathcal{L}_{\perp}(p)$ into some subspace in $L_2(B)$,

$$\ell\mathcal{L}_{\perp}(p) \subset L_2(B).$$

Then it is the operator $U_{q,p}(m)$, as an operator in $L_2(B)$, acting on $\ell\psi$, which will map $\ell\mathcal{L}_{\perp}(p)$ into $\mathcal{L}(q)$:

$$U_{q,p}(m)\ell p_j = q_j, \quad j = 1, \ldots, m.$$

It would be better to consider why the mapping of any testing problem for intensities of the point process, with only usual regularity assumptions, is basically the same problem always. This is true because in any model with these regularity assumptions we will end up with a Brownian motion in some time B – it will be specific for the model, and with a projection of this Brownian motion, parallel to the functions $q_0, \ldots, q_{m-1}$ – also specific for the model. While the method of unitary mapping remains applicable and the same.

We will need to apply the operator $U_{q,p}(m)$ to empirical processes with estimated parameters, that is, to the situation with finite n. Then we need to be sure that the transformed process $w_{n,\hat{\theta}}(U_{q,p}\ell\psi)$, $\psi \in \Psi$, where $\Psi \subset L_2(M)$ is a class of functions of our choice, does converge in distribution to the limiting process $\hat{w}_M(\psi)$, $\psi \in \Psi$. The most natural choice will be the set of indicator functions $\psi_t(s) = \mathbb{1}_{\{s \leq t\}})$ indexed by $t \geq 0$. It is obvious that as the function-parametric process, $\hat{w}_M(\psi_t)$ coincides with its point-parametric version $\hat{w}_M(t)$, and therefore the transformed empirical process $w_{n,\hat{\theta}}(U_{q,p}\ell\psi_t)$ should asymptotically behave as the point-parametric projected Brownian motion $\hat{w}_M(t)$.

It is very interesting to see what will be the graph of "rotated" ψ_t, that is, the graph of $U_{q,p}\ell\psi_t$. A sample of three graphs is shown in Figures 5.1 and 5.2 in the case of a point process model described in Example A of the next section. There the parameter is two-dimensional and we wished to transform the process into the projected Poisson process described above. The graphs have been calculated by S. Umut Can.

5. Some Specific Examples

Before we turn to specific examples, let us have a look on the expression of the limiting process (22) in the situation when the parameter of the intensity $\lambda_{n,\theta}$ of the point process is one-dimensional. In this situation α is a scalar function and $R_\theta = \int_0^T \alpha^2(s)\beta(s)\mathrm{d}s$ is a number. Then, considering the integral from α with respect to $\widehat{w}_B$:

$$\int_0^t \alpha(s)\mathrm{d}\widehat{w}_B(s) = \int_0^t \alpha(s)\mathrm{d}w_B(s) - \frac{\int_0^t \alpha^2(s)\beta(s)\,\mathrm{d}s}{\int_0^T \alpha^2(s)\beta(s)\mathrm{d}s} \int_0^T \alpha(s)\,\mathrm{d}w_B(s),$$

we see that the right-hand side is just the Brownian bridge in time

$$\tau = \frac{\int_0^t \alpha^2(s)\beta(s)\,\mathrm{d}s}{\int_0^T \alpha^2(s)\beta(s)\mathrm{d}s}, \quad t \in [0, T].$$

Therefore, all classical goodness of fit statistics from the Brownian bridge will be distribution free as statistics from the process $\int_0^t \alpha(s)\mathrm{d}\widehat{w}_B(s)$. The projection argument behind $\widehat{w}_B$ was used here, but to achieve distribution freeness no "rotation" was necessary. Full details are given in [9].

Example A. Consider a sequence of point processes $N_n(t)$ with compensated form

$$N_n(t) - \int_0^t c_\theta(t')[n - N_n(t')]dt';$$

in other words, the difference above is a martingale. Here we choose c_θ as the failure rate of Weibull distribution

$$c_\theta(t) = \frac{f_\theta(t)}{1 - F_\theta(t)} = \frac{\theta_1}{\theta_0}\left(\frac{t}{\theta_0}\right)^{\theta_1 - 1},$$

with parameters such that the corresponding Weibull's distribution behaves close to the distribution of life-times of, say, New Zealand population. These values are $\theta_0 = 86$ and $\theta_1 = 9$.

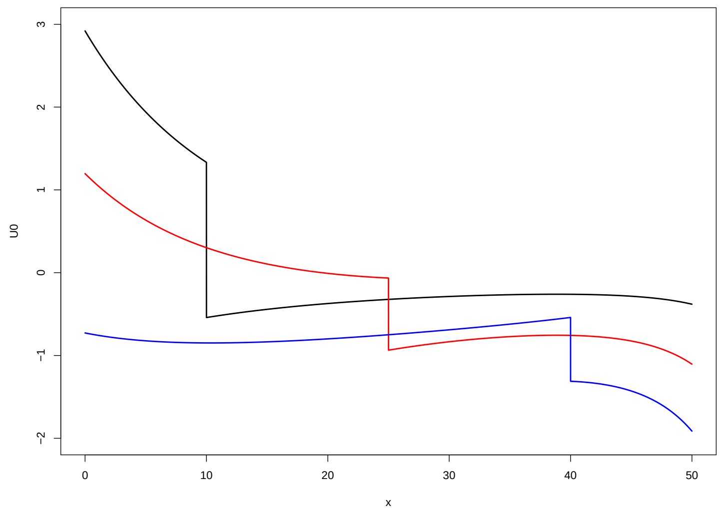


FIGURE 5.1. These are images of indicator functions $\mathbb{1}_{\{s\leq t\}}$ for $t = 10, 25$ and 40 after first rotation by the operator $K_{q_0,\ell p_0}\ell$.

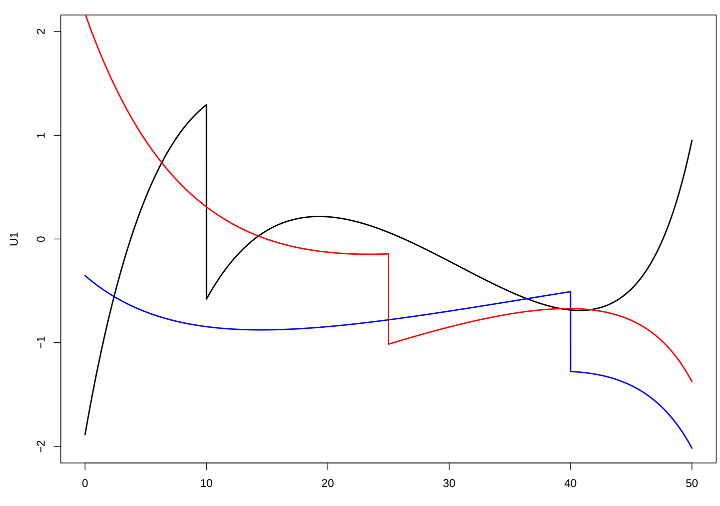


FIGURE 5.2. These are images of indicator functions $\mathbb{1}_{\{s\leq t\}}$ for $t = 10, 25$ and 40 after two rotations, i.e., by the operator $U_{q,p}\ell$. Who would think that if you integrate these three functions with respect to $dw_{\hat{\theta},n}(s)$ the resulting three integrals will asymptotically jointly behave as $\hat{w}_M(t), t = 10, 25,$ and 40?

We know that our process is, actually, a binomial process based on n i.i.d. observations from the distribution with the failure rate c_θ, i.e., from Weibull's distribution. If we would center N_n by $nF_\theta(t)$ and normalize by $\sqrt{n}$, we would obtain an empirical process, of which the limiting process will be the F_θ-Brownian bridge. Centered as in the above display, and again normalized by $\sqrt{n}$, we obtain a basic martingale (cf. [1,5,10]), and its weak limit will be the F_θ-Brownian motion.

The vector-function $[\dot\lambda/\lambda]_{n,\theta}(t)$ is now two-dimensional,

$$\frac{\dot\lambda_{n,\theta}}{\lambda_{n,\theta}}(t) = \frac{\dot c_{n,\theta}}{c_{n,\theta}}(t) = \left(-\frac{\theta_1}{\theta_0}, \frac{1}{\theta_1} + \ln\frac{t}{\theta_0}\right)^{\mathsf{T}}.$$

The function $\lambda_{n,\theta}/n$ and its limit is

$$\frac{1}{n}\lambda_{n,\theta}(t) = c_\theta(t)\frac{n - N_n(t)}{n} \to \beta_{\theta_0,\theta_1}(t),$$

where

$$\beta_{\theta_0,\theta_1}(t) = \frac{\theta_1}{\theta_0}\left(\frac{t}{\theta_0}\right)^{\theta_1 - 1}\exp\left(-(\frac{t}{\theta_0})^{\theta_1}\right)$$

is density of the Weibull distribution. The distribution function itself, in the current parametrisation, is $F_\theta(t) = 1 - \exp\left(-(\frac{t}{\theta_0})^{\theta_1}\right)$.

The covariance matrix R in its limiting form becomes

$$R_\theta = \int \begin{bmatrix} (\theta_1/\theta_0)^2, & -(1+\theta_1\ln(t/\theta_0))/\theta_0 \\ -(1+\theta_1\ln(t/\theta_0))/\theta_0, & (1+\theta_1\ln(t/\theta_0))^2/\theta_1^2 \end{bmatrix} \beta_{\theta_0,\theta_1}(t)dt$$

or, changing the variable t to $\tau = t/\theta_0$ and separating the constant terms, we obtain a slightly simpler expression

$$R_\theta = \int \begin{bmatrix} (\theta_1/\theta_0)^2, & -(1+\theta_1\ln\tau)/\theta_0 \\ -(1+\theta_1\ln\tau)/\theta_0, & (1+\theta_1\ln\tau)^2/\theta_1^2 \end{bmatrix} \beta_{1,\theta_1}(\tau)d\tau.$$

We note, as a side remark, that the matrix under the integral sign is, certainly, degenerate for every t, but the matrix R_θ is non-degenerate, it is invertible.

Now consider the integral on the anti-diagonal of this matrix. Since

$$\frac{d}{d\vartheta}\vartheta t^{\vartheta-1} = (1+\vartheta\ln t)t^{\vartheta-1},$$

one can write

$$\int (1+\vartheta\ln t)t^{\vartheta-1}\theta_1\exp\left(-t^{\theta_1}\right)dt = \frac{d}{d\vartheta}\int \vartheta t^{\vartheta-1}\theta_1\exp\left(-t^{\theta_1}\right)dt.$$

In the last integral we change the variable $t^{\theta_1} = z$ so that $t = z^{1/\theta_1}, dt = (1/\theta_1)z^{1/\theta_1-1}$. This leads to

$$\begin{aligned}\frac{d}{d\vartheta}\int \vartheta t^{\vartheta-1}\theta_1\exp\left(-t^{\theta_1}\right)dt &= \frac{d}{d\vartheta}\int \vartheta z^{(\vartheta-1)/\theta_1}\exp\left(-z\right)z^{1/\theta_1-1}dz \\ &= \frac{d}{d\vartheta}\int \vartheta z^{(\vartheta/\theta_1-1)}\exp\left(-z\right)dz = \frac{d}{d\vartheta}\vartheta\Gamma(\frac{\vartheta}{\theta_1})\end{aligned}$$

which at $\vartheta = \theta_1$ becomes $\dot\Gamma(1)$. This implies that we know explicitly the elements of the matrix R_θ, except one integral

$$R_\theta = \begin{bmatrix} \left(\frac{\theta_1}{\theta_0}\right)^2, & -\frac{1}{\theta_0}\dot\Gamma(1) \\ -\frac{1}{\theta_0}\dot\Gamma(1), & \frac{1}{\theta_1^2}\int(1+\theta_1\ln\tau)^2\beta_{1,\theta_1}(\tau)d\tau \end{bmatrix}.$$

Example B. One real life situation where this process appears is, as we said, the analysis of life times in human populations. However, in general human populations the huge bulk of life times belongs to the interval of 50-100 years. For example, according to New Zealand life tables for 2012-14 for general populations 50 (years) is only 4%-point and 100 (years) is about 99%-point. Therefore, it makes sense to analyse the life times only after age of fifty. If X_i is a life time of an i-th individual, then we consider $X_{i,x_0} = \max(0, X_i - x_0)$ and the point process N_{n,x_0} based on these "over the threshold" values, and then we can choose x_0 equal to 50, or to any other value of interest. We can also assume that we know how many people of age over x_0 we have in the population under study. Thus, for

$$N_{n,x_0}(t) = \sum_{i=1}^{n} \mathbb{1}(X_{i,x_0} \le t),$$

we have the representation

$$N_{n,x_0}(t) - \int_0^t c_\theta(x_0 + t')[n - N_{n,x_0}(t')]dt',$$

where the difference is a martingale. The functions $[\dot\lambda/\lambda]_{n,\theta}(t)$ and $\lambda_{n,\theta}/n$ now take the form

$$\frac{\dot\lambda_{n,\theta}}{\lambda_{n,\theta}}(t) = \frac{\dot c_{n,\theta}}{c_{n,\theta}}(x_0 + t) = \left(-\frac{\theta_1}{\theta_0}, \frac{1}{\theta_1} + \ln(\frac{t + x_0}{\theta_0})\right)^{\mathsf{T}}$$

and

$$\frac{1}{n}\lambda_{n,\theta}(t) = c_\theta(x_0 + t)\frac{n - N_{n,x_0}(t)}{n} \to \frac{\beta_{\theta_0,\theta_1}(x_0 + t)}{1 - F_{\theta_0,\theta_1}(x_0)}.$$

The matrix R_θ will also change, but in an obvious way. It is more interesting to note that we will need, in applications, to consider life-times not exceeding some value x_1, say, $x_1 = 100$, so that $N_{n,x_0}(t)$ will be stopped at some duration $x_1 - x_0$, equal, say, to 50 years (of life over age 50).

Example C. Now consider the same situation, but with $n - N_n(t)$ replaced by the process of "those at risk" (see, e.g., [1]). More specifically, consider a sequence of pairs $(X_i, C_i)_{i=1}^n$, where X_i is the survival time of i-th individual, and C_i is a censoring random variable of this survival time. Our main interest is in these survival times, however, one can only observe $\tilde X_i = \min(X_i, C_i)$ together with the indicator function $\delta_i = \mathbb{I}(X_i = \tilde X_i) = \mathbb{I}(X_i < C_i)$. The point process of interest is given as

$$N_n^c(t) = \sum_{i=1}^{n} \mathbb{1}(\tilde X_i \le t)\delta_i$$

which counts the number of "genuine" survival times observed no later than t. Another point process, of those at risk at time t is given as

$$Y_n(t) = \sum_{i=1}^{n} \mathbb{1}(\tilde X_i \ge t).$$

With the help of this process, the process N_n^c can be compensated to the martingale as follows:

$$N_n^c(t) - \int_0^t c_\theta(t')Y_n(t')dt', \tag{24}$$

and the difference is a martingale (see [1]). Here c_θ, as in Example A, is the failure rate (or the force of mortality in demographic applications) of the hypothetical distribution F_θ, depending on parameter θ.

If one is interested in computer simulation of N_n^c, one should somehow choose not only parametric family F_θ, of interests for practitioner, but also a distribution G of truncating variables C_i's. Evolution in time of Y_n will strongly depend on this choice. However, this evolution is not looked at too much and evolution of N_n^c is studied, as it is implied by (24).

With $\lambda_{n,\theta}(t) = c_\theta(t)Y_n(t)$, let us clarify now the limit behaviour of the functions $\lambda_{n,\theta}(t)/n$ and $[\dot\lambda/\lambda]_{n,\theta}$. From the definition of $Y_n(t)$ and the Law of Large Numbers, it follows that, as $n \to \infty$,

$$c_\theta(t)\frac{Y_n(t)}{n} \to c_\theta(t)[1 - F_\theta(t)][1 - G(t)] = f_\theta(t)[1 - G(t)],$$

while

$$[\dot\lambda/\lambda]_{n,\theta}(t) = [\dot c/c]_\theta(t).$$

Therefore, if we choose $c_\theta(t)$ the same as in Example A, i.e., the failure rate of Weibull's distribution, then all will not differ from what we said in that example. However, this time the limit of $\lambda_{n,\theta}(t)/n$ is not a probability density.

Example D. Marked point processes. The example is interesting and has many applications, but is not treated here. We think it will be another example of regular models which permit the treatment as described in Section 3.

Acknowledgement

I would like to thank Prof. Robert Mnatsakanov, WVU, for exceptionally careful reading of the text at the final stages and for useful and fruitful discussions that followed.

References

1. P. K. Andersen, O. Borgan, R. D. Gill, N. Keiding, *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer-Verlag, New York, 1993.
2. D. J. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes. vol. I. Elementary Theory and Methods*. Second edition. Probability and its Applications (New York). Springer-Verlag, New York, 2003; *vol. II. General Theory and Structure*. Second edition. Probability and its Applications (New York). Springer, New York, 2008.
3. A. F. Karr, *Point Processes and their Statistical Inference*. Second edition. Probability: Pure and Applied, 7. Marcel Dekker, Inc., New York, 1991.
4. E. Khmaladze, The use of Omega-square Tests for Testing Parametric Hypotheses. *Theory Probab. Appl.* **24** (1979), no. 2, 283–302.
5. E. Khmaladze, Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability and Its Applications* **26** (1982), no. 2, 240–257.
6. E. Khmaladze, Note on distribution free testing for discrete distributions. *Ann. Statist.* **41** (2013), no. 6, 2979–2993.
7. E. Khmaladze, Unitary transformations, empirical processes and distribution free testing. *Bernoulli* **22** (2016), no. 1, 563–588.
8. E. Khmaladze, Distribution free testing for conditional distributions given covariates. *Statist. Probab. Lett.* **129** (2017), 348–354.
9. L. Roberts, Distribution free testing of goodness of fit in a one dimensional parameter space. *Statist. Probab. Lett.* **99** (2015), 215–222.
10. G. R. Shorack, J. A. Wellner, *Empirical Processes with Applications to Statistics*. SIAM, Philadelphia, PA, 2009.

(Received 17.02.2020)

Victoria University of Wellington, Wellington, New Zealand
E-mail address: Estate.Khmaladze@vuw.ac.nz

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 175–188

OPTIMAL RATE ESTIMATION OF THE MIXING DISTRIBUTION IN POISSON MIXTURE MODELS VIA LAPLACE INVERSION

CHRIS A. J. KLAASSEN[1] AND ROBERT M. MNATSAKANOV[2]

Dedicated to Estate Khmaladze on the occasion of his 75th birthday

Abstract. Consistent estimators of the mixing distribution in Poisson mixture models are constructed for both the right censored and the uncensored case. The estimators are based on a kind of Laplace inversion via factorial moments. The rate of convergence of the mean integrated squared error of these estimators is $(\log n/\log\log n)^2$. It is also shown that there do not exist estimators for which this rate is better.

1. Introduction

Consider independent and identically distributed random variables $X, X_1, X_2, \ldots, X_n$ with discrete distribution

$$p(x) = P(X = x) = \int_0^\infty e^{-\lambda}\frac{\lambda^x}{x!}dG(\lambda), \quad x = 0, 1, \ldots . \tag{1.1}$$

In this Poisson mixture model we shall study nonparametric estimation of the unknown mixing distribution G. This estimation problem is discussed by H. Robbins in [12] (pages 162–163) who suggests to estimate the distribution p of the observations and to solve (1.1) for G with p replaced by its estimate. We apply this approach and solve (1.1) via a kind of Laplace inversion as in Section 4 of [11]. We will investigate the rate of convergence of our estimators as measured by their mean integrated squared error, both in the censored and in the uncensored case. We will also prove this rate to be optimal.

Papers [6] and [8] define multinomial models with a large number of rare events, introduce the concept of a structural function, and discuss its estimation. For polynomial distributions and occupancy problems with a large number of rare events, asymptotic results for the relevant statistics are obtained in [9] and [7], respectively. In [14], the kernel type estimators of the structural distribution function in the multinomial scheme of [6] and [8] are studied via Poissonization.

Approximating the binomial marginals of such multinomial models by Poisson distributions, one arrives at the Poisson mixture model with the distribution function G as a structural function.

2. Construction of the Inverse Transformation

Consider the inhomogeneous Fredholm equation of the first kind (cf. [3])

$$\mathcal{K}G = p, \tag{2.1}$$

where the probability mass function $p(x)$, $x = 0, 1, \ldots$, denotes the Poisson mixture distribution from (1.1). Our construction of estimators of the unknown mixing distribution G is based on a particular type of Laplace inversion as in [2] (Section VII.6, formulae (6.1)–(6.4)). For (2.1) it can be written as follows:

$$(\mathcal{K}_\alpha^{-1}\mathcal{K}G)(z) = (\mathcal{K}_\alpha^{-1}p)(z) = \sum_{k=0}^{\lfloor \alpha z\rfloor} \frac{\alpha^k}{k!}\sum_{j=k}^\infty \frac{(-\alpha)^{j-k}}{(j-k)!}\sum_{x=j}^\infty \frac{x!}{(x-j)!}\, p(x), \tag{2.2}$$

where $\lfloor y\rfloor$ denotes the integer part of y.

2020 *Mathematics Subject Classification.* Primary: 62G05, Secondary: 62G20.
Key words and phrases. Demixing; Poisson mixture; Asymptotics; Mean integrated squared error.

Lemma 2.1. *Assume* $\int_0^\infty e^{\beta\lambda} dG(\lambda) < \infty$ *for all* $\beta \in \mathbb{R}$. *As* α *tends to infinity,* $\mathcal{K}_\alpha^{-1}$ *from* (2.2) *represents inversion of* $\mathcal{K}$ *from* (2.1) *in the weak sense, i.e.,*

$$\mathcal{K}_\alpha^{-1}\mathcal{K}G \xrightarrow{w} G, \quad \text{as} \quad \alpha \to \infty.$$

Moreover, the transformation $\mathcal{K}_\alpha^{-1}$ *can be written as*

$$(\mathcal{K}_\alpha^{-1}\mathcal{K}G)(z) = (\mathcal{K}_\alpha^{-1}p)(z) = \sum_{x=0}^{\infty} \sum_{k=0}^{\lfloor \alpha z \rfloor \wedge x} \binom{x}{k} \alpha^k (1-\alpha)^{x-k} \; p(x). \tag{2.3}$$

Proof. The factorial moments of the Poisson distribution are powers of its parameter. Consequently, Fubini's theorem (actually Tonelli's theorem) implies

$$\sum_{x=j}^{\infty} \frac{x!}{(x-j)!} p(x) = \int_0^\infty \sum_{x=j}^{\infty} \frac{x!}{(x-j)!} e^{-\lambda} \frac{\lambda^x}{x!} dG(\lambda) = \int_0^\infty \lambda^j dG(\lambda), \; j = 0, 1, \dots . \tag{2.4}$$

Note that retrieving G from these moments is called the moment problem. Subsequently, as the Laplace transform $\int_0^\infty e^{\beta\lambda} dG(\lambda)$ of G is finite for all $\beta \in \mathbb{R}$, Fubini's theorem yields

$$\sum_{j=k}^{\infty} \frac{(-\alpha)^{j-k}}{(j-k)!} \int_0^\infty \lambda^j dG(\lambda) = \int_0^\infty \sum_{j=k}^{\infty} \frac{(-\alpha\lambda)^{j-k}}{(j-k)!} \lambda^k dG(\lambda) = \int_0^\infty e^{-\alpha\lambda} \lambda^k dG(\lambda). \tag{2.5}$$

From (2.2), (2.4) and (2.5) it follows, again by Tonelli's theorem, that

$$(\mathcal{K}_\alpha^{-1}\mathcal{K}G)(z) = \sum_{0 \le k \le \alpha z} \frac{\alpha^k}{k!} \int_0^\infty e^{-\alpha\lambda} \lambda^k dG(\lambda) = \int_0^\infty \sum_{0 \le k \le \alpha z} e^{-\alpha\lambda} \frac{(\alpha\lambda)^k}{k!} dG(\lambda).$$

Chebyshev's inequality for a Poisson random variable Z with parameter $\alpha\lambda$ yields
$P(|Z - \alpha\lambda| \ge \alpha|z - \lambda|) \le \lambda/(\alpha(z-\lambda)^2)$. Hence we obtain

$$\left| \sum_{0 \le k \le \alpha z} e^{-\alpha\lambda} \frac{(\alpha\lambda)^k}{k!} - \mathbf{1}_{[\alpha\lambda \le \alpha z]} \right| \le \left(\frac{\lambda}{\alpha(z-\lambda)^2} \right) \wedge 1.$$

Consequently, at any point of continuity z of G we have

$$(\mathcal{K}_\alpha^{-1}\mathcal{K}G)(z) \to \int_0^\infty \mathbf{1}_{[\lambda \le z]} dG(\lambda) = G(z),$$

as $\alpha \to \infty$; cf. formula (6.1) from Section VII.6 of [2].

Furthermore, from (2.4) and (2.5) with $-\alpha$ replaced by α it follows that Fubini's theorem may be applied to

$$\begin{aligned}\sum_{j=k}^{\infty} \frac{(-\alpha)^{j-k}}{(j-k)!} \sum_{x=j}^{\infty} \frac{x!}{(x-j)!} \; p(x) &= \sum_{x=k}^{\infty} \sum_{j=k}^{x} \binom{x-k}{j-k} (-\alpha)^{j-k} \frac{x!}{(x-k)!} \; p(x) \\ &= \sum_{x=k}^{\infty} (1-\alpha)^{x-k} \frac{x!}{(x-k)!} \; p(x),\end{aligned}$$

which implies

$$(\mathcal{K}_\alpha^{-1}p)(z) = \sum_{k=0}^{\lfloor \alpha z \rfloor} \sum_{x=k}^{\infty} \binom{x}{k} \alpha^k (1-\alpha)^{x-k} \; p(x). \tag{2.6}$$

For $\alpha > 1$, Tonelli's theorem yields

$$\begin{aligned}\sum_{k=0}^{\lfloor \alpha z\rfloor}\sum_{x=k}^{\infty}\binom{x}{k}\alpha^k(\alpha-1)^{x-k}\ p(x) &= \sum_{x=0}^{\infty}\sum_{k=0}^{\lfloor \alpha z\rfloor\wedge x}\binom{x}{k}\alpha^k(\alpha-1)^{x-k}\ p(x)\\ &\le \sum_{x=0}^{\infty}(2\alpha-1)^x\ p(x) = \sum_{x=0}^{\infty}(2\alpha-1)^x\int_0^\infty e^{-\lambda}\frac{(\lambda)^x}{x!}dG(\lambda)\\ &= \int_0^\infty e^{-\lambda}\sum_{x=0}^{\infty}\frac{((2\alpha-1)\lambda)^x}{x!}dG(\lambda) = \int_0^\infty e^{2(\alpha-1)\lambda}dG(\lambda). \end{aligned} \tag{2.7}$$

By the finiteness of the Laplace transform of G the right hand side of (2.7) is finite. Consequently, Fubini's theorem can be applied to (2.6), which yields (2.3). □

Estimating the probability mass function $\mathcal{K}G = p$ from the observations and applying (2.3), we can see by Lemma 2.1 that we might obtain consistent estimators of the mixing distribution G. This is verified for the case of i.i.d. uncensored random variables in Section 3, and under random right censoring in Section 4.

We remark here that the estimator of the so-called structural distribution function for a multinomial random variable discussed in Section 4 of [11] is also based on inversion (2.3) with p replaced by an appropriate empirical version of p.

3. Uncensored Data

Let $X, X_1, \ldots, X_n$ be i.i.d. random variables with the Poisson mixture distribution p as in (1.1); cf. (2.1). Replacing the marginal distribution $p(x) = P(X = x)$ in (2.3) by the corresponding empirical version

$$\hat{p}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{[X_i=x]}, \tag{3.1}$$

restricting the sum over x in (2.3) to $x \le K_n$, and taking $\alpha = \alpha_n > 1$ dependent on n, we obtain the estimator $\hat{G}_n$ of G with

$$\hat{G}_n(z) = \sum_{x=0}^{K_n}\sum_{k=0}^{\lfloor \alpha_n z\rfloor\wedge x}\binom{x}{k}\alpha_n^k(1-\alpha_n)^{x-k}\ \hat{p}_n(x), \quad z \ge 0. \tag{3.2}$$

In view of Lemma 2.1 and (2.3), the estimator $\hat{G}_n$ should be consistent for appropriately chosen α_n and K_n that tend to infinity when n does. Under reasonable assumptions on the class of mixing distribution functions G it is consistent indeed.

Theorem 3.1. *Let C, D and L be positive constants. Let $G(D) = 1$ hold and let G have a density g that is bounded by C and is Lipschitz continuous with the Lipschitz constant L. Then the mean integrated squared error of $\hat{G}_n$ with $K_n \ge 2\alpha_n De^2$ and $\alpha_n \ge 1$ satisfies*

$$E\int_0^\infty\left(\hat{G}_n(z)-G(z)\right)^2 dG(z) \le \frac{1}{n}(2\alpha_n)^{2K_n} + 2\frac{(C+\frac{1}{2}L(D+2))^2}{\alpha_n^2} + 2e^{-2K_n}. \tag{3.3}$$

Furthermore,

$$E\int_0^\infty\left(\hat{G}_n(z)-G(z)\right)^2 dG(z) = \mathcal{O}\left(\left(\frac{\log\log n}{\log n}\right)^2\right) \tag{3.4}$$

holds as $n \to \infty$, when α_n and K_n are chosen as

$$\alpha_n = \frac{\log n}{\gamma\log\log n}, \quad K_n = \left\lceil\frac{\log n}{\kappa\log\log n}\right\rceil \tag{3.5}$$

with the constants γ and κ satisfying $\gamma \ge 2De^2\kappa,\ \kappa > 2$.

Our proof of this theorem is based on the representation of $\hat{G}_n$ as an average, to wit

$$\hat{G}_n(z) = \frac{1}{n}\sum_{i=1}^{n} B_n(z, X_i) \tag{3.6}$$

with

$$B_n(z,x) = \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge x} \binom{x}{k} {\alpha_n}^k \,(1-\alpha_n)^{x-k}\, \mathbf{1}_{[x \le K_n]}. \tag{3.7}$$

Subsequently, both the variance and the bias part of the mean integrated squared error are studied in Appendix A.

4. Randomly Right Censored Data

Suppose now that $X, X_1, \ldots, X_n$ are i.i.d. random variables with distribution $p(x) = P(X = x)$ given by (1.1) and that $Y, Y_1, \ldots, Y_n$ are i.i.d. nonnegative random variables distributed according to some distribution function H. Assume that the X's and Y's are independent and that one observes $Z_i = \min(X_i, Y_i)$ and $\Delta_i = \mathbf{1}_{[X_i \le Y_i]}$ only. We are interested in estimation of the unknown mixing distribution function G in this random censoring model.

It is known that the distribution of the X_i's can be estimated at the same $\sqrt{n}$ rate as in the uncensored case, provided the right censoring is not too strict (cf. [4]). Therefore, it should be possible to estimate the mixing distribution under right censoring at the same rate as without censoring. Our results here confirm this heuristic.

First consider the case where the censoring distribution function H is known. Observe

$$\begin{aligned} P(Z_i = x, \Delta_i = 1) &= P(X_i = x, X_i \le Y_i) \\ &= P(X_i = x)(1 - H(x-)) = p(x)(1 - H(x-)), \quad x = 0, 1, \ldots . \end{aligned} \tag{4.1}$$

Consequently, using the observations Z_i and Δ_i, we can estimate $p(x)$ by the following empirical expression:

$$\tilde{p}_n(x) = \frac{1}{1 - H(x-)}\ \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[Z_i = x, \Delta_i = 1]} \tag{4.2}$$

for those x for which $1 - H(x-)$ is positive. In analogy to (3.1), (3.2) and (3.6) we construct our estimator of the unknown mixing distribution function G as follows. For $\alpha_n > 1$ and K_n a positive integer, we define

$$\begin{aligned} \tilde{G}_n(z) &= \sum_{x=0}^{K_n} \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge x} \binom{x}{k} \alpha_n^k (1-\alpha_n)^{x-k}\ \tilde{p}_n(x) \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(Z_i-)} \sum_{x=0}^{K_n} \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge x} \binom{x}{k} \alpha_n^k (1-\alpha_n)^{x-k}\ \mathbf{1}_{[Z_i = x]} \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(Z_i-)} \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge Z_i} \binom{Z_i}{k} \alpha_n^k (1-\alpha_n)^{Z_i-k}\ \mathbf{1}_{[Z_i \le K_n]}. \end{aligned} \tag{4.3}$$

Note that this estimator has the form

$$\tilde{G}_n(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(Z_i-)} B_n(z, Z_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(X_i-)} B_n(z, X_i), \tag{4.4}$$

where again $B_n(z,x)$ is defined by (3.7).

Studying (4.3), we see that if the censoring random variables Y_i have bounded support, then for $\alpha_n z$ and K_n large our estimator reduces to

$$\tilde{G}_n(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{(1 - H(X_i-))}\,,$$

which by the Law of Large Numbers converges to 1. Consequently, it is crucial for the consistency of our estimator that the right-hand tail of H be not too thin. In fact, no estimator can behave properly if H has bounded support $[0,\tau]$, say, as in this case it is possible to estimate $p(x)$ consistently only for $x=0,1,\dots,\lfloor\tau\rfloor$. However, the mixing distribution G is **not** identifiable from the $\lfloor\tau\rfloor+1$ equations

$$p(x)=\int_0^\infty e^{-\lambda}\,\frac{\lambda^x}{x!}\,dG(\lambda),\quad x=0,1,\dots,\lfloor\tau\rfloor. \tag{4.5}$$

Actually, in Section 5, we construct mixing densities g and g_n such that they differ, but yield the same values of $p(x)$ in (4.5) for $x=0,1,\dots,m-5$, where m may be chosen arbitrarily large (see (5.5) up to but not including (5.9)). Because of this unidentifiability phenomenon we will assume that the right-hand tail of the censoring distribution H does not decrease too fast. More precisely, we will assume condition (1.1) from [4].

Assumption 4.1. There exists a finite constant M with

$$\sum_{x=0}^{\infty}\frac{1}{P(Y\ge x)}\,p(x)=\sum_{x=0}^{\infty}\frac{1}{1-H(x-)}\,p(x)\le M. \tag{4.6}$$

As $\tilde G_n$ and $\hat G_n$ are similar averages (cf. (3.6), (4.3) and (4.4)), we can establish the consistency of $\tilde G_n$ along the lines of the proof of Theorem 3.1 as given in Appendix A. In the censored case, $\tilde G_n$ attains the same rate as $\hat G_n$ in the uncensored case.

Theorem 4.1. *Let the conditions of Theorem* 3.1 *be satisfied and let α_n and K_n be chosen as in* (3.5)*. If the censoring distribution H is known and fulfills Assumption* 4.1*, then the mean integrated squared error of $\tilde G_n$ is of the order* $(\log\log n/\log n)^2$ *as $n\to\infty$. More precisely,*

$$E\int_0^\infty\Big(\tilde G_n(z)-G(z)\Big)^2\,dG(z)\le\frac{1}{n}(2\alpha_n)^{2K_n}M+2\frac{(C+\frac12 L(D+2))^2}{\alpha_n^2}+2e^{-2K_n} \tag{4.7}$$

holds.

Proof. First, we estimate the variance of $\tilde G_n(z)$ under Assumption 4.1 (see (A.4)) as follows:

$$\begin{aligned}\operatorname{var}\tilde G_n(z)&=\frac1n\operatorname{var}\left(\frac{\Delta_1}{1-H(Z_1-)}B_n(z,Z_1)\right)\le\frac1n E\left(\frac{B_n^2(z,X)}{(1-H(X-))^2}E(\mathbf{1}_{[X\le Y]}\mid X)\right)\\&\le\frac1n E\left(\frac{(2\alpha_n)^{2K_n}}{(1-H(X-))}\right)\le\frac1n(2\alpha_n)^{2K_n}M.\end{aligned} \tag{4.8}$$

Furthermore, the bias of $\tilde G_n$ equals

$$E\left(\frac{1}{1-H(X-)}B_n(z,X)E(\mathbf{1}_{[X\le Y]}\mid X)\right)-G(z)=E\big[B_n(z,X)\big]-G(z),$$

which in view of (A.6) is the same expression as in (A.7). Together with (A.1), (A.2), (4.8) and (A.11) this yields (4.7) and hence the Theorem. □

Next, we consider the case where the survival function $S=1-H$ of the censoring variable Y is unknown, but is known to be continuous. Observe that the estimator $\hat G_n$ for the non-censored case (cf. (3.6)) can de written as $\hat G_n(z)=\int B_n(z,x)d\hat F_n(x)$ with $\hat F_n$ the empirical distribution function of X. So it is natural in the censored case to consider $\tilde G_n^{\mathrm{KM}}(z)=\int B_n(z,x)d\tilde F_n(x)$ with $\tilde F_n$ the Kaplan-Meier estimator of the distribution function of X. However, we have not been able to study the asymptotic performance of the mean integrated squared error of this estimator of G.

Therefore, we construct another estimator. It is based on the technique of sample splitting as in [10]. To explain the idea, we assume for the time being that we have an extra sample $(\tilde\Delta,\tilde Z)=((\tilde\Delta_1,\tilde Z_1),\dots,(\tilde\Delta_n,\tilde Z_n))$ available of size n, that is, independent of and identically distributed to

$((\Delta_1, Z_1), \ldots, (\Delta_n, Z_n))$. The product-limit or Kaplan-Meier estimator $\tilde{S}_n$ of the survival function S based on this extra sample is defined as

$$\tilde{S}_n(x) = \begin{cases} 1, & 0 \le x \le \tilde{Z}_{(1)}, \\ \prod_{i=1}^{k-1} (\frac{n-i}{n-i+1})^{1-\tilde{\Delta}_{(i)}}, & \tilde{Z}_{(k-1)} < x \le \tilde{Z}_{(k)}, \quad k = 2, \ldots, n, \\ 0, & \tilde{Z}_{(n)} < x, \end{cases}$$

where $\tilde{Z}_{(i)}$ and $\tilde{\Delta}_{(i)}$ denote the ordered $\tilde{Z}_i$'s and corresponding $\tilde{\Delta}_i$'s. Note that $\tilde{S}_n$ is well defined, as there are no ties among the $\tilde{Z}_{(i)}$'s for which the $\tilde{\Delta}_{(i)}$'s vanish in view of the continuity of H.

We define δ_n and redefine K_n as follows:

$$\delta_n = n^{\frac{1}{2\kappa_0} - \frac{1}{2}} \sqrt{\log n}, \quad K_n = \left\lceil \frac{\log n}{\kappa \log \log n} \right\rceil, \quad \kappa > \kappa_0 > 0. \tag{4.9}$$

Replacing in (4.2), (4.3) and (4.4) the survival function $S = 1 - H$ by its estimator $\tilde{S}_n$ with δ_n added to it, we obtain our estimator

$$\tilde{G}_n^*(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{\tilde{S}_n(Z_i-) + \delta_n} B_n(z, Z_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{\tilde{S}_n(X_i-) + \delta_n} B_n(z, X_i),$$

which is based on the original sample together with the extra one.

Following the proof of Theorem 4.1, we estimate the variance of $\tilde{G}_n^*(z)$ as follows:

$$\begin{aligned} \operatorname{var}\left(\tilde{G}_n^*(z) \,\middle|\, \tilde{\Delta}, \tilde{Z}\right) &= \frac{1}{n} \operatorname{var}\left(\frac{\Delta_1}{\tilde{S}_n(X_1-) + \delta_n} B_n(z, X_1) \,\middle|\, \tilde{\Delta}, \tilde{Z}\right) \\ &\le \frac{1}{n} E\left(\frac{B_n^2(z, X)}{\left(\tilde{S}_n(X-) + \delta_n\right)^2} E(\mathbf{1}_{[X \le Y]} \mid X) \,\middle|\, \tilde{\Delta}, \tilde{Z}\right) \\ &\le \frac{1}{n} E\left(\frac{S(X-)(2\alpha_n)^{2K_n}}{\left(\tilde{S}_n(X-) + \delta_n\right)^2} \,\middle|\, \tilde{\Delta}, \tilde{Z}\right) \le \frac{1}{n} E\left(\frac{S(X-)(2\alpha_n)^{2K_n}}{\left(S(X-)[1 - \tilde{D}_n] + \delta_n\right)^2} \,\middle|\, \tilde{\Delta}, \tilde{Z}\right), \end{aligned} \tag{4.10}$$

where $\tilde{D}_n$ is defined as

$$\tilde{D}_n = \sup_{0 \le y \le K_n} \left| \frac{\tilde{S}_n(y) - S(y)}{S(y)} \right|.$$

From (B.3) in Appendix B we know that for large n

$$-S(X-)\tilde{D}_n + \delta_n \ge 0 \quad \text{almost surely}$$

holds, which combined with (4.10) and Assumption 4.1 yields

$$\operatorname{var}\left(\tilde{G}_n^*(z) \,\middle|\, \tilde{\Delta}, \tilde{Z}\right) \le \frac{1}{n} E\left(\frac{(2\alpha_n)^{2K_n}}{S(X-)}\right) \le \frac{1}{n} (2\alpha_n)^{2K_n} M, \quad \text{a.s.} \tag{4.11}$$

Note that for $0 \le x \le K_n$ and large n formula (B.3) implies $\tilde{S}_n(x-) + \delta_n \in [S(x-), S(x-) + 2\delta_n]$ a.s. and hence

$$1 - \frac{S(x-)}{\tilde{S}_n(x-) + \delta_n} \in \left[0, \frac{2\delta_n}{S(x-)}\right] \quad \text{a.s.}$$

By (A.4) and Assumption 4.1, this implies that conditionally on the extra sample the bias of our estimator satisfies

$$\begin{aligned} \left| E\, B_n(z, X) - E\left(\tilde{G}_n^*(z) \mid \tilde{\Delta}, \tilde{Z}\right) \right| &\le E\left(\left| \left[1 - \frac{S(X-)}{\tilde{S}_n(X-) + \delta_n}\right] \; B_n(z, X) \right| \mid \tilde{\Delta}, \tilde{Z} \right) \\ &\le 2\delta_n (2\alpha_n)^{K_n} E\left(\frac{1}{S(X-)}\right) \le 2\delta_n (2\alpha_n)^{K_n} M \quad \text{a.s.} \end{aligned}$$

Together with (A.11) this means (see also (4.9))

$$\int_0^\infty \left[E\left(\tilde G_n^*(z) \mid \tilde\Delta, \tilde Z\right) - G(z)\right]^2 dG(z)$$

$$\le 4\frac{\log n}{n^{1-1/\kappa_0}}(2\alpha_n)^{2K_n}M^2 + 4\frac{(C+\frac12 L(D+2))^2}{\alpha_n^2} + 4e^{-2K_n}, \quad \text{a.s.},$$

which in combination with (4.11) results in

$$E\int_0^\infty \left(\tilde G_n^*(z) - G(z)\right)^2 dG(z)$$

$$\le \frac1n(2\alpha_n)^{2K_n}M + 4\frac{\log n}{n^{1-1/\kappa_0}}(2\alpha_n)^{2K_n}M^2 + 4\frac{(C+\frac12 L(D+2))^2}{\alpha_n^2} + 4e^{-2K_n}.$$

With α_n and K_n defined as in (3.5) and (4.9), where the constants γ and κ satisfy $\gamma \ge 2De^2\kappa$, $\kappa > \kappa_0 = 3$, we obtain

$$\left(\frac{\log n}{\log\log n}\right)^2 E\int_0^\infty \left(\tilde G_n^*(z) - G(z)\right)^2 dG(z) = \mathcal{O}(1),$$

as n tends to infinity.

However, $\tilde G_n^*$ is based on $2n$ observations, as it is based on the original and the extra samples. But, if one has a sample of n observations available, the rate $\log n/(\log\log n)$ can still be obtained by splitting the sample into two subsamples of about the same size and applying the natural modification of our construction in order to get an estimate of G. Here, the first subsample plays the role of the original sample in our construction and the second subsample the role of the extra sample. Interchanging the roles of the two subsamples one gets another estimate of G and it makes sense to average these two estimates to obtain $\tilde G_n^{**}$ (cf. [13], Section 2 of [10], or page 396 of [1]). In summary,

Theorem 4.2. *Fix positive constants $C, D, L,$ and β with $CD \ge 1$. Consider the class of mixing distributions G that have support contained in $[0, D]$ and have a density bounded by C that is Lipschitz continuous with Lipschitz constant L. For the class of censoring distributions H that are continuous and fulfill Assumption 4.1, there exists an estimator $\tilde G_n^{**}$ of G based on $(\Delta_1, Z_1), \ldots, (\Delta_n, Z_n)$, for which the mean integrated squared error is of the order $(\log\log n/\log n)^2$.*

5. Lower Bound to the Mean Integrated Squared Error

Information in the data about the mixing distribution G in the right censored case equals at most the information in the uncensored case. Therefore, the optimal lower bound to the mean integrated squared error for estimators of G in the censored case should have a convergence rate at most as large as the rate in the uncensored case. As we have seen in the preceding Section, our estimators for these two cases attain the same convergence rate $(\log n/\log\log n)^2$. In this Section we shall prove that the convergence rate equals at most this rate $(\log n/\log\log n)^2$ in the uncensored case, and hence in the censored case. Thus we have shown that our estimators attain the optimal rate and that our bound on the rate is also optimal, in both the uncensored and censored cases.

We study the minimax risk and note that it is bounded from below by a Bayes risk. Namely, we have

$$\inf_{\hat G_n}\sup_{G} E_G\int_0^\infty \left(\hat G_n(\lambda; X) - G(\lambda)\right)^2 dG(\lambda)$$

$$= \inf_{\hat G_n}\sup_{\alpha, G_0, G_n}\left\{\alpha E_{G_0}\int_0^\infty \left(\hat G_n(\lambda; X) - G_0(\lambda)\right)^2 dG_0(\lambda)\right.$$

$$+ (1-\alpha)E_{G_n} \int_0^\infty \left(\hat{G}_n(\lambda;X) - G_n(\lambda)\right)^2 dG_n(\lambda)\Big\}$$

$$\geq \sup_{\alpha,G_0,G_n} \inf_{\hat{G}_n} \Big\{ \alpha E_{G_0} \int_0^\infty \left(\hat{G}_n(\lambda;X) - G_0(\lambda)\right)^2 g_0(\lambda) d\lambda$$

$$+ (1-\alpha)E_{G_n} \int_0^\infty \left(\hat{G}_n(\lambda;X) - G_n(\lambda)\right)^2 g_n(\lambda) d\lambda \Big\}, \tag{5.1}$$

where the G's are supposed to have densities g with respect to the Lebesgue measure. We introduce the notation

$$p_{n0}(x) = p_{n0}(x_1,\dots,x_n) = \prod_{i=1}^n \left(\int_0^\infty e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}\, g_0(\lambda) d\lambda \right) \tag{5.2}$$

and similarly for $p_{nn}(x)$. Now the right-hand side of (5.1) can be written as

$$\sup_{\alpha,g_0,g_n} \inf_{\hat{G}_n} \sum_x \int_0^\infty \Big\{ \alpha \left(\hat{G}_n(\lambda;x) - G_0(\lambda)\right)^2 p_{n0}(x) g_0(\lambda)$$

$$+ (1-\alpha) \left(\hat{G}_n(\lambda;x) - G_n(\lambda)\right)^2 p_{nn}(x) g_n(\lambda) \Big\} d\lambda$$

and this infimum is attained by

$$\hat{G}_n(\lambda;x) = \frac{G_0(\lambda)\alpha p_{n0}(x) g_0(\lambda) + G_n(\lambda)(1-\alpha)p_{nn}(x)g_n(\lambda)}{\alpha p_{n0}(x)g_0(\lambda) + (1-\alpha)p_{nn}(x)g_n(\lambda)},$$

which results into

$$\sup_{\alpha,g_0,g_n} \int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 \sum_x \frac{\alpha(1-\alpha)p_{n0}(x)p_{nn}(x)g_0(\lambda)g_n(\lambda)}{\alpha p_{n0}(x)g_0(\lambda) + (1-\alpha)p_{nn}(x)g_n(\lambda)}\, d\lambda\,. \tag{5.3}$$

For positive reals s and t we have $st/(s+t) \geq \frac{1}{2}(s \wedge t)$. Consequently, the right-hand side of (5.3) is bounded from below by

$$\sup_{\alpha,g_0,g_n} \frac{1}{2} \int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 \sum_x \left\{ (\alpha p_{n0}(x) g_0(\lambda)) \wedge ((1-\alpha)p_{nn}(x)g_n(\lambda)) \right\} d\lambda\,,$$

which for $\alpha = \frac{1}{2}$ and combined with (5.1) through (5.3) results in

$$\inf_{\hat{G}_n} \sup_G E_G \int_0^\infty \left(\hat{G}_n(\lambda;X) - G(\lambda)\right)^2 dG(\lambda)$$

$$\geq \sup_{g_0,g_n} \frac{1}{4} \int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 (g_0(\lambda) \wedge g_n(\lambda))\, d\lambda \sum_x p_{n0}(x) \wedge p_{nn}(x)\,. \tag{5.4}$$

In order to come close to this supremum one has to choose g_0 and g_n in such a way that p_{n0} and p_{nn} are close together and that simultaneously G_0 and G_n are as different as possible. We shall choose g_0 and g_n with the help of the orthogonal system of Chebyshev polynomials C_m on $[-1,1]$, $m = 0,1,\dots$. They are defined as

$$C_m(z) = \cos(m \arccos z), \quad -1 \leq z \leq 1,$$

and are orthogonal with respect to the weight function $1/\sqrt{1-z^2}$, $-1 < z < 1$. Now we choose

$$g_0(\lambda) = \mathbf{1}_{(0,1)}(\lambda), \quad \lambda > 0,$$

$$H_n(\lambda) = \int_0^\lambda h_n(\mu) g_0(\mu) d\mu = [\lambda(1-\lambda)]^{3/2} e^\lambda C_m(2\lambda - 1), \tag{5.5}$$
$$g_n(\lambda) = g_0(\lambda)(1 + a_n h_n(\lambda)),$$

where $m = m_n$ depends on n in an appropriate way to be determined below. By differentiation, for $0 < \lambda < 1$, we obtain

$$h_n(\lambda) = \tfrac{1}{2}\sqrt{\lambda(1-\lambda)}(3 - 4\lambda - 2\lambda^2) e^\lambda C_m(2\lambda - 1)$$
$$- m\lambda(1-\lambda) e^\lambda \sin(m \arccos(2\lambda - 1)),$$

which we may bound by

$$|h_n(\lambda)| < \tfrac{1}{4} e(m+3). \tag{5.6}$$

In view of

$$\int_0^1 h_n(\lambda) g_0(\lambda) d\lambda = H_n(1) = 0,$$

equations (5.5) and (5.6) imply that g_n is a proper density provided

$$|a_n| \le \frac{4}{e(m+3)} \tag{5.7}$$

holds. With (5.5) in mind, by partial integration, we compute

$$\begin{aligned}
\int_0^\infty e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} h_n(\lambda) g_0(\lambda) d\lambda &= \left[e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} H_n(\lambda) \right]_0^1 - \int_0^1 e^{-\lambda} \frac{\lambda^{x_i - 1}}{x_i!} (x_i - \lambda) H_n(\lambda) d\lambda \\
&= \int_0^1 \frac{\lambda^{x_i - 1}}{x_i!} (\lambda - x_i) \frac{2\lambda^2(1-\lambda)^2}{\sqrt{1 - (2\lambda - 1)^2}} C_m(2\lambda - 1) d\lambda \\
&= \int_{-1}^1 \frac{(1+z)^{x_i - 1}}{2^{x_i + 4} x_i!} (z + 1 - 2x_i) \frac{(1 - z^2)^2}{\sqrt{1 - z^2}} C_m(z) dz.
\end{aligned} \tag{5.8}$$

As the Chebyshev polynomial of degree m is orthogonal with respect to the weight function $1/\sqrt{1-z^2}$, $-1 < z < 1$, to all polynomials of degree at most $m-1$, the integrals in (5.8) vanish for $x_i \le m-5$. Hence,

$$p_{nn}(x) = p_{n0}(x) \tag{5.9}$$

holds (cf. (5.2)), unless at least one of the x_i's equals $m-4$ or more. Actually, the probability q_n that X_i equals at least $m-4$, may be bounded both under g_0 and g_n via

$$q_n = P(X_i \ge m-4) = \int_0^1 \sum_{k=m-4}^\infty e^{-\lambda} \frac{\lambda^k}{k!} d\lambda \le \int_0^1 \frac{\lambda^{m-4}}{(m-4)!} d\lambda = \frac{1}{(m-3)!}. \tag{5.10}$$

Let Z_n be the random variable denoting the number of X_i that equal at least $m-4$. Note that Z_n has a binomial distribution with parameters n and q_n.

Combining (5.6) and (5.10), we arrive at

$$\begin{aligned}
\sum_x p_{n0}(x) \wedge p_{nn}(x) &\ge E_{g_0}\left(\left(1 - a_n \sup_{0<\lambda<1} |h_n(\lambda)| \right)^{Z_n} \right) \\
&= \left(1 - q_n a_n \sup_{0<\lambda<1} |h_n(\lambda)| \right)^n \ge \left(1 - \frac{e a_n (m+3)}{4(m-3)!} \right)^n,
\end{aligned} \tag{5.11}$$

which converges to $1/\sqrt{e}$ by the choice

$$a_n = \frac{2(m-3)!}{en(m+3)}. \tag{5.12}$$

For the time being we assume that

$$n = (m-3)! \tag{5.13}$$

holds. By Stirling's formula this means

$$\lim_{m\to\infty} \frac{(m-3)\log\log n}{\log n} = 1. \tag{5.14}$$

Note that these choices of n, m and a_n satisfy (5.7). Some computation shows that all together the above choices imply

$$\begin{aligned}
&\int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 (g_0(\lambda) \wedge g_n(\lambda))\, d\lambda = \int_0^\infty a_n^2 H_n^2(\lambda) (g_0(\lambda) \wedge g_n(\lambda))\, d\lambda \\
&\quad \geq \tfrac{1}{2} a_n^2 \int_0^1 \lambda^3 (1-\lambda)^3 e^{2\lambda} \cos^2(m \arccos(2\lambda - 1)) d\lambda \\
&\quad = 2^{-8} a_n^2 \int_0^\pi [(1+\cos\alpha)(1-\cos\alpha)]^3 e^{1+\cos\alpha} \cos^2(m\alpha) \sin\alpha\, d\alpha \\
&\quad \geq 2^{-8} a_n^2 \int_0^\pi \sin^7\alpha \cos^2(m\alpha)\, d\alpha \geq 2^{-23/2} a_n^2 \int_{\pi/4}^{3\pi/4} \cos^2(m\alpha)\, d\alpha.
\end{aligned} \tag{5.15}$$

Because this last integral converges to $\pi/4$ as m tends to infinity, the relations (5.15), (5.12), (5.13), and (5.14) imply

$$\liminf_{n=(m-3)!, m\to\infty} \left(\frac{\log n}{\log\log n}\right)^2 \int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 (g_0(\lambda) \wedge g_n(\lambda))\, d\lambda \geq 2^{-15/2} \pi e^2 > 0.$$

Together with (5.11), this yields

$$\begin{aligned}
\liminf_{n=(m-3)!, m\to\infty} &\left(\frac{\log n}{\log\log n}\right)^2 \int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 (g_0(\lambda) \wedge g_n(\lambda))\, d\lambda \\
&\sum_x p_{n0}(x) \wedge p_{nn}(x) \geq 2^{-15/2} \pi e^{3/2} > 0.
\end{aligned} \tag{5.16}$$

If $\tilde{n}$ satisfies $n = (m-3)! \leq \tilde{n} < (m-2)!$, then $1 \leq \tilde{n}/n < m-2$ holds and hence

$$1 \leq \frac{\log \tilde{n}}{\log n} < 1 + \frac{\log(m-2)}{\log n} \to 1, \quad \text{as } m \to \infty. \tag{5.17}$$

Combining (5.4), (5.16) and (5.17), we obtain the following lower bound.

Theorem 5.1. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with the Poisson mixture distribution* (1.1) *from the class of mixing distributions G that have density bounded by $C \in [2, \infty)$ and have $G(D) = 1$ for some $D \in [1, \infty)$. With $\hat{G}_n$ an estimator of G based on $X_1, \ldots, X_n$ the minimax value of the mean integrated squared error of $\hat{G}_n$ in estimating G does not tend to* 0 *faster than* $(\log\log n/\log n)^2$ *as n tends to infinity, more precisely,*

$$\liminf_{n\to\infty} \left(\frac{\log n}{\log\log n}\right)^2 \inf_{\hat{G}_n} \sup_G E_G \int_0^\infty \left(\hat{G}_n(\lambda; X) - G(\lambda)\right)^2 dG(\lambda) > 0.$$

Appendix A. Proof of Theorem 3.1

The proof in this Appendix of Theorem 3.1 will be based on (3.6) with (3.7). In view of

$$E\int_0^\infty \left(\hat{G}_n(z)-G(z)\right)^2 dG(z) = \int_0^\infty E\left(\hat{G}_n(z)-G(z)\right)^2 dG(z), \tag{A.1}$$

we first fix z, study

$$E\left(\hat{G}_n(z)-G(z)\right)^2 = \text{var}\ \hat{G}_n(z) + \left(E\hat{G}_n(z)-G(z)\right)^2 \tag{A.2}$$

and note that

$$\text{var}\ \hat{G}_n(z) = \frac{1}{n}\ \text{var}\ B_n(z,X). \tag{A.3}$$

As $B_n(z,x)$ from (3.7) satisfies

$$|B_n(z,x)| \le \sum_{k=0}^{x}\binom{x}{k}\alpha_n^k\,|1-\alpha_n|^{x-k}\,\mathbf{1}_{[x\le K_n]} \le (2\alpha_n)^{K_n}, \tag{A.4}$$

we have

$$\text{var}\ B_n(z,X) \le E\left(B_n^2(z,X)\right) \le (2\alpha_n)^{2K_n}. \tag{A.5}$$

The study of the bias is more involved. We choose the random variable Λ with a distribution function G in such a way that the conditional distribution of X given Λ is Poisson (Λ); so

$$P(X=x\mid\Lambda=\lambda) = e^{-\lambda}\frac{\lambda^x}{x!},\quad x=0,1,\ldots.$$

By Taylor's theorem (or partial integration),

$$e^x = \sum_{k=0}^{K}\frac{x^k}{k!} + \int_0^x \frac{(x-y)^K}{K!}e^y\,dy$$

holds. Consequently, we have

$$\begin{aligned}
E\left(B_n(z,X)\,|\,\Lambda=\lambda\right) &= \sum_{x=0}^{K_n}\sum_{k=0}^{\lfloor\alpha_n z\rfloor\wedge x}\binom{x}{k}{\alpha_n}^k\,(1-\alpha_n)^{x-k}\,e^{-\lambda}\,\frac{\lambda^x}{x!}\\
&= \sum_{k=0}^{\lfloor\alpha_n z\rfloor\wedge K_n} e^{-\lambda}\,\frac{(\alpha_n\lambda)^k}{k!}\sum_{x=k}^{K_n}\frac{1}{(x-k)!}\,((1-\alpha_n)\lambda)^{x-k}\\
&= \sum_{k=0}^{\lfloor\alpha_n z\rfloor\wedge K_n} e^{-\alpha_n\lambda}\,\frac{(\alpha_n\lambda)^k}{k!}\\
&\quad - \sum_{k=0}^{\lfloor\alpha_n z\rfloor\wedge K_n} e^{-\lambda}\,\frac{(\alpha_n\lambda)^k}{k!}\int_0^{(1-\alpha_n)\lambda}\frac{((1-\alpha_n)\lambda-y)^{K_n-k}}{(K_n-k)!}\,e^y\,dy\\
&= P\left(U_{\alpha_n\lambda}\le\alpha_n z\wedge K_n\right) - R_n(z,\lambda)
\end{aligned} \tag{A.6}$$

with U_μ, distributed as Poisson(μ). In view of $G(D)=1$, only z that are at most D, are relevant and for such z we have $\alpha_n z < K_n$. Hence, the bias of $\hat{G}_n$ equals

$$E\hat{G}_n(z)-G(z) = \int_0^\infty\left(P(U_{\alpha_n\lambda}\le\alpha_n z)-\mathbf{1}_{[\lambda\le z]}\right)dG(\lambda) - \int_0^\infty R_n(z,\lambda)dG(\lambda). \tag{A.7}$$

First, we note

$$|R_n(z,\lambda)| \le \sum_{k=0}^{\lfloor\alpha_n z\rfloor} e^{-\lambda}\,\frac{(\alpha_n\lambda)^k}{k!}\frac{((\alpha_n-1)\lambda)^{K_n-k}}{(K_n-k)!} \le e^{-\lambda}\,\frac{(\alpha_n\lambda)^{K_n}}{K_n!}\sum_{k=0}^{\lfloor\alpha_n z\rfloor}\binom{K_n}{k}$$

$$\leq e^{-\lambda}\,\frac{(2\alpha_n\lambda)^{K_n}}{K_n!} \leq \frac{e^{-\lambda}}{\sqrt{2\pi K_n}}\left(\frac{2\alpha_n\lambda e}{K_n}\right)^{K_n} \leq e^{-K_n}, \tag{A.8}$$

where the second to last inequality stems from Stirling's formula.

Furthermore, we note

$$G(0)=0, \quad \lim_{\lambda\to\infty} P(U_{\alpha_n\lambda} \leq \alpha_n z) = 0$$

and

$$\frac{\partial}{\partial\lambda} P(U_{\alpha_n\lambda} \leq \alpha_n z) = -\alpha_n e^{-\alpha_n\lambda}\,\frac{(\alpha_n\lambda)^{\lfloor\alpha_n z\rfloor}}{\lfloor\alpha_n z\rfloor!}.$$

Consequently, partial integration yields

$$\int_0^\infty P(U_{\alpha_n\lambda} \leq \alpha_n z)\,dG(\lambda) = \int_0^\infty G(\lambda)\alpha_n e^{-\alpha_n\lambda}\,\frac{(\alpha_n\lambda)^{\lfloor\alpha_n z\rfloor}}{\lfloor\alpha_n z\rfloor!}\,d\lambda = EG(\Lambda_n), \tag{A.9}$$

where Λ_n has a gamma distribution with shape parameter $\lfloor\alpha_n z\rfloor+1$ and rate parameter α_n. Hence we have

$$E\Lambda_n - z = \frac{1+\lfloor\alpha_n z\rfloor-\alpha_n z}{\alpha_n} \in (0,1/\alpha_n], \quad \mathrm{var}(\Lambda_n) = \frac{\lfloor\alpha_n z\rfloor+1}{\alpha_n^2},$$

$$E(\Lambda_n - z)^2 = \frac{\lfloor\alpha_n z\rfloor+1}{\alpha_n^2} + \left(\frac{1+\lfloor\alpha_n z\rfloor-\alpha_n z}{\alpha_n}\right)^2 \leq \frac{z+2}{\alpha_n}.$$

As G has a density g that is Lipschitz continuous with Lipschitz constant L, we have

$$|G(\lambda)-G(z)-(\lambda-z)g(z)| = \left|\int_z^\lambda (g(y)-g(z))\,dy\right| \leq \tfrac{1}{2}L(\lambda-z)^2. \tag{A.10}$$

Equations (A.9) through (A.10) yield

$$\left|\int_0^\infty \left(P(U_{\alpha_n\lambda} \leq \alpha_n z) - \mathbf{1}_{[\lambda\leq z]}\right) dG(\lambda)\right| \leq \frac{g(z)+\frac{1}{2}L(z+2)}{\alpha_n},$$

which together with (A.7) and (A.8) shows that the bias of $\hat{G}_n(z)$ satisfies

$$\left(E\hat{G}_n(z) - G(z)\right)^2 = \left(EB_n(z,X)-G(z)\right)^2$$

$$\leq 2\frac{(g(z)+\frac{1}{2}L(z+2))^2}{\alpha_n^2} + 2e^{-2K_n} \leq 2\frac{(C+\frac{1}{2}L(D+2))^2}{\alpha_n^2} + 2e^{-2K_n}, \quad z \leq D. \tag{A.11}$$

Together with (A.1)–(A.3) and (A.5) this yields (3.3) and consequently (3.4) when α_n is chosen as in (3.5). We have chosen $\kappa > 2$ because the first term at the right hand side of (3.3) is of the order $n^{-1+2/\kappa}$.

Appendix B. Kaplan–Meier

In Section 4, we have used the Kaplan-Meier estimator $\tilde{S}_n$ of the survival function S of the censoring distribution H. In this appendix we study the consistency of $\tilde{S}_n$ by applying Theorem 7 of [5]. We choose their d_n to be equal to $\log n/(\delta \log\log n)$ and their T_n to our K_n. Their ε_n is related to K_n via

$$\varepsilon_n = 8P(X \wedge Y > K_n). \tag{B.1}$$

As $\sum_{x\geq K_n+1} e^{-\lambda}\lambda^x/x!$ is decreasing in λ for $\lambda \leq K_n+1$, the support of G is contained in $[0,D]$, and $K_n+1 > D$ holds for n large, we have for such n

$$P(X > K_n) = \int_0^\infty \sum_{x=K_n+1}^\infty e^{-\lambda}\frac{\lambda^x}{x!}dG(\lambda) \geq C\int_0^{1/C} \sum_{x=K_n+1}^\infty e^{-\lambda}\frac{\lambda^x}{x!}\,d\lambda$$

$$\geq \sum_{x=K_n+1}^{\infty} e^{-1/C} \frac{(1/C)^x}{x!} \geq e^{-1/C} \frac{(1/C)^{K_n+1}}{(K_n+1)!}. \tag{B.2}$$

From (B.1), (B.2) and Assumption 4.1, we derive

$$\varepsilon_n \geq 8e^{-1/C} \frac{(1/C)^{K_n+1}}{(K_n+1)!} \beta^{-K_n}.$$

With K_n as in (4.9), some computation with the help of Stirling's formula shows

$$\varepsilon_n \geq \exp\left(-(\log n)/\kappa(1+o(1))\right),$$

which in view of $\kappa > \kappa_0$ implies that for sufficiently large n

$$\varepsilon_{2n} \geq n^{-1/\kappa_0}$$

holds. Now, formula (4.16) from Theorem 7 of [5] shows that almost surely

$$\sqrt{\frac{n^{1-1/\kappa_0}}{\log n}}\ \tilde{D}_n = \sqrt{\frac{n^{1-1/\kappa_0}}{\log n}} \sup_{0\leq y\leq K_n} \left| \frac{\tilde{S}_n(y) - S(y)}{S(y)} \right| \to 0 \tag{B.3}$$

holds as n tends to infinity.

Acknowledgement

This paper has been prepared under NSF grant DMS-0906639, partly during a visit of the first author to the Department of Statistics, West Virginia University, in August 2010. The excellent suggestions by the referee are gratefully acknowledged.

References

1. P. J. Bickel, C. A. J. Klaassen, Y. Ritov, J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 1993.
2. W. Feller, *An Introduction to Probability Theory and Its Applications*. vol. II. Second edition John Wiley & Sons, Inc., New York-London-Sydney, 1971.
3. E. I. Fredholm, Sur une classe d'équations fonctionnelles. (French) *Acta Math.* **27** (1903), no. 1, 365–390. doi:10.1007/bf02421317
4. R. D. Gill, Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11** (1983), no. 1, 49–58.
5. E.Giné, A. Guillou, Laws of the iterated logarithm for censored data. *Ann. Probab.* **27** (1999), no. 4, 2042–2067.
6. E. V. Khmaladze, *The statistical Analysis of a Large Number of Rare Events.* CWI Report MS-R8804, 1988.
7. E. V. Khmaladze, Convergence properties in certain occupancy problems including the Karlin–Rouault law. *J. Appl. Probab.* **48** (2011), no. 4, 1095–1113.
8. E. V. Khmaladze, R. Ya. Chitashvili, Statistical analysis of a large number of rare events and related problems. Probability theory and mathematical statistics. (Russian) *Trudy Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. SSR* **92** (1989), 196–245.
9. E. V. Khmaladze, Z. P. Tsigroshvili, On polynomial distributions with a large number of rare events. *Math. Methods Statist.* **2** (1993), no. 3, 240–247.
10. C. A. J. Klaassen, Consistent estimation of the influence function of locally asymptotically linear estimators. *Ann. Statist.* **15** (1987), no. 4, 1548–1562.
11. C. A. J. Klaassen, R. M. Mnatsakanov, Consistent estimation of the structural distribution function. *Scand. J. Statist.* **27** (2000), no. 4, 733–746.
12. H. Robbins, An empirical Bayes approach to statistics. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954-1955, vol. I, pp. 157–163. University of California Press, Berkeley and Los Angeles, 1956.
13. A. Schick, On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** (1986), no. 3, 1139–1151.
14. B. Van Es, C. A. J. Klaassen, R. M. Mnatsakanov, Estimating the structural distribution function of cell probabilities. *Austrian Journal of Statistics* **32** (2003), 85–98.

(Received 30.10.2019)

[1]Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248 1090 GE, The Netherlands
E-mail address: C.A.J.Klaassen@uva.nl

[2]Department of Mathematics, West Virginia University, P.O. Box 6310, Morgantown, WV 26506, USA
E-mail address: Robert.Mnatsakanov@mail.wvu.edu

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 189–197

ALMOST PARAMETRIC SMOOTHING

ROGER KOENKER[1] AND JIAYING GU[2]

Dedicated to Estate Khmaladze on the occasion of his 75th birthday

Abstract. Shape constraints may be a powerful aid in nonparametric function estimation, often regularizing problems without any pesky choice of tuning parameters. In some special circumstances they also achieve a remarkable, adaptive, nearly parametric convergence rate. After reviewing some prominent examples of this phenomenon, we briefly consider a closely related problem arising in the context of monotone single index models for conditional quantile functions.

1. Introduction

In regular, finite-dimensional parametric models we expect that estimated parameters converge at a rate, proportional to $1/\sqrt{n}$ for sample size n. A nonparametric estimation of densities and regression functions is, generally, more challenging, and this is typically reflected at slower rates of convergence. Of course, a higher order kernel density estimation enables one to achieve nearly parametric rates at the price of producing embarrassing estimates that may violate the basic non-negativity requirement for estimated densities; consequently, they will not be considered further here. Instead, we will focus on settings where shape constraints enable nearly parametric convergence in various related smoothing problems.

2. Monotone Density Estimation

The leading example of the phenomenon that we wish to study is the celebrated monotone density estimator of [10]. Given independent observations, $X_1, \dots, X_n$ from a distribution F_0 with a monotone decreasing density f_0, the classical prescription for the Grenander estimator is characterized as the left derivative of the least concave majorant of the empirical distribution function,

$$\mathbb{F}_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \le x).$$

This is illustrated in Figure 1, where the piecewise linear least concave majorant yields a piecewise constant density estimate. An especially appealing feature of this estimator is that it is fully automatic, not depending on any choice of tuning parameters. The location and mass associated with the resulting "histogram bins" are determined entirely from the data. This can be seen geometrically in the Figure: it is as if we have stretched a string over the empirical distribution function, and once this is done, the left derivative is determined. This may seem rather *ad hoc* on first encounter, so it is perhaps appropriate to find that the estimator can also be viewed as a nonparametric maximum likelihood estimator.

Consider the shape constrained density estimation problem,

$$\max_f \left\{ \int \log f(x) d\mathbb{F}_n(x) \mid f \text{ decreasing}, \int f(x)dx = 1 \right\}.$$

Lemma 2.2 of [13] establishes that the solution to this problem is the Grenander estimator, provided that we adopt the convention that $\hat{f}(x) = 0$ for $x < 0$. Jumps in $\hat{f}$ occur at the order statistics of a

2020 *Mathematics Subject Classification.* 62G07, 62G08.
Key words and phrases. Density estimation; Penalized likelihood, Quantile regression.

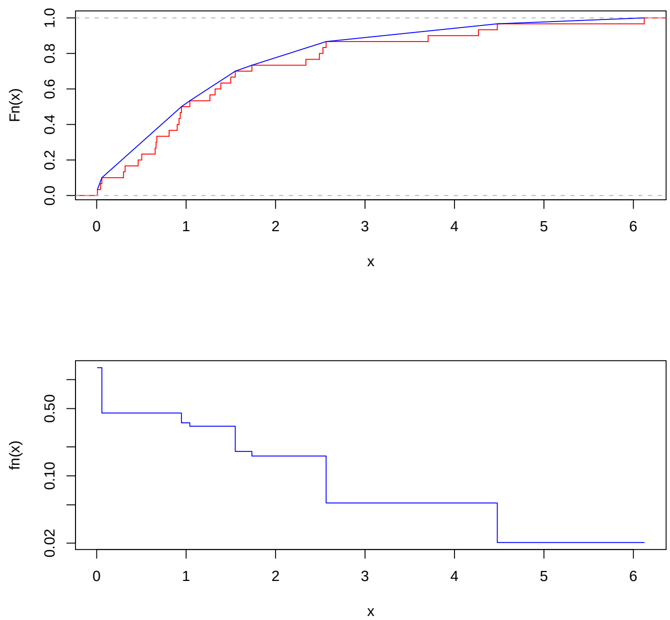


FIGURE 1. Grenander Estimator: The least concave majorant of the empirical distribution function in the upper panel, when differentiated yields the piecewise constant density estimate in the lower panel.

sample and at the origin. An alternative formulation, also grounded in maximum likelihood, involves the writing of our target density f, as a scale mixture of uniforms,

$$\max_{G \in \mathcal{G}} \left\{ \int \log f(x) d\mathbb{F}_n(x) \mid f(x) = \int t^{-1} I(0 \le x \le t) dG(t) \right\},$$

where $\mathcal{G}$ constitutes the set of proper distribution functions. In this case, solutions $\hat{G}$ assign a mass to a few discrete order statistics that then yield a mixture density $\hat{f}$, that is equal to the previous solutions.[1] The scale mixture formulation automatically imposes the constraint that the mixture density is supported on the positive half-line.

There is an extensive literature on the asymptotic behavior of the Grenander estimator beginning with [26], who established that pointwise,

$$n^{1/3}(\hat{f}_n(x_0) - f(x_0))/[4f(x_0)f'(x_0)]^{1/3} \rightsquigarrow Z$$

where Z is the maximizer of two-sided Brownian motion minus a parabola,

$$Z = \text{argmax}_t \{W(t) - t^2\}$$

The $\mathcal{O}(n^{-1/3})$ rate may be somewhat disappointing, however, it should be kept in mind that this result applies to the entire class of decreasing densities without the (second-order) differentiability conditions routinely assumed by kernel estimators to achieve their familiar $\mathcal{O}(n^{-2/5})$ rate.

Global convergence of the Grenander estimator was studied by [11], who established that for any bounded decreasing density f, with compact support on $[a, \infty)$ and continuous first derivative,

$$\lim_{n \to \infty} n^{1/3} R_n(f, \hat{f}_n) = K \int_a^\infty |f(x)f'(x)/2|^{1/3} dx$$

[1]For further computational details on these alternative formulations see `demo(Grenander) in the R package REBayes.`

where $R_n(f, \hat{f}_n) = \mathbb{E}_f \int |f(x) - \hat{f}_n(x)| dx$. Here, the constant K is more explicitly expressed as $2\mathbb{E}|V(0)| \approx 0.82$, where

$$V(a) = \sup\{t \in \mathbb{R} \mid W(t) = (t-a)^2 = \max!\}$$

and $W(t)$ is a two-sided Brownian motion on $\mathbb{R}$. [3] provide a more refined analysis of the local behavior at zero including the possibility of an unbounded target density. As is noted by [4], the uniformity is still problematic, so it is of considerable interest to have non-asymptotic risk bounds for $R_n(f, \hat{f}_n)$. To this end, L. Birgé shows that the piecewise constant, histogram-like nature of the Grenander estimator is adaptive in the sense that it tends to select an optimal partition for the binning strategy of the histogram. For smooth target densities this still yields a $\mathcal{O}(n^{-1/3})$ convergence rate, however, in the very special case that the target density is piecewise constant with a finite number of jumps, the results imply that $\hat{f}_n$ achieves the parametric rate $\mathcal{O}(n^{-1/2})$. The piecewise constant, histogram-like nature of the Grenander estimator is adaptive in the stronger sense that it selects a binning strategy suited to histogram nature of the true density as if it were a parametric object, which of course in a sense it is. Birgé is very careful to stress the special character of this result, so it may be easy to lose sight of this truly remarkable feature. In contrast to an adaptive kernel density estimation that requires a pilot estimate to guide the choice of the local bandwidth selection, the Grenander estimator constitutes its own pilot estimator, automatically selecting bins without the benefit of any preliminary bandwidth selection.[2]

3. Unimodal Density Estimation

This parametric rate performance of the Grenander estimator, however special its circumstances may be, turns out to have interesting extensions and counterparts in a wide variety of other shape constrained smoothing problems. For unimodal densities with the known mode results for the Grenander estimator can be immediately extended, and with some further effort an estimated mode can be accommodated. Closely related is the problem of estimating *strongly* unimodal, i.e., log-concave, densities. This is also a shape constrained problem susceptible to a maximum likelihood treatment,

$$\max_f \left\{ \sum_{i=1}^n \log f(x_i) \mid \log f \text{ concave}, \int f(x)dx = 1 \right\},$$

and can be reformulated as the convex optimization problem,

$$\min_g \left\{ \sum_{i=1}^n g(x_i) \mid g \in \mathcal{K}, \int e^{g(x)}dx = 1 \right\},$$

where $\mathcal{K}$ denotes the closed convex cone of convex functions. Solutions $\hat{g}_n$ are now piecewise linear with knots at the data points, so $\hat{f}_n = e^{\hat{g}_n}$ is piecewise exponential, and vanishes off the empirical support of the observations. Recently, [19] have proved that $\hat{f}_n$ achieves the minimax rate of convergence,

$$\inf_{f_n} \sup_{f_0 \in \mathcal{F}} \mathbb{E}_{f_0} d_H^2(f_n, f_0) \asymp n^{-4/5},$$

where $d_H^2(f,g) = \int(\sqrt{f(x)} - \sqrt{g(x)})^2 dx$ is the squared Hellinger distance, $\mathcal{F}$ denotes the set of all upper semi-continuous log concave densities, and f_n is any estimator of f_0. Again, it may be tempting to ask, "So what? Can't I achieve this same rate with conventional kernel methods?" When the target density f_0 is strictly log concave, the shape constraint is eventually rendered irrelevant, since any reasonable estimator would remain in the interior of the constraint set. What if, instead, f_0 lies in the boundary of the constraint set? In the log concave case this would mean that $g_0 = \log f_0$ was itself piecewise affine with k distinct pieces. In such k-affine cases, [17] establish that the non-parametric

[2]Indeed, this may lead one to wonder whether, in circumstances where the monotonicity assumption is plausible, it might be advantageous to use the Grenander $\hat{f}_n$ as a pilot estimator, simply convolving it with a smooth density if its piecewise constant appearance was deemed unattractive.

maximum likelihood estimator $\hat{f}_n$ achieves a nearly parametric rate of convergence, that is, there is a universal constant C such that for every $n \geq 2$ and every k-affine f_0,

$$\mathbb{E}_{f_0} d_H^2(f_n, f_0) \leq \frac{Ck}{n} \log^{5/4} n.$$

Thus, again without any prior knowledge about the number of affine pieces, the NPMLE achieves the almost parametric rate of $\mathcal{O}(1/\sqrt{n})$, without any required tuning parameter selection. In fact, something considerably more general is proved for f_0 that are nearly k-affine. It would also be possible to generalize to weaker forms of concavity as in [22], but we will resist going into the details. Instead, we will turn our attention to the estimation of a general class of mixture models.

4. Nonparametric Estimation of Mixture Densities

Many statistical problems can be formulated as parametric mixtures, leading examples are the Gaussian location mixture

$$f(x) = \int \varphi(x - \theta) dG(\theta)$$

and the Gaussian scale mixture

$$f(x) = \int \theta^{-1} \varphi(x/\theta) dG(\theta).$$

Given a sample of independent observations, $X_1, X_2, \ldots, X_n$, we can consider these as models with $X_i \sim \mathcal{N}(\theta_i, 1)$ and $X_i \sim \mathcal{N}(0, \theta_i^2)$, respectively. We would like to estimate the mixing distribution G when the observations are assumed to be exchangeable. [16] proposed estimating G by a nonparametric maximum likelihood,

$$\max_{G \in \mathcal{G}} \left\{ \sum_{i=1}^n \log f(X_i) \mid f(x) = \int \varphi(x, \theta) dG(\theta) \right\}, \tag{1}$$

and proved consistency of the resulting $\hat{G}$. Computation by the EM algorithm was suggested by [23], but remained quite challenging. Modern convex optimization methods provide a much more efficient and scalable approach to computation as is shown in [21]. However, many problems regarding the statistical performance of these methods remain open.

An important step forward in this respect is the recent work of [27] who consider the Gaussian location mixture model in $\mathbb{R}^d$. They evaluate performance relative to the oracle Bayes estimator that knows the empirical measure of the true θ's, $\mathbb{G}_n(t) = n^{-1} \sum_{i=1}^n I(t - \theta_i)$. Their Proposition 2.3 establishes that when $\mathbb{G}_n$ is discrete, supported on a set of cardinality k, there exists a constant C_d such that

$$\mathbb{E} d_H^2(\hat{f}_n, f_{\mathbb{G}_n}) \leq C_d \left(\frac{k}{n} (\sqrt{\log n})^{d + (4-d)_+} \right).$$

It follows easily that this is the minimax attainable rate. Again, we have an almost parametric convergence rate up to the logarithmic factor for the nonparametric MLE of the mixture density.

5. Shape Constrained Regression

It should not come as a big surprise that the shape constraints can also play an important role in regression, as well as in density estimation. Most of the literature has focused on the least squares fidelity criterion. The simplest setting is the isotonic regression model,

$$Y_i = \theta_i + u_i \quad i = 1, 2, \ldots, n,$$

where the θ_i are assumed to satisfy $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_n$. Implicitly, we can think of this model as one in which we observe Y_i's at a sequence of increasing design points. The apparently more general formulation of the model with $Y_i = g(x_i) + u_i$ reduces to the former model under general convex loss; if the observations are not ordered in the covariate x_i, we can simply reorder the Y_i's according to the order of the x_i's and proceed as before. Under the monotonicity constraint, the solutions are piecewise constant with jumps at the design points and loss depends only on the estimated function values at these design points.

For i.i.d. Gaussian u_i with variance $\sigma^2 < \infty$ the nonparametric MLE can again be formulated as a convex optimization problem

$$\min_\theta \left\{ \sum_{i=1}^n (Y_i - \theta_i)^2 \mid \theta \in \mathcal{K}_n \right\},$$

where $\mathcal{K}_n$ is the convex polyhedral cone of nondecreasing sequences. This problem has a long history going back to [5] and perhaps even before. Computation of solutions are typically carried out with the pool-adjacent-violaters algorithm (PAVA), although various modern variants of quadratic programming could also be used.

[30] showed that the empirical risk of the nonparametric MLE, $\hat{\theta}_n$

$$R_n = n^{-1} \sum_{i=1}^n (\theta_i - \hat{\theta}_i)^2 \le C \left[\left(\frac{\sigma^2 V_n}{n} \right)^{2/3} + \frac{\sigma^2 \log n}{n} \right],$$

where $V_n = \theta_n - \theta_1$ and C is a fixed constant. However, more recent refinements establish that improvement over this $\mathcal{O}(n^{-1/3})$ rate can be achieved under the special circumstances that the θ_i are piecewise constant with a small number k of pieces. In that case it is proved in [7] that

$$R_n \le \inf_k \left(\frac{4\sigma^2(1+k)}{n} \log \frac{en}{1+k} \right).$$

Thus, up to the log factor we again have almost parametric convergence determined by the number of distinct piecewise constant elements in the target function. And again, it is worth stressing that adaptation is achieved over the number and locations of these pieces without any intervention of tuning parameters. When the monotonicity is misspecified, there is, obviously, a bias effect and this is also characterized in the general formulation of this result.

When the monotonicity constraint is replaced by a convexity (or concavity) constraint, the nonparametric MLE under Gaussian error is piecewise linear with knots at the observed design points. In the simplest setting with equally spaced design points this imposes the constraint that the second differences of θ_i's are nonnegative. [14] and [7] prove that when the target regression function is k-affine, that is, piecewise linear with k distinct pieces, the NPMLE again achieves an adaptive parametric rate of convergence up to a log factor.

Although the prior literature has focused exclusively on the least squares, i.i.d. Gaussian noise setting, as has most of the PAVA literature, there is nothing that prohibits us from entertaining other fidelity criteria. A natural alternative is the family of quantile loss functions that yield estimates of the conditional quantile functions of the response. Again, we have a convex optimization problem

$$\min \left\{ \sum_{i=1}^n \rho_\tau(y_i - g(x_i)) \mid g \in \mathcal{K} \right\}$$

, where $\rho_\tau(u) = u(\tau - I(u < 0))$, and $\mathcal{K}$ is the closed convex cone representing either monotone, convex or concave functions. An implementation of such estimators is available in the R package `quantreg` with the function `rqss`. In contrast to the least squares version of PAVA, the algorithmic complexity of the quantile implementation via interior point methods is not carefully analyzed, but sparsity of the underlying constraint matrix assures efficient practical performance. This implementation expands the formulation in several respects: (i) there is an option to impose further smoothness on the shape constrained estimate; (ii) general, unequally spaced design points are permitted; and (iii) additive models with several shape constrained components are permitted. To elaborate briefly on the first point, the general form of the `rqss` function permits the user to impose a total variation penalty on the first derivative of the fitted function

$$TV(g') = \int |g''(x)| dx$$

, controlled by a tuning parameter λ. When λ is sufficiently large, the $\hat{g}_n$ is constrained to be linear, while when λ is sufficiently close to zero, the TV penalty has no effect, and only the shape constraint determines the fit.

From a computational viewpoint the polyhedral cone and total variation constraints are especially appealing in the quantile regression setting because they maintain the linear programming structure of the estimation problem. Due to the relative sparsity of the design matrices in such problems, modern interior point algorithms are quite efficient even for large scale problems. It should be noted that the form of the solutions, that is whether they are piecewise constant, piecewise linear, etc., is entirely determined by the form of the constraints and, in particular, by the order of the differential operator appearing there. Thus, if $g \in \mathcal{K}$ requires that $Dg \geq 0$ to impose monotonicity, then solutions will be piecewise constant. If instead $D^2 g \geq 0$ is imposed to achieve convexity, then solutions will be piecewise linear. Likewise, total variation penalties on g, itself, yield piecewise constant solutions, while total variation penalties on Dg, thereby controlling the L_1 norm of $D^2 g$ yield piecewise linear solutions. Although such methods have a long history in imaging and actuarial science, they only have become widely appreciated in statistics through the relatively recent works of [18] and [28].

We conjecture that these shape constrained conditional quantile function estimators enjoy the same almost parametric convergence as their least squares counterparts and hope to report on this at a later time.

6. Shape Constrained Transformation Models

This brings us to our final category of shape constrained estimators: transformation models take a variety of forms, but typically they have a single index structure like

$$\mathbb{E}Y_i|X_i = \Psi(X_i^\top \beta). \tag{2}$$

The covariates and the parameter $\beta \in \mathbb{R}^p$ are wrapped in a function $\Psi : \mathbb{R} \to \mathbb{R}$ that may be parametric or nonparametric. Motivated by the revival of interest in the Grenander estimator, there has been an increased interest in transformation models with *monotonic* Ψ. Clearly, when $p = 1$, the β parameter is irrelevant and with Ψ monotonic, we are back to the methods described in the previous section. When $p > 1$, we can regard such models as an heroic attempt to circumvent the curse of dimensionality by assuming a simple form for the way the covariates enter the model while preserving some semblance of nonlinear structure. There are a variety of closely related models, some of which replace Ψ on the right-hand side by some transformation of the response variable itself. The monograph [6] provides a systematic treatment of many of these models, both parametric and nonparametric.

Recently, [2] and [12] have very thoroughly explored various approaches to estimating the model (2). They argue that it is preferable to avoid the direct profiling approach and focus on methods that find approximate zeros of the score (gradient) equations. It is clear that the vector β is identified only up to scale, so it is natural to impose the constraint that its Euclidean norm is one, $\|\beta\| = 1$. This can be accomplished in a variety of ways, either by transformation to spherical coordinates, or by adding a Lagrangian term. They employ the former scheme for their asymptotics, but prefer the latter from a practical, computational standpoint. Another option is to simply set one of coordinates of β equal to 1 or -1.

A drawback of the conditional mean formulation of the model, one that also afflicts a much broader class of nonlinear transformation models for conditional means, is the necessity of assuming that in the additive error formulation of the model

$$Y_i = \Psi(X_i^\top \beta) + U_i, \tag{3}$$

there is a full independence between the observed covariates X_i and U_i. One way to circumvent this requirement is to replace the mean formulation by a conditional quantile formulation

$$\mathbb{Q}_{Y|X}(\tau|X) = \Psi_\tau(X^\top \beta_\tau). \tag{4}$$

The quantile formulation also renders superfluous the unsightly moment conditions that appear inevitably in the analysis of the mean formulation. Such models were first considered by [8] who provide a very thorough motivation and contextualization for this class of models. Drawing on earlier work of Chaudhuri, they propose an average derivative estimator for β based on the nonparametric kernel weighted quantile regression.

When the τth conditional quantile function of Y given X is postulated to be a monotone function of a linear predictor in X_i, as seems plausible in many applications, we can try to exploit shape constrained methods to estimate both Ψ and β. Since quantiles are equivariant to monotone transformation, interpretation of the family of such models is also much more straightforward, than their mean counterparts. Our initial computational strategy arose immediately from the equivariance property of the quantiles, (4) implies

$$\mathbb{Q}_{\Psi_\tau^{-1}(Y)|X}(\tau|X) = X^\top \beta_\tau. \tag{5}$$

Since this linear quantile regression formulation can be efficiently estimated even for a high-dimensional β, a simple iterative strategy in which alternate back and forth from estimation of Ψ to estimation of β seems attractive. At each iteration we can modify the resulting β so that it has norm one. Given a β, an estimate of Ψ can be obtained by solving the monotone quantile regression problem described above. Both steps are linear programs. The biconvex structure of the problem is common to many mathematical contexts (see [1] and [9] for further details). Unfortunately, there is no general assurance that such an iterative procedure converges to a global optimum. Indeed, contrary to our initial, naive expectations, it performed abysmally.

Thus, following the lead of [12], but not without some trepidation, we turned to the global methods of optimization, in particular, the patterned search method of [15]. Convergence of such pattern search algorithms to a stationary point was established in [29]. An R implementation is available from the `optimx` package of [24], and an Rcpp implementation is available from the github site of Piet Groeneboom. Provisionally, we have experimented with the former implementation which has been performed quite well. In Figure 2, we illustrate three realizations from a sample in which the true Ψ is piecewise constant with only one jump; there are 5 covariates drawn as independent standard Gaussians. It is apparent from this Figure that the location and magnitude of the jump in Ψ is quite accurately estimated, this is hardly surprising in view of the fact that all the information about the linear predictor is contained in the neighborhood around this jump. Estimation of the level of Ψ before and after the jump is more problematic, which is again not surprising given that in the absence of a jump we would not be able to consistently estimate the linear index at all.

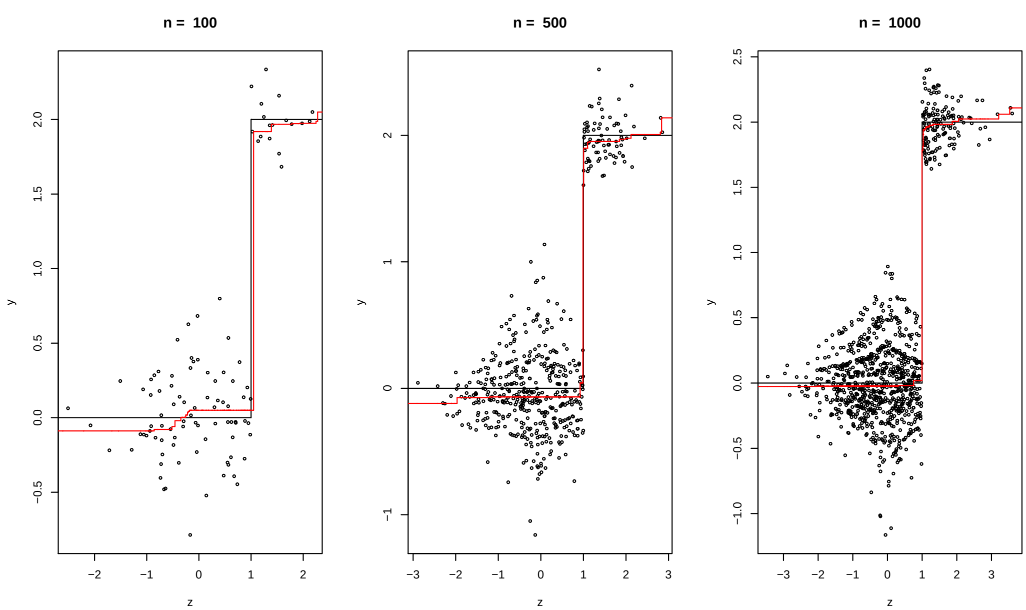


FIGURE 2. Single Index Median Regression Monotone Transformation Model: The plotted points depict the pairs $(\Psi_\tau(x_i^\top \beta_0), y_i)$. Three realizations for difference sample sizes of the final estimate of the monotone $\hat{\Psi}_n$ based on five Gaussian covariates in the single index. The true Ψ is depicted in black, while the estimate is in red.

To begin to explore the asymptotic behavior of our proposed estimator it is useful to reconsider the case for a known transformation, Ψ. As is described in [20], Section 4.4, if we adopt the model

$$\mathbb{Q}_{Y_i|X_i=x_i}(\tau|x_i) = g(x_i, \beta_0),$$

it is natural to try to estimate β_0 by

$$\hat{\beta}_n = \operatorname{argmin}_{b \in \mathcal{B}} \sum \rho_\tau(y_i - g(x_i, b)).$$

We emphasize the verb "try" since optimization need no longer be an assured attack on a convex problem with a unique solution. In keeping with the vast literature on nonlinear least squares, we will assume that the domain $\mathcal{B}$ is compact. In addition, we will assume that the conditional distribution functions F_i of $Y_i|X_i$ are absolutely continuous with continuous derivatives $f_i(\xi_i)$ at the points $\xi_i = g(x_i, \beta_0)$, and the following conditions on design.

G1: There exist constants k_0, k_1 and n_0 such that for $\beta_1, \beta_2 \in \mathcal{B}$ and $n > n_0$,

$$k_0 \parallel \beta_1 - \beta_2 \parallel \leq \left(n^{-1} \sum_{i=1}^{n} (g(x_i, \beta_1) - g(x_i, \beta_2))^2 \right)^{1/2} \leq k_1 \parallel \beta_1 - \beta_2 \parallel .$$

G2: There exist positive definite matrices D_0 and $D_1(\tau)$ such that with $\dot{g}_i = \partial g(x_i, \beta)/\partial \beta|_{\beta=\beta_0}$,

$$\begin{aligned} &\text{(i)} \quad \mathbb{E}\dot{g}_i\dot{g}_i^\top = D_0/ \\ &\text{(ii)} \quad \mathbb{E}f_i(\xi_i)\dot{g}_i\dot{g}_i^\top = D_1(\tau), \\ &\text{(iii)} \quad \max_{i=1,\ldots,n} \parallel \dot{g}_i \parallel / \sqrt{n} \to 0. \end{aligned}$$

Under these conditions, it can be shown that we have the Bahadur representation

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = D_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{g}_i \psi_\tau(u_i) + o_p(1),$$

where $\psi_\tau = \rho_\tau'$ and $u_i = y_i - g(x_i, \beta_0)$. Consequently,

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \rightsquigarrow \mathcal{N}(0, \tau(1-\tau) D_1^{-1} D_0 D_1^{-1}),$$

(for further details see [25]).

In the special case of the single index model, $g(x, \beta) = \Psi(x^\top \beta)$ and $\beta \in \mathcal{B}$ is replaced by $\beta \in \mathcal{S}^{p-1} \equiv \{b \in \mathbb{R}^p \mid \|b\| = 1\}$. Thus, $\dot{g} = \partial g/\partial \beta$ becomes $J\dot{\Psi}X$, where J denotes the Jacobian of the transformation that maps β into its $(p-1)$-dimensional counterpart. When Ψ is strictly increasing as is commonly assumed in the literature, this returns the expressions for D_0 and D_1 to something closely resembling their linear quantile regression equivalents, except for the weighting factors from the $\dot{\Psi}_i$ terms and the dimension reduction effect of the Jacobian terms. Inverses in the sandwich formulae now of course need to be interpreted as generalized inverses due to the dimension reduction.

At this point the obvious question is: How does all this change when Ψ is *estimated*? Surprisingly, the answer would seem to be: very little. Following the arguments of [2] and several prior authors cited there in the mean regression setting, this would entail replacing $XX^\top$ in the modified expressions for D_0 and D_1 by the conditional covariance $\mathrm{Cov}(X|X^\top\beta = x^\top\beta)$. This change reflects a reduction in the precision of the estimator $\hat{\beta}_n$. For smoothly increasing Ψ, as in the least squares theory, it is inevitable that we would obtain cube root convergence for $\hat{\Psi}_n$. A much more intriguing question, but a considerably more difficult one, is this: Can $\sqrt{n}$ convergence of $\hat{\Psi}_n$ be salvaged if we are willing to assume that the true Ψ is piecewise constant? The highly accurate estimates of the jump component of Ψ in Figure 2 offers a hint that this may indeed be plausible. Unfortunately, we must leave this intriguing problem for a future research.

References

1. R. J. Aumann, S. Hart, Bi-convexity and bi-martingales. *Israel J. Math.* **54** (1986), no. 2, 159–180.
2. F. Balabdaoui, P. Groeneboom, K. Hendrickx, Score estimation in the monotone single-index model. *Scand. J. Stat.* **46** (2019), no. 2, 517–544.

3. F. Balabdaoui, H. Jankowski, M. Pavlides, A. Seregin, J. Wellner, On the Grenander estimator at zero. *Statist. Sinica* **21** (2011), no. 2, 873–899.
4. L. Birgé, The Grenander estimator: a nonasymptotic approach. *Ann. Statist.* **17** (1989), no. 4, 1532–1549.
5. H. D. Brunk, Maximum likelihood estimates of monotone parameters. *Ann. Math. Statist.* **26** (1955), 607–616.
6. R. J. Carroll, D. Ruppert, *Transformation and Weighting in Regression.* Monographs on Statistics and Applied Probability. Chapman and Hall, New York, 1988.
7. S. Chatterjee, A. Guntuboyina, B. Sen, On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** (2015), no. 4, 1774–1800.
8. P. Chaudhuri, K. Doksum, A. Samarov, On average derivative quantile regression. *Ann. Statist.* **25** (1997), no. 2, 715–744.
9. J. Gorski, F. Pfeuffer, K. Klamroth, Biconvex sets and optimization with biconvex functions: a survey and extensions. *Math. Methods Oper. Res.* **66** (2007), no. 3, 373–407.
10. U. Grenander, On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** (1956), 125–153 (1957).
11. P. Groeneboom, *Estimating a Monotone Density.* Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, vol. II (Berkeley, Calif., 1983), 539–555, Wadsworth Statist./Probab. Ser., Wadsworth, Belmont, CA, 1985.
12. P. Groeneboom, K. Hendrickx, Estimation in monotone single-index models. *Stat. Neerl.* **73** (2019), no. 1, 78–99.
13. P. Groeneboom, G. Jongbloed, *Nonparametric Estimation Under Shape Constraints. Estimators, Algorithms and Asymptotics.* Cambridge Series in Statistical and Probabilistic Mathematics, 38. Cambridge University Press, New York, 2014.
14. A. Guntuboyina, B. Sen, Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** (2015), no. 1-2, 379–411.
15. R. Hooke, T. A. Jeeves, Direct search solution of numerical and statistical problems. *Journal of the ACM* **8** (1961), 212–229.
16. J. Kiefer, J. Wolfowitz, Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** (1956), 887–906.
17. A. K. H. Kim, A. Guntuboyina, R. J. Samworth, Adaptation in log-concave density estimation. *Ann. Statist.* **46** (2018), no. 5, 2279–2306.
18. S.-J. Kim, K. Koh, S. Boyd, D. Gorinevsky, l_1 trend filtering. *SIAM Rev.* **51** (2009), no. 2, 339–360.
19. A. K. H. Kim, R. J. Samworth, Global rates of convergence in log-concave density estimation. *Ann. Statist.* **44** (2016), no. 6, 2756–2779.
20. R. Koenker, *Quantile Regression.* Econometric Society Monographs, 38. Cambridge University Press, Cambridge, 2005.
21. R. Koenker, I. Mizera, Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *J. Amer. Statist. Assoc.* **109** (2014), no. 506, 674–685.
22. R. Koenker, I. Mizera, Shape constrained density estimation via penalized Rnyi divergence. *Statist. Sci.* **33** (2018), no. 4, 510–526.
23. N. Laird, Nonparametric maximum likelihood estimation of a mixed distribution. *J. Amer. Statist. Assoc.* **73** (1978), no. 364, 805–811.
24. J. C. Nash, R. Varadhan, Unifying Optimization Algorithms to Aid Software System Users: optimx for R. *Journal of Statistical Software* **43** (2011), 1–14.
25. W. Oberhofer, H. Haupt, Asymptotic theory for nonlinear quantile regression under weak dependence. *Econometric Theory* **32** (2016), no. 3, 686–713.
26. B. L. S. Prakasa Rao, Estimation of a unimodal density. *Sankhyā Ser. A* **31** (1969), 23–36.
27. S. Saha, A. Guntuboyina, On the nonparametric maximum likelihood estimator for Gaussian location mixture densities with application to Gaussian denoising. *Ann. Statist.* **48** (2020), no. 2, 738–762.
28. R. J. Tibshirani, Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.* **42** (2014), no. 1, 285–323.
29. V. J. Torczon, On the convergence of pattern search algorithms. *SIAM J. Optim.* **7** (1997), no. 1, 1–25.
30. C.-H. Zhang, Risk bounds in isotonic regression. *Ann. Statist.* **30** (2002), no. 2, 528–555.

(Received 19.11.2019)

[1]Roger Koenker, Department of Economics, UCL, London, WC1H OAX, UK
E-mail address: r.koenker@ucl.ac.uk

[2]Jiaying Gu, Department of Economics, University of Toronto, Ontario, M5S 3G7, Canada
E-mail address: jiaying.gu@utoronto.ca

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 199–205

LIMITING DISTRIBUTION OF A SEQUENCE OF FUNCTIONS DEFINED ON A MARKOV CHAIN

ZURAB KVATADZE[1] AND TSIALA KVATADZE[2]

Abstract. The present article shows the limiting distribution of partial sums of a functional sequence defined on a Markov Chain in case the chain is ergodic, with one class of ergodicity and contains cyclical subclasses.

Limiting behavior of sums of random variables is a classical problem in the probability theory, which is intensely studied by contemporaneous researchers both for independent variables and for the case of certain relationships between the terms of sequences. There exists a rich theory of sums of independent random variables (see, e.g., [5, 12, 13]). The problem of extending this case to the sums of dependent random variables introduces naturally the Markovian dependence, which in turn represents particular type of a weak dependency. The limiting theorems by Rosenblatt, Ibragimov and others concerning weakly dependent sequences are usually stated in terms of σ-algebras generated by asymptotically separable intervals of the sequence. The process of their investigation involves the so-called S. Bernstein's "sectioning" method based on the weakening effect taking place during separation of groups of dependent variables (see [6]). Contemporaneous situation in the theory of sums of dependent random variables is expressed by using limiting theorems for martingales and semi-martingales (see [7]).

Different authors considered sums of random variables, whose joint distribution is determined by the controlling sequence of random variables (see [2, 3, 9]). An important part of these comprise problems regarding the sums of variables is defined directly on a chain (see [1, 3, 8, 11]). This paper considers the limiting theorem for functions defined on a stationary, finite, ergodic Markov chain.

We consider stationary, homogeneous, finite $\{\xi_i\}_{i\geq 1}$ ergodic Marcov chain with one class of ergodicity (might containing cyclic subclasses) defined on a probability space (Ω, F, P). The chains have a set of states $\Xi = \{b_1, b_2, \dots, b_r\}$, a matrix of transient probabilities $P = \|P_{\alpha\beta}\|_{\alpha,\beta=\overline{1,r}}$ and a vector of limiting distribution of stationary probabilities $\pi = (\pi_1, \pi_2, \dots, \pi_r)$ representing a solution of the following matrix equation:

$$\pi = \pi P.$$

We suggest that the initial distribution is stationary, the distribution

$$P(\xi_1 = b_\alpha) = \pi_\alpha, \quad \alpha = \overline{1, r},$$

is based on stationarity means and the chain has the same distribution for each step

$$P(\xi_n = b_\alpha) = \pi_\alpha, \quad \alpha = \overline{1, r}, \quad n = 1, 2, \dots .$$

Next, we introduce the Cezaro definition for convergence of the sequence and, relying on that definition, we establish all types of convergence when the chain has cyclical subclasses.

The sequence $\{t_n\}_{n\geq 1}$ is Cezaro convergent to t, and we write

$$(\lim_{n\to\infty} t_n)_c = t,$$

if the means of the first n terms of the sequence converge to t:

$$\lim_{n\to\infty} T_n = t,$$

2020 *Mathematics Subject Classification.* 62G05, 62G07.

Key words and phrases. Markov Chain; Limiting distribution; Class of ergodicity; Cyclical subclasses.

where $T_n = \frac{1}{n} \sum\limits_{i=1}^{n-1} t_i$.

Cezaro convergence may be considered upon analyzing the convergence of series.

The series $\sum\limits_{k=1}^{\infty} a_k$ is said to be Cezaro convergent and the sum be equal to a, if a is the Cezaro limit of the sequence of the partial sum $S_n = \sum\limits_{k=1}^{n} a_k$,

$$\lim_{n\to\infty} (S_n)_c = a,$$

which implies that there exists the limit of the sequence $\widetilde{a_n}$,

$$\lim_{n\to\infty} \widetilde{a_n} = a,$$

where $\widetilde{a_n} = \frac{1}{n} \sum\limits_{k=0}^{n} (n-k)a_k$ and, at the same time, this a represents the Cezaro sum of the series under consideration which can be written as

$$\left(\sum_{k=1}^{n} a_k \right)_c = a.$$

We denote by Π the limit (see [8])

$$\lim_{n\to\infty} (P^n)_c = \Pi = \begin{pmatrix} \pi_1, & \pi_2, & \dots, & \pi_r \\ \pi_1, & \pi_2, & \dots, & \pi_r \\ \dots & \dots & \dots & \dots \\ \pi_1, & \pi_2, & \dots, & \pi_r. \end{pmatrix}$$

It is obvious that

$$\Pi = \|\pi_{\alpha\beta}\|_{\alpha,\beta=\overline{1,r}}; \quad \pi_{\alpha\beta} = \pi_\beta; \quad \alpha,\beta = \overline{1,r},$$
$$\lim_{n\to\infty} \left(p^n_{\alpha\beta}\right)_c = \pi_\beta, \quad \alpha,\beta = \overline{1,r}.$$

Let the fundamental matrix of the chain be

$$Z = \|z_{\alpha\beta}\|_{\alpha,\beta=\overline{1,r}},$$
$$Z = [I - (P - \Pi)]^{-1} = I + \left(\sum_{j=1}^{\infty} (P^j - \Pi) \right)_c = \|z_{\alpha\beta}\|_{\alpha,\beta=\overline{1,r}},$$

where I is the identity matrix of $r \times r$ dimensions. For the regular chain, the convergence of series is implied to be a standard convergence.

Let us consider a vector function defined on the Ξ space

$$f(\xi_i) : \Xi \to R^k,$$
$$f(\xi_i) = (f_1(\xi_i), f_2(\xi_i), \dots, f_k(\xi_i))$$

and introduce the notation:

$$f(b_\alpha) = f(\alpha) = (f_1(\alpha), f_2(\alpha), \dots, f_k(\alpha)), \quad \alpha = \overline{1,r},$$
$$f_i(\alpha) = f_i(b_\alpha), \quad i = \overline{1,k}, \quad \alpha = \overline{1,r}.$$

Theorem 1. *When $\{\xi_i\}_{i\geq 1}$ is the above-mentioned Markov chain and $f = (f_1, f_2, \dots, f_k)$ is the k-dimensional vector function from Ξ to R^k, then if the limiting covariance matrix of the sum is*

$$U_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [f(\xi_j) - Ef(\xi_j)]$$
$$T_f = \left\| t_{f_{i,j}} \right\|_{i,j=\overline{1,k}},$$
$$t_{f_{i,j}} = \sum_{\alpha,\beta}^{r} (\pi_\alpha z_{\alpha\beta} + \pi_\beta z_{\beta\alpha} - \pi_\alpha \pi_\beta - \pi_\alpha \delta_{\alpha\beta}) f_i(\alpha) f_i(\beta) \quad i,j = \overline{1,k} \tag{1}$$

(*where* $\delta_{\alpha\beta}$ *is the Kronecker symbol*) *is positively defined, as* $n \to \infty$, *there is a convergence*

$$P_{U_n} \xrightarrow{W} \Phi_{T_f}.$$

The case for $k = 1$, when $\lim_{n\to\infty} D(U_n) > 0$ (where $D(\cdot)$ denotes variance), is a famous fact (see [4,10]) (when $\lim_{n\to\infty} D(U_n) = 0$, then U_n converges to zero in probability) and T_f can be written explicitly as a sum of components of the chain (see [8])

$$t = \lim_{n\to\infty} D(U_n) = \sum_{\alpha,\beta=1}^{r} (\pi_\alpha z_{\alpha\beta} + \pi_\beta z_{\beta\alpha} - \pi_\alpha \pi_\beta - \pi_\alpha \delta_{\alpha\beta}) f(\alpha) f(\beta).$$

Proof. Using the Kramer-Wold method, we can derive the multidimensional case. Using the chain characteristic, we derive a matrix representation of the matrix T_f. Let us introduce a $k \times r$ matrix F,

$$F = \begin{pmatrix} f_1(b_1), & f_1(b_2), & \dots, & f_1(b_r) \\ f_2(b_1), & f_2(b_2), & \dots, & f_2(b_r) \\ \dots & \dots & \dots & \dots \\ f_k(b_1), & f_k(b_2), & \dots, & f_k(b_r) \end{pmatrix} = \begin{pmatrix} f_1(1), & f_1(2), & \dots, & f_1(r) \\ f_2(1), & f_2(2), & \dots, & f_2(r) \\ \dots & \dots & \dots & \dots \\ f_k(1), & f_k(2), & \dots, & f_k(r) \end{pmatrix}$$

$$= \|f_{ij}\|_{\substack{i=\overline{1,r} \\ j=\overline{1,r}}}, \quad f_{ij} = f_i(b_j)$$

and denote

$$\begin{aligned} V_0 =&\mathrm{cov}[f(\xi_1)] = E\{[f(\xi_1) - Ef(\xi_1)]^T [f(\xi_1) - Ef(\xi_1)]\}, \\ V_j =&E\{[f(\xi_1) - Ef(\xi_1)]^T [f(\xi_{1+j}) - Ef(\xi_{1+j})]\}, \quad j > 0, \\ V_{-j} =&E\{[f(\xi_{1+j}) - Ef(\xi_{1+j})]^T [f(\xi_1) - Ef(\xi_1)]\}, \quad j > 0. \end{aligned}$$

Based on the stationarity of the sequence $\{\xi_i\}_{i\geq 1}$, as $n \to \infty$, we have

$$E[U_n^T, U_n] = \frac{1}{n}\left[nV_0 + \sum_{j=1}^{n-1}(n-j)(V_j + V_{-j})\right]$$

$$= V_0 + \frac{1}{n}\sum_{j=1}^{n}(n-j)V_j + \frac{1}{n}\sum_{j=1}^{n}(n-j)V_{-j} \xrightarrow{n\to\infty} V_0 + \left(\sum_{j=1}^{\infty} V_j\right)_c + \left(\sum_{j=1}^{\infty} V_{-j}\right)_c, \tag{2}$$

where $()_c$ denotes the Cezaro convergence of the sum in the parenthesis. It is obvious that if the chain is regular, this convergence is equivalent to the standard case of convergence of partial sums.

Thus, T_f represents the limiting covariance of the sum U_n and we have

$$T_f = V_0 + \left(\sum_{j=1}^{\infty} V_j\right)_c + \left(\sum_{j=1}^{\infty} V_{-j}\right)_c. \tag{3}$$

In the right-hand side, the convergence of matrix series is equivalent to that of a regular chain by virtue of a common definition of the convergence.

We now express V_0 and V_j matrices based on the components of the chain

$$Ef(\xi_1) = \sum_{\alpha=1}^{r} \pi_\alpha f(\alpha) = (\sum_{\alpha=1}^{r} \pi_\alpha f_1(\alpha), \dots, \sum_{\alpha=1}^{r} \pi_\alpha f_k(\alpha))$$

$$= (\pi_1, \pi_2, \dots, \pi_r) \begin{pmatrix} f_1(1), & f_2(1), & \dots, & f_k(1) \\ f_1(2), & f_2(2), & \dots, & f_k(2) \\ \dots & \dots & \dots & \dots \\ f_1(r), & f_2(r), & \dots, & f_k(r) \end{pmatrix} = \pi F^T;$$

$$E\{f(\xi_1)^T f(\xi_1)\} = E\left\{\begin{pmatrix} f_1(\xi_1) \\ f_2(\xi_1) \\ \vdots \\ f_k(\xi_1) \end{pmatrix} (f_1(\xi_1), f_2(\xi_1), \dots, f_k(\xi_1))\right\}$$

$$= \|Ef_i(\xi_1)f_j(\xi_1)\|_{i,j=\overline{1,k}} = \left\|\sum_{\alpha=1}^{r} \pi_\alpha f_i(\alpha) f_j(\alpha)\right\|_{i,j=\overline{1,k}}$$

$$= \begin{pmatrix} \sum\limits_{\alpha=1}^{r} \pi_\alpha f_1(\alpha) f_1(\alpha), & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_1(\alpha) f_2(\alpha), & \dots, & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_1(\alpha) f_k(\alpha) \\ \sum\limits_{\alpha=1}^{r} \pi_\alpha f_2(\alpha) f_1(\alpha), & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_2(\alpha) f_2(\alpha), & \dots, & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_2(\alpha) f_k(\alpha) \\ \dots & \dots & \dots & \dots \\ \sum\limits_{\alpha=1}^{r} \pi_\alpha f_k(\alpha) f_1(\alpha), & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_k(\alpha) f_2(\alpha), & \dots, & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_k(\alpha) f_k(\alpha) \end{pmatrix} = F\Pi_{dg}F^T;$$

$$E\{f(\xi_1)^T f(\xi_{1+j})\} = \|Ef_i(\xi_1)f_s(\xi_{1+j})\|_{i,s=\overline{1,k}}$$

$$= \begin{pmatrix} \sum\limits_{\alpha=1}^{r} \pi_\alpha f_1(\alpha) \sum\limits_{\beta=1}^{r} P_{\alpha\beta}^j f_1(\beta), & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_1(\alpha) \sum\limits_{\beta=1}^{r} P_{\alpha\beta}^j f_2(\beta), & \dots, & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_1(\alpha) \sum\limits_{\beta=1}^{r} P_{\alpha\beta}^j f_k(\beta) \\ \sum\limits_{\alpha=1}^{r} \pi_\alpha f_2(\alpha) \sum\limits_{\beta=1}^{r} P_{\alpha\beta}^j f_1(\beta), & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_2(\alpha) \sum\limits_{\beta=1}^{r} P_{\alpha\beta}^j f_2(\beta), & \dots, & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_2(\alpha) \sum\limits_{\beta=1}^{r} P_{\alpha\beta}^j f_k(\beta) \\ \dots & \dots & \dots & \dots \\ \sum\limits_{\alpha=1}^{r} \pi_\alpha f_k(\alpha) \sum\limits_{\beta=1}^{r} P_{\alpha\beta}^j f_1(\beta), & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_k(\alpha) \sum\limits_{\beta=1}^{r} P_{\alpha\beta}^j f_2(\beta), & \cdots, & \sum\limits_{\alpha=1}^{r} \pi_\alpha f_k(\alpha) \sum\limits_{\beta=1}^{r} P_{\alpha\beta}^j f_k(\beta) \end{pmatrix}$$

$$= \left\|\sum_{\alpha,\beta=1}^{r} \pi_\alpha f_i(\alpha) P_{\alpha\beta}^j f_s(\beta)\right\|_{i,s=\overline{1,k}} = F\Pi_{dg}P^jF^T,$$

where $(\cdot)_{dg}$ denotes the matrix obtained by replacing each element of the matrix in the parenthesis by zero, except ones located on the main diagonal.

The following equality

$$\pi^T\pi = \Pi_{dg}\Pi$$

holds and the derived equations will be taken into account in the expression for V_j. When $j = 0$, we obtain

$$\begin{aligned} V_0 &= E\left\{f^T(\xi_1)f(\xi_1)\right\} - E\left\{f^T(\xi_1)\right\}E\left\{f(\xi_1)\right\} \\ &= F\Pi_{dg}F^T - (\pi F^T)^T\pi F^T = F\Pi_{dg}F^T - F\pi^T\pi F^T \\ &= F\Pi_{dg}F^T - F\Pi_{dg}\Pi F^T = F(\Pi_{dg} - \Pi_{dg}\Pi)F^T. \end{aligned}$$

By the stationarity $E\left\{f(\xi_{1+j})\right\} = E\left\{f(\xi_1)\right\}$, when $j > 0$, the equalities

$$\begin{aligned} V_j &= E\left\{f^T(\xi_1)f(\xi_{1+j})\right\} - E\left\{f^T(\xi_1)\right\}E\left\{f(\xi_{1+j})\right\} \\ &= E\left\{f^T(\xi_1)f(\xi_{1+j})\right\} - E\left\{f^T(\xi_1)\right\}E\left\{f(\xi_1)\right\} \\ &= F\Pi_{dg}P^jF^T - (\pi F^T)^T\pi F^T = F\Pi_{dg}P^jF^T - F\pi^T\pi F^T \\ &= F\Pi_{dg}P^jF^T - F\Pi_{dg}\Pi F^T = F\Pi_{dg}(P^j - \Pi)F^T \end{aligned}$$

are true.

Thus, the sum in the right-hand side of (2) can be expressed as

$$\left(\sum_{j=1}^{\infty} V_j\right)_c = \left(\sum_{j=1}^{\infty} F\Pi_{dg}(P^j - \Pi)F^T\right)_c = F\Pi_{dg}\left(\sum_{j=1}^{\infty}(P^j - \Pi)\right)_c F^T.$$

By the property of the fundamental matrix, we have

$$\left(\sum_{j=1}^{\infty}(P^j - \Pi)\right)_c = Z - I.$$

Thus we get the following equation:

$$\left(\sum_{j=1}^{\infty} V_j\right)_c = F\Pi_{dg}(Z-I)F^T = F(\Pi_{dg}Z - \Pi_{dg})F^T. \tag{4}$$

Like equation (4), the following sum can be computed by using stationarity of the chain

$$\begin{aligned}\left(\sum_{j=1}^{\infty} V_{-j}\right)_c &= \left(\sum_{j=1}^{\infty} E\{[f(\xi_{1+j}) - Ef(\xi_{1+j})]^T[f(\xi_1) - Ef(\xi_1)]\}\right)_c \\ &= \left(\sum_{j=1}^{\infty} E(\{[f(\xi_1) - Ef(\xi_1)]^T[f(\xi_{1+j}) - Ef(\xi_{1+j})]\})^T\right)_c \\ = \left[\left(\sum_{j=1}^{\infty} E\{[f(\xi_1) - Ef(\xi_1)]^T[f(\xi_{1+j}) - Ef(\xi_{1+j})]\}\right)_c\right]^T &= \left[\left(\sum_{j=1}^{\infty} V_j\right)_c\right]^T \\ &= \left[F\left(\Pi_{dg}Z - \Pi_{dg}\right)_c F^T\right]^T = F((\Pi_{dg}Z)^T - \Pi_{dg})_c F^T.\end{aligned}$$

Substituting the obtained results into (3) and using characteristic matrices corresponding to the chain, we get the following matrix expression for Tf,

$$T_f = F[\Pi_{dg}Z + (\Pi_{dg}Z)^T - \Pi_{dg}\Pi - \Pi_{dg}]F^T.$$

Obviously, the $t_{f_{i,j}}$ elements of the matrix T_f can be expressed by virtue of (1).

Next, we introduce a characteristic of time moments quantity elapsed by the chain at the first n steps in different b_α, $\alpha = \overline{1,r}$ positions.

Let $\nu_n(\alpha) = \nu_n(b_\alpha)$, $(\alpha = \overline{1,r})$ be a random variable representing the amount of time intervals during the first n steps when the chain is in position b_α, $(\alpha = \overline{1,r})$ on a fixed trajectory $\bar{\xi}_{1n} = (\xi_1, \xi_2, \dots, \xi_n)$. Then it is obvious that the equation

$$\nu_n(1) + \nu_n(2) + \cdots + \nu_n(r) = n$$

holds.

The quantity $\frac{\nu_n(\alpha)}{n}$ is a part of time n during which the chain at the first n steps spends in condition b_α. $\square$

Theorem 2. *The $\nu_n(\alpha)$, $(\alpha = \overline{1,r})$, random variable is measurable with respect to the sigma algebra induced by dividing the Ω space during fixation of a $\bar{\xi}_{1n}$ trajectory.*

Proof. We show that a discrete random variable $\nu_n(\alpha)$ attains constant values on sets generated by partitioning the Ω space during fixation of a $\bar{\xi}_{1n}$ trajectory.

Conditions set of the chain is $\Xi = \{b_1, b_2, \dots, b_r\}$. On a fixed $\bar{\xi}_{1n}$ trajectory, possible values will be the Cartesian product $\Xi^n = \Xi \times \Xi \times \cdots \times \Xi$. Let us show how the Ω space will be partitioned.

Introduce the following sets:

$$D_{1,2,\dots,n,m_1,m_2,\dots,m_n = \left\{\omega \middle| \xi_1 = b_{m_1}, \xi_2 = b_{m_2}, \dots, \xi_n = b_{m_n}\right\}}, \quad b_{m_i} \in \Xi, \quad i = \overline{1,r}.$$

Fixation of a $\bar{\xi}_{1n}$ trajectory will result in a partition of the Ω space,

$$\bar{D} = \{D_{1,2,\dots,n,m_1,m_2,\dots,m_n} | m_i \in \{1, 2, \dots, r\}\}.$$

It is clear that

$$D_{ik} = \{\omega\,|\xi_i = b_k\} = \sum_{\substack{m_\alpha \in \Xi\setminus\{b_i\} \\ \alpha \neq i}} \{D_{1,2,\dots,n,m_1,m_2,\dots,m_{i-1},k,m_{i+1},\dots,m_n}\},$$

$$\xi_i = \sum_{k=1}^{r} b_k I_{(D_{ik})}.$$

To derive analytical expression for the sum $\nu_n(i)$, consider the sets

$$A^i_{n,j_1,j_2,\dots,j_k} = \left\{\omega \left| \begin{array}{l} j_1 < j_2 < \cdots < j_k \\ \xi_\alpha = b_i \quad \alpha \in \{j_1,\dots,j_k\} \\ \xi_\alpha \in \Xi\setminus\{b_i\} \quad \alpha \notin \{j_1,\dots,j_k\}, \alpha = \overline{1,n} \end{array}\right.\right\}$$

$$= \sum_{\substack{j_1 < j_2 < \cdots < j_k \\ m_{j_1} = m_{j_2} = \cdots = m_{j_k} = b_i \\ m_{j_\alpha} \in \Xi\setminus\{b_i\} \quad \alpha = \overline{k+1.n}}} D_{j_1,\dots,j_n,m_{j_1},\dots,m_{j_n}};$$

$$A^i_{n,k} = \{\nu_n(b_i) = k\} = \sum_{\substack{j_1 < j_2 < \cdots < j_k \\ \{j_1,j_2,\dots,j_k\} \subset \{1,2,\dots,n\}}} A^i_{n,j_1,\dots,j_k}$$

$$= \sum_{\substack{j_1 < j_2 < \cdots < j_k \\ \{j_1,j_2,\dots,j_k\} \subset \{1,2,\dots,n\}}} \sum_{\substack{j_1 < j_2 < \cdots < j_k \\ m_{j_1} = m_{j_2} = \cdots = m_{j_k} = b_i \\ m_{j_\alpha} \in \Xi\setminus\{b_i\}, \quad \alpha = \overline{k+1,n}}} D_{j_1,\dots,j_n,m_{j_1},\dots,m_{j_n}}.$$

Clearly, the $A^i_{n,j_1,j_2,\dots,j_k}$ type sets are $(r-1)^{n-k}$ in total, while there are $C_n^k \cdot (r-1)^{n-k}$ $A^i_{n,k}$ type sets. $\square$

Relying on the above-said, we easily find that

$$\nu_n(i) = \nu_n(b_i) = \sum_{k=0}^{n} k I_{A^i_{n,k}}.$$

Thus, the measurability of a $\nu_n(i)$ random variable with respect to partition $\bar{D}$ is shown. Clearly, this implies that the variable is measurable with respect to the sigma algebra generated by that partition. Note finally that any function $f(\nu_n(i))$ is also measurable, where $f(\circ)$ is a continuous function.

References

1. G. YU. Aleshkyavichus, On the central limit theorem for sums of random variables given on a Markov chain. (Russian) *Lithuanian Mathematical Collected Works, Vilnius* **6** (1966), no. 1, 15–22.
2. I. V. Bokuchava, Limit theorems for conditionally independent sequences. (Russian) *Theory of Probability and its Application. Moskva* **29** (1984) no. 1, 192–193.
3. I. V. Bokuchava, Z. Kvatadze, T. Shervashidze, On limit theorems for random vectors controlled by a Markov chain. *Probability theory and mathematical statistics, Vilnius* **1** (1985), 231–250, *VNU Sci. Press, Utrecht*, 1987.
4. J. L. Doob, *Stochastic Processes.* Wiley Classics Library. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1990.
5. B. V. Gnedenko, A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables.* Gostekhizdat, Moscow-Leningrad, 1949.
6. I. A. Ibragimov, YU. V. Linnik, *Independent and Stationarily Connected Variables.* (Russian) Nauka, Moscow, 1965.
7. J. Jakod, A. N. Shiryaev, *Limit Theorems for Stochastic Processes.* (Russian) Nauka, Moscow, 1994.
8. J. G. Kemeny, J. L. Snell, *Finite Markov chains.* The university series in undergraduate mathematics. Springer-Verlag, 1963.

9. R. Z. Khas'minskiĭ, On limiting distributions for sums of conditionally independent random variables. (Russian) *Teor. Verojatnost. i Primenen.* **6** (1961), 119–125.
10. S. V. Nagaev, Some limit theorems for stationary Markov chains. (Russian) *Teor. Veroyatnost. i Primenen.* **2** (1957), 389–416.
11. G. L. O'Brien, Limit theorems for sums of chain-dependent processes. *J. Appl. Probability* **11** (1974), 582–587.
12. V. V. Petrov, *Sums of Independent Random Variables.* (Russian) Nauka, Moscow, 1972.
13. V. M. Zolotariov, *Contemporary Theory of Summation of Independent Stochastic Variables.* (Russian) Nauka, Moscow, 1962.

(Received 20.12.2019)

[1]Department of Mathematics, Georgian Technical University, 77 Kostava Str., Tbilisi 0175, Georgia
E-mail address: zurakvatadze@yahoo.com

[2]Faculty of Informatics and Management Systems, Georgian Technical University, 77 Kostava Str., Tbilisi 0175, Georgia
E-mail address: ttkvatadze@gmail.com

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 207–216

BANACH SPACE VALUED FUNCTIONALS OF THE WIENER PROCESS

BADRI MAMPORIA[1] AND OMAR PURTUKHIA[2]

Abstract. The problem of representation of the Banach space-valued functionals of the one- dimensional Wiener process by the Ito stochastic integral is considered. Earlier, in [5] we have developed this problem in case the joint distribution of the Wiener process and its functional is Gaussian. In this article we consider the general case: firstly, for the weak second order Banach space-valued functional the generalized random process is found as an integrand. Further, for the one-dimensional functional of the Wiener process the sequence of adapted step functions converging to the integrand function, generalizing the corresponding result for the Gaussian case, is obtained (see [2]); the sequence of adapted step functions of generalized random elements converging to the integrand generalized random process is constructed for a Banach space-valued functional.

In developing the Ito stochastic analysis in a Banach space the main goal of the problem is to construct the stochastic integral in an arbitrary separable Banach space. This problem is considered in the following cases: (a) the integrand adapted to the σ-algebra generated by the Wiener process is Banach space-valued and the stochastic integral is constructed by the one-dimensional Wiener process; (b) the integrand adaptive process is operator-valued (from the Banach space to the Banach space), and the stochastic integral is constructed by the Wiener process in a Banach space; (c) the integrand adaptive process is operator-valued (from the Hilbert space to the Banach space), and the stochastic integral is constructed by the cylindrical Wiener process in a Hilbert space. In all the above-mentioned cases the main difficulties are the same. Therefore, to realize simply all these difficulties, in the previous article [5] and here we consider the first case (the Wiener process is one-dimensional).

Using traditional methods, it becomes possible to find the suitable conditions that guarantee the construction of the Ito stochastic integral in a Banach space only in a very narrow class of Banach spaces. This class is called the class of UMD Banach spaces (for survey, see [8]). In our approach, the generalized stochastic integral for a wide class of adapted Banach space-valued random processes is constructed and the problem of the existence of the stochastic integral is reduced to the problem of decomposability of the generalized random element (cylindrical random element, or random linear function) (see [4]).

In this article we consider the problem of representation of the functional of the Wiener process by the stochastic integral in an arbitrary separable Banach space. This problem is, in some sense, opposite to the problem of the existence of the stochastic integral: in this case we have the Banach space-valued random element and the problem of finding the integrand Banach space-valued adapted process whose stochastic integral is this random element. In [5], we considered this problem in the case, where the joint distribution of the Wiener process and its functional is Gaussian. In [3], this problem is considered for the case of UMD Banach space, where under special conditions the Wiener functional is represented by the stochastic integral and the Clark–Ocone formula of representation of the functional of the Wiener process is generalized.

Let X be a real separable Banach space, X^* be its conjugate, and (Ω, B, P) a probability space.

Remember that the continuous linear operator $T : X^* \to L_2(\Omega, B, P)$ is called the generalized random element (GRE) Denote by $\mathcal{M}_1 := L(X^*, L_2(\Omega, B, P)$ the Banach space of GRE with the norm

$$\|T\|^2 = \sup_{\|x^*\| \leq 1} E(Tx^*)^2.$$

2020 *Mathematics Subject Classification.* 60B11, 60H5, 60H10, 37L55.

Key words and phrases. Wiener process; Functional of the Wiener process; Ito stochastic integral; Generalized random element; Covariance operator of the (generalized) random element in a Banach space.

We can realize the weak second order random element ξ as an element of $\mathcal{M}_1$, $T_\xi x^* = \langle \xi, x^* \rangle$ (the boundedness of this operator follows by the closed graph theorem), but not conversely: in an infinite-dimensional Banach space, for any $T : X^* \to L_2(\Omega, B, P)$, there does not always exist the random element $\xi : \Omega \to X$ such that $Tx^* = \langle \xi, x^* \rangle$ for all $x^* \in X^*$. The problem of the existence of such random element is well known as the problem of decomposability (radonizability) of the GRE. Denote by $\mathcal{M}_2$ the linear normed space of all random elements of weak second order with the norm

$$\|\xi\|^2 = \sup_{\|x^*\| \le 1} E\langle \xi, x^* \rangle^2.$$

Thus we have $\mathcal{M}_2 \subset \mathcal{M}_1$.

The family of linear operators $(T_t)_{t \in [0,1]}$ is called the weak second order generalized random process (GRP) if $T_t x^*$ is $B([0,1]) \times B(\Omega)$ measurable and

$$\|T_t\|^2 \equiv \sup_{\|x^*\| \le 1} \int_0^1 E(T_t x^*)^2 dt < \infty.$$

Denote by $\mathcal{M}_1^{(\lambda,P)}$ the Banach space of such GRP.

The Banach space-valued stochastic process $f(t, \omega)$, $t \in [0,1]$ is called a weak second order random process, if for all $x^* \in X^*$,

$$\int_0^1 E\langle f(t,\omega), x^* \rangle^2 dt < \infty.$$

The weak second order random process realizes the GRP $T_f : X^* \to L_2([0,1] \times \Omega)$: $T_f x^* = \langle f(t, \omega), x^* \rangle$.

Denote by $\mathcal{M}_2^{(\lambda,P)}$ the normed linear spaces of $f(t, \omega)$, $t \in [0,1]$, with the norm

$$\sup_{\|x^*\| \le 1} \left(\int_0^1 E\langle f(t,\omega), x^* \rangle^2 dt \right)^{\frac{1}{2}}.$$

We have $\mathcal{M}_2^{(\lambda,P)} \subset \mathcal{M}_1^{(\lambda,P)}$.

Let $(W_t)_{t \in [0,1]}$ be a real-valued Wiener process. Denote by F_t^W the σ-algebra generated by the random variables $(W_s)_{s \le t}$ ($F_t^W = \sigma(W_s, s \le t)$), which are completed by P-null sets. Suppose that ξ is F_1^W-measurable weak second order random element i.e., ξ is the functional of the Wiener process. Our main aim is to represent the random element ξ by the Ito stochastic integral

$$\xi = E\xi + \int_0^1 f(t, \omega) dW_t,$$

where $f(t, \omega)$ is the Banach space-valued F_t^W-adapted random process, but this is, in general, impossible. We have the following positive result: For all weak second order Wiener functional we always have integrand as a GRP, that is, an element of the Banach space $\mathcal{M}_1^{(\lambda,P)}$. In developing this problem, we considered firstly in [5] the case, where ξ is a Gaussian random element which together with the Wiener process generates the mutually Gaussian system. Even in this case we have constructed an example (see [5, Example 1]), where a) the integrand function $f(t)$ (in this case the integrand is nonrandom) is X-valued; b) the integrand function is not X-valued, but it is X^{**}-valued and c) the integrand function is not X^{**}-valued, but it is a GRE $T : X^* \to L_2[0,1]$.

The following result gives representation of the Banach space-valued functional of the Wiener process by the stochastic integral from the F_t^W-adapted GRP.

Proposition 1. *Let ξ be a Banach space-valued weak second order functional of the one-dimensional Wiener process. There exists the F_t^W-adapted GRP $T : X^* \to L_2([0,1] \times \Omega)$ such that for all $x^* \in X^*$*

$$\langle \xi, x^* \rangle = E\langle \xi, x^* \rangle + \int_0^1 Tx^*(t,\omega) dW_t. \tag{0.1}$$

Proof. Let ξ be a Banach space-valued weak second order functional of the one-dimensional Wiener process. For any $x^* \in X^*$, $\langle \xi, x^* \rangle$ is one-dimensional functional of the Wiener process. By the one-dimensional theorem, there exists the unique F_t^W-adapted one-dimensional random process $f(x^*, t, \omega)$ such that

$$\langle \xi, x^* \rangle = E\langle \xi, x^* \rangle + \int_0^1 f(x^*, t, \omega) dW_t.$$

Consider the map $T : X^* \to L_2([0,1], \Omega)$, $Tx^* = f(x^*, t, \omega)$. It is easy to see that T is a linear operator. Further,

$$\begin{aligned} \infty > \sup_{\|x^*\| \le 1} E\langle \xi - E\xi, x^* \rangle^2 &= \sup_{\|x^*\| \le 1} E\Big(\int_0^1 f(x^*, t, \omega) dW_t \Big)^2 \\ &= \sup_{\|x^*\| \le 1} \int_0^1 Ef(x^*, t, \omega)^2 dt = \sup_{\|x^*\| \le 1} \int_0^1 E(T_t x^*)^2 dt. \end{aligned}$$

That is, $T : X^* \to L_2([0,1], \Omega)$ is bounded, and therefore, this is the GRP. □

Remark 1. The representation (0.1) of the Wiener functional is unique for any x^* $dt \otimes dP$-almost everywhere. Indeed, if we have two representations of ξ,

$$\langle \xi, x^* \rangle = E\langle \xi, x^* \rangle + \int_0^1 T_1 x^*(t,\omega) dW_t = E\langle \xi, x^* \rangle + \int_0^1 T_2 x^*(t,\omega) dW_t,$$

then

$$\begin{aligned} 0 &= \sup_{\|x^*\| \le 1} E\Big(\int_0^1 (T_1 x^*(t,\omega) - T_2 x^*(t,\omega)) dW_t \Big)^2 \\ &= \sup_{\|x^*\| \le 1} E \int_0^1 (T_1 x^*(t,\omega) - T_2 x^*(t,\omega))^2 dt. \end{aligned}$$

For any GRP $T : X^* \to L_2([0,1] \times \Omega)$ from $\mathcal{M}_1^{(\lambda,P)}$, the correlation operator of T is called the linear, bounded operator from X^* to X^{**}, $R_T = T^*T$.

Proposition 2. *If for any functional of the Wiener process ξ,*

$$\langle \xi, x^* \rangle = \int_0^1 Tx^*(t,\omega) dW_t,$$

then $R_T = T^*T$ maps X^* onto X.

Proof. For any x^* and y^*,

$$\langle R_T x^*, y^*\rangle = \langle T^*Tx^*, y^*\rangle = \langle Tx^*, Ty^*\rangle = \int_0^1 ETx^*(t,\omega)Ty^*(t,\omega)dt$$
$$= E\Big(\int_0^1 Tx^*(t,\omega)dW_t \times \int_0^1 Ty^*(t,\omega)dW_t\Big)$$
$$= E(\langle(\xi - E\xi), x^*\rangle \times \langle(\xi - E\xi), y^*\rangle) = \langle R_\xi x^*, y^*\rangle,$$

where R_ξ is a covariance operator of ξ, which maps X^* onto X (see [7, Theorem 3.2.1]). Therefore, R_T maps X^* onto X. □

As is known (see [2, Theorem 5.6]), for the one-dimensional case, if the joint distribution of the Wiener process and its one-dimensional functional is Gaussian, then the sequence of step functions

$$f_n(t) = \sum_{i=0}^{2^n-1} 2^n E(\xi - E\xi)(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)$$

converges in $L_2[0,1]$ to the integrand function $f(t)$, $\int_0^1 f^2(t)dt < \infty$ and

$$\xi_n = E\xi + \int_0^1 f_n(t)dW_t$$

converges in $L_2(\Omega, B, P)$ to

$$\xi = E\xi + \int_0^1 f(t)dW_t.$$

First, we give the generalization of this theorem for an arbitrary (non Gaussian) case when the functional of the Wiener process is one-dimensional.

Theorem 1. *Let the square integrable random variable ξ be a functional of the Wiener process. The sequence of step functions*

$$f_n(t,\omega) = \sum_{i=0}^{2^n-1} 2^n E((\xi_{\frac{i+1}{2^n}} - \xi_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})/F^W_{\frac{i}{2^n}})I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t) \tag{0.2}$$

converges in $L_2([0,1],\Omega)$ to the F^W_t-adapted random process $f(t,\omega)$ and

$$\xi = \lim_{n\to\infty} \int_0^1 f_n(t,\omega)dW_t = \int_0^1 f(t,\omega)dW_t.$$

Proof. First of all, we prove the following

Lemma 1. *Let $\xi = \int_0^1 f(t,\omega)dW(t)$ be a real-valued functional of the Wiener process. Then for any $0 \le a \le b$,*

$$E((\xi_b - \xi_a)(W_b - W_a)/F^W_a) = E\Big(\int_a^b f(t,\omega)dt/F^W_a\Big),$$

where $\xi_t = E(\xi/F^W_t) = \int_0^t f(s,\omega)dW(s)$.

Proof of Lemma 1. Consider the left part of the equality and denote $m \equiv (b-a)^{-1}$. Remember that by Lemma 1.1.3 from [7], for any $f(t,\omega) \in L_2([0,1]\times\Omega)$, the sum

$$\sum_{i=1}^{2^n-1} 2^n \int\limits_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} f(s,\omega)ds I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)$$

converges to $f(t,\omega)$ in $L_2([0,1]\times\Omega)$.

Next, we have

$$E((\xi_b-\xi_a)(W_b-W_a)/F_a^W) = \lim_{n\to\infty} E\Bigg(\Bigg(\Bigg(\sum_{i=1}^{2^n-1}(2^n m \int\limits_{(a+\frac{(i-1)}{2^n m})\vee 0}^{a+\frac{i}{2^n m}} f(s,\omega)ds)$$

$$\times(W_{a+\frac{(i+1)}{2^n m}} - W_{a+\frac{i}{2^n m}})\Bigg)\Bigg(\sum_{i=0}^{2^n-1}(W_{a+\frac{(i+1)}{2^n m}} - W_{a+\frac{i}{2^n m}})\Bigg)\Bigg)/F_a^W\Bigg)$$

$$= \lim_{n\to\infty} E\Bigg(\sum_{i=1}^{2^n-1}\Bigg(\Bigg(2^n m \int\limits_{(a+\frac{(i-1)}{m2^n})\vee 0}^{a+\frac{i}{m2^n}} f(s,\omega)ds\Bigg)/F_a^W\Bigg)$$

$$\times E(W_{a+\frac{(i+1)}{m2^n}} - W_{a+\frac{i}{m2^n}})^2\Bigg) = \lim_{n\to\infty}\sum_{i=1}^{2^n-1} E\Bigg(2^n m \int\limits_{(a+\frac{(i-1)}{m2^n}\vee 0}^{a+\frac{i}{m2^n}} f(s,\omega)ds)/F_a^W\Bigg)\frac{1}{m2^n}$$

$$= \lim_{n\to\infty} E\Bigg(\sum_{i=1}^{2^n-1}\int\limits_{a+\frac{i}{m2^n}}^{a+\frac{(i+1)}{m2^n}}\Bigg(2^n m \int\limits_{(a+\frac{(i-1)}{m2^n}\vee 0}^{a+\frac{i}{m2^n}} f(s,\omega)ds)I_{(a+\frac{i}{m2^n},a+\frac{(i+1)}{m2^n}]}(t)\Bigg)dt/F_a^W\Bigg)$$

$$= \lim_{n\to\infty} E\Bigg(\sum_{i=1}^{2^n-1}\Bigg(2^n m \int\limits_{(a+\frac{(i-1)}{m2^n}\vee 0}^{a+\frac{i}{m2^n}} f(s,\omega)ds)\frac{1}{m2^n}\Bigg)/F_a^W\Bigg)$$

$$= \lim_{n\to\infty} E\Bigg(\sum_{i=1}^{2^n-1}\int\limits_{(a+\frac{(i-1)}{m2^n})\vee 0}^{a+\frac{i}{m2^n}} f(s,\omega)ds/F_a^W\Bigg) = E\Bigg(\int\limits_a^{a+\frac{1}{m}} f(t,\omega)dt/F_a^W\Bigg),$$

as

$$E\Bigg(\int\limits_{(a+\frac{(i-1)}{m2^n})\vee 0}^{a+\frac{i}{m2^n}} f(s,\omega)ds(W_{a+\frac{(i+1)}{2^n m}} - W_{a+\frac{i}{2^n m}})(W_{a+\frac{(j+1)}{2^n m}} - W_{a+\frac{j}{2^n m}})/F_a^W\Bigg) = 0,$$

when $i \neq j$.

Thus, the proof of the lemma 1 is completed.

Consider now the following sum

$$\sum_{i=1}^{2^n-1} 2^n \int\limits_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} f(s,\omega)ds I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t).$$

According to Lemma 1.1.3 of [7], this sum converges likewise to $f(t,\omega)$ in $L_2([0,1]\times\Omega)$. Therefore,

$$\sum_{i=1}^{2^n-1} 2^n \int\limits_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} f(s,\omega)ds I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t) - \sum_{i=1}^{2^n-1} 2^n \int\limits_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} f(s,\omega)ds) I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)$$

tends to 0 in $L_2([0,1]\times\Omega)$. That is,

$$\int_0^1 E\Big(\sum_{i=1}^{2^n-1} 2^n\Big(\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} f(s,\omega)ds-\int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} f(s,\omega)ds\Big)I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)\Big)^2 dt\to 0.$$

Hence,

$$\sum_{i=1}^{2^n-1} 2^{2n}E\Big(\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} f(s,\omega)ds-\int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} f(s,\omega)ds\Big)^2\Big(\frac{1}{2^n}\Big)\to 0.$$

Therefore,

$$\sum_{i=1}^{2^n-1} 2^n E\Big(E\Big(\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} f(s,\omega)ds/F^W_{\frac{i}{2^n}}\Big)-E\Big(\int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} f(s,\omega)ds/F^W_{\frac{i}{2^n}}\Big)\Big)^2\to 0.$$

That is,

$$\begin{aligned}\int_0^1 E\Big(&\sum_{i=1}^{2^n-1} 2^n\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} f(s,\omega)dsI_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)\\ &-\sum_{i=1}^{2^n-1} 2^n E\Big(\int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} f(s,\omega)ds/F^W_{\frac{i}{2^n}}\Big)I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)\Big)^2\to 0.\end{aligned}$$

But the first sum converges to the $f(t,\omega)$. Therefore the sequence of F_t^W-adapted step functions

$$\sum_{i=1}^{2^n-1} 2^n E\Big(\int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} f(s,\omega)ds/F^W_{\frac{i}{2^n}}\Big)I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)$$

converges to $f(t,\omega)$ in $L_2([0,1]\times\Omega)$.

Now we can construct the sequence of step functions $f_n(t,\omega)$, $n\in N$, the stochastic integral of which converges to ξ: let us consider

$$\xi_n=\sum_{i=1}^{2^n-1} 2^n E((\xi_{\frac{i+1}{2^n}}-\xi_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}}-W_{\frac{i}{2^n}})/F^W_{\frac{i}{2^n}})\times(W_{\frac{i+1}{2^n}}-W_{\frac{i}{2^n}})=\int_0^1 f_n(t,\omega)dW_t,$$

where

$$\begin{aligned}f_n(t,\omega)=&\sum_{i=1}^{2^n-1} 2^n E((\xi_{\frac{i+1}{2^n}}-\xi_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}}-W_{\frac{i}{2^n}})/F^W_{\frac{i}{2^n}})I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)\\ &\sum_{i=1}^{2^n-1} 2^n E\Big(\int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} f(s,\omega)ds/F^W_{\frac{i}{2^n}}\Big)I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t).\end{aligned}\tag{0.3}$$

Then we have $f_n(t,\omega)\to f(t,\omega)$ in $L_2([0,1]\times\Omega)$ and

$$\int_0^1 f_n(t,\omega)dW_t\to\int_0^1 f(t,\omega)dW_t\quad\text{in}\ \ L_2(\Omega).\qquad\square$$

Remark 2. In case when the joint distribution of functional of the Wiener process and of the Wiener process is Gaussian, then

$$f_n(t) = \sum_{i=0}^{2^n-1} 2^n E\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}}) I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)$$

$$= \sum_{i=0}^{2^n-1} 2^n E((\xi_{\frac{i}{2^n}} + (\xi_{\frac{i+1}{2^n}} - \xi_{\frac{i}{2^n}}) + (\xi - \xi_{\frac{i+1}{2^n}})))(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}}) I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)$$

$$= \sum_{i=0}^{2^n-1} 2^n E((\xi_{\frac{i+1}{2^n}} - \xi_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})) I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)$$

$$= \sum_{i=0}^{2^n-1} 2^n E(((\xi_{\frac{i+1}{2^n}} - \xi_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}}))/F^W_{\frac{i}{2^n}} I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t).$$

Therefore formula (0.3) is the generalization of formula (0.2) for an arbitrary (nonGaussian) case.

Let now ξ be a Banach space-valued functional of the Wiener process. As in the one-dimensional case, consider the sequence of step functions

$$f_n(t,\omega) = \sum_{i=0}^{2^n-1} 2^n E((\xi_{\frac{i+1}{2^n}} - \xi_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})/F^W_{\frac{i}{2^n}}) I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t).$$

The random step function $f_n(t,\omega)$ does not always exist as a X-valued random process, because the conditional expectation $E((\xi_t - \xi_s)(W_t - W_s)/F_s^W)$ for the weak second order random element $\xi_t - \xi_s$ does not exist, in general. Nevertheless, we can consider the GRE

$$Tx^* = E(\langle(\xi_t - \xi_s), x^*\rangle(W_t - W_s)/F_s^W).$$

From Proposition 1 and Theorem 1 we immediately obtain the following

Proposition 3. *For any weak second order Banach space-valued functional of the Wiener process $\xi : \Omega \to X$, there exists the sequence of step generalized random functions $(T_n)_{n\in N}$, such that for all $x^* \in X^*$,*

$$\int_0^1 E(T_n x^* - Tx^*)^2 dt \to 0,$$

when $n \to \infty$, where T is the GRP such that

$$\langle \xi, x^*\rangle = E\langle \xi, x^*\rangle + \int_0^1 Tx^*(t,\omega)dW_t.$$

Proof. For any $x^* \in X^*$, let $T_n x^* = \langle f_n(t,\omega), x^*\rangle$. By Theorem 1, $T_n x^* \to Tx^*$, when $n \to \infty$. By Proposition 1, we have

$$\langle \xi, x^*\rangle = E\langle \xi, x^*\rangle + \int_0^1 Tx^*(t,\omega)dW_t. \qquad \square$$

The following theorem is a generalization of the one-dimensional Theorem 1 for the Banach space-valued functionals of the one-dimensional Wiener process.

Theorem 2. *Let ξ be a Banach space-valued F_1^W measurable weak second order random element such that in the representation* (0.1) *the GRP $T \in \mathcal{M}_1^{\lambda,P}$ $T : [0,1] \to \mathcal{M}_1$ is separable-valued and*

$$\int_0^1 \|T\|^2_{\mathcal{M}_1} < \infty.$$

There exists the sequence of F_t^W-adapted step functions $T_n(t,\omega)$, $n \in N$ converging in $\mathcal{M}_1^{\lambda,P}$ to the F_t^W-adapted GRP $T : X^* \to L_2([0,1],\Omega)$ such that the sequence of the stochastic integrals

$$\int_0^1 T_n x^*(t,\omega)dW_t$$

converges to

$$\int_0^1 Tx^*(t,\omega)dW_t = \langle \xi - E\xi, x^* \rangle \quad in \ \ \mathcal{M}_1.$$

Proof. By Proposition 1, for the weak second order random element ξ, there exists the unique GRP $T : X^* \to L_2([0,1],\Omega)$ such that for all $x^* \in X^*$

$$\langle \xi, x^* \rangle = E\langle \xi, x^* \rangle + \int_0^1 Tx^*(t,\omega)dW_t.$$

Consider

$$\langle T_n(t,\omega), x^* \rangle = \sum_{i=0}^{2^n-1} 2^n E((\langle \xi_{\frac{i+1}{2^n}}, x^* \rangle - \langle \xi_{\frac{i}{2^n}}, x^* \rangle)(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})/F^W_{\frac{i}{2^n}}) I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t).$$

By Theorem 1, $\langle T_n(t,\omega), x^* \rangle$, $n \in N$ converges in $L_2([0,1],\Omega)$ to the one-dimensional functional of the Wiener process $Tx^*(t,\omega)$ and we have

$$T_n x^*(t,\omega) = \sum_{i=1}^{2^n-1} 2^n \Bigg(\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} Tx^*(s,\omega)ds \Bigg) I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t).$$

Further, it is easy to see that $T_n \in \mathcal{M}_1^{\lambda,P}$ for all $n \in N$ and $\|T_n\| \le \|T\|$. Indeed,

$$\|T_n\|^2 = \sup_{\|x^*\|\le 1} \int_0^1 E\Bigg(\sum_{i=1}^{2^n-1} 2^n \Bigg(\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} Tx^*(s,\omega)ds I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t) \Bigg)^2 dt \Bigg)$$

$$= \sup_{\|x^*\|\le 1} \int_0^1 E\Bigg(\sum_{i=1}^{2^n-1} 2^{2n} \Bigg(\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} Tx^*(s,\omega)ds \Bigg)^2 I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t) \Bigg)dt$$

$$= \sup_{\|x^*\|\le 1} \Bigg(\sum_{i=1}^{2^n-1} 2^{2n} \frac{1}{2^n} E\Bigg(\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} Tx^*(s,\omega)ds \Bigg)^2 \Bigg)$$

$$\le \sup_{\|x^*\|\le 1} \sum_{i=1}^{2^n-1} E \int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} (Tx^*(s,\omega))^2 ds$$

$$= \sup_{\|x^*\|\le 1} E \int_0^{1-\frac{1}{2^n}} (Tx^*(s,\omega))^2 ds \le \|Tx^*(t,\omega)\|^2.$$

If $T \in \mathcal{M}_1^{\lambda,P}$ is a continuous function $T : [0,1] \to \mathcal{M}_1$, then $T_n \to T$ in $\mathcal{M}_1^{\lambda,P}$. Really,

$$\|T - T_n\|^2 = \sup_{\|x^*\|\le 1} E \int_0^1 \Bigg(\sum_{i=1}^{2^n-1} 2^n \Bigg(\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} Tx^*(s,\omega)ds I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t) \Bigg) - Tx^*(t,\omega) \Bigg)^2 dt$$

$$= \sup_{\|x^*\|\le 1} E\int_0^1 \Big(\sum_{i=1}^{2^n-1}\Big(2^n\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} Tx^*(s,\omega)ds - Tx^*(t,\omega)\Big)I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)\Big)^2 dt$$

$$= \sup_{\|x^*\|\le 1} E\int_0^1 \Big(\sum_{i=1}^{2^n-1}\Big(2^n\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} Tx^*(s,\omega)ds - Tx^*(t,\omega)\Big)\Big)^2 I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t)dt$$

$$= \sup_{\|x^*\|\le 1} E\sum_{i=1}^{2^n-1}\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}}\Big(Tx^*(t,\omega)-\Big(2^n\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} Tx^*(s,\omega)ds\Big)\Big)^2 dt$$

$$\le \sum_{i=1}^{2^n-1}\sup_{\|x^*\|\le 1} E\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}}\Big(Tx^*(t,\omega)-\Big(2^n\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} Tx^*(s,\omega)ds\Big)\Big)^2 dt$$

$$\le \sum_{i=1}^{2^n-1}\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}}\sum_{i=1}^{2^n-1}\sup_{\|x^*\|\le 1} E\Big(2^n\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}}(Tx^*(t,\omega)-Tx^*(s,\omega)ds)\Big)^2 dt$$

$$\le \sum_{i=1}^{2^n-1}\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}}\sum_{i=1}^{2^n-1}\Big(2^{2n}\frac{1}{2^n}\int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}}\sup_{\|x^*\|\le 1} E(Tx^*(t,\omega)-Tx^*(s,\omega))^2 ds\Big)dt < \varepsilon,$$

as the function $T:[0,1]\to\mathcal{M}_1$ is continuous, for any $\varepsilon>0$ and sufficiently large n,

$$\sup_{\|x^*\|\le 1} E(Tx^*(t,\omega)-Tx^*(s,\omega))^2 ds)) < \varepsilon,$$

when $|t-s|<\frac{1}{2^n}$.

Consider now an arbitrary separable-valued $T:[0,1]\to\mathcal{M}_1$. Any fixed $x^*\in X^*$ and $g\in L_2(\Omega)$, generates the linear continuous functional $f:\mathcal{M}_1\to R$,

$$f(x^*,g)(T)=\int_\Omega Tx^*(\omega)g(\omega)dP.$$

The set of such functionals separates the points of the Banach space $\mathcal{M}_1$. As $T:[0,1]\to\mathcal{M}_1$ is separable-valued and $f(x^*,g)T(t)$ is measurable, by the Pettis theorem (see [6, Proposition 1.1.10]), $T:[0,1]\to\mathcal{M}_1$ is measurable. As $\int_0^1\|T(t)\|dt<\infty$, the Bochner integral $\int_s^t T(t)dt$ exists for all $0\le s<t\le 1$. Let $T(t)$ be a bounded function. Consider $T_m(t):=m\int_{(t-\frac{1}{m})\vee 0}^t T(s)ds$, $m\in N$. $T_m(t)\to T(t)$ a.s. (see [1, Corollary 2 of Theorem 3.8.5]). By the Lebesgue theorem, $\int_0^1\|T_m(t)-T(t)\|^2dt\to 0$. As $T_m(t)$ is continuous for all $m\in N$, there exists the sequence of F_t^W- adapted step functions T_{mn}, $n\in N$ such that $\int_0^1\|T_{mn}(t)-T_m(t)\|^2dt\to 0$. Therefore we can choose the sequence of step functions $(T_n)_{n\in N}$ such that $\int_0^1\|T_n(t)-T(t)\|^2dt\to 0$. It is now easy to get the sequence of step functions converging to the arbitrary separable-valued $T:[0,1]\to\mathcal{M}_1$, with $\int_0^1\|T(t)\|^2dt<\infty$. □

Remark 3. By Proposition 1, for the X-valued weak second order functional of the Wiener process the integrand $T(t,\omega)$ belongs to the Banach space $\mathcal{M}_1^{\lambda,P}$. The existence of step functions converging to the integrand we prove in the case for $T\in L_2([0,1],\mathcal{M}_1)$ which is separable-valued. We prove the convergence in $L_2([0,1],\mathcal{M}_1)$, but there arises the question whether this theorem is true for $\mathcal{M}_1^{\lambda,P}$ without the above restrictions? The answer is unknown.

Remark 4. If the sequence $f_n(t,\omega)$, $n\in N$ is such that the members of it as X-valued random processes exist (for example, the functional ξ has strong p-th moment for any $p>1$), then from

Theorem 2 it follows that the integrand process $T(t,\omega)$ belongs to $\mathcal{M}_2^{(\lambda,P)} \subset \mathcal{M}_1^{(\lambda,P)}$. If the sequence $f_n(t,\omega)$, $n \in N$, of X-valued random processes converges in X, then the integrand process is X-valued and in this case we have the representation of the Banach space-valued functional by the stochastic integral from the Banach space F_t^W-adapted X-valued random process.

Remark 5. It is easy to see that

$$\begin{aligned}\sum_{i=0}^{2^n-1} 2^n E((\xi_{\frac{i+1}{2^n}} - \xi_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})/F^W_{\frac{i}{2^n}}) I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t) \\ = \sum_{i=0}^{2^n-1} 2^n E(\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})/F^W_{\frac{i}{2^n}}) I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t).\end{aligned}$$

As

$$\begin{aligned}E(\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})/F^W_{\frac{i}{2^n}}) \\ = E(((\xi - \xi_{\frac{i+1}{2^n}}) + (\xi_{\frac{i+1}{2^n}} - \xi_{\frac{i}{2^n}}) + \xi_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})/F^W_{\frac{i}{2^n}}) \\ = E((\xi_{\frac{i+1}{2^n}} - \xi_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})/F^W_{\frac{i}{2^n}}).\end{aligned}$$

Acknowledgement

The first author was supported by the Shota Rustaveli National Science Foundation of Georgia (SRNSFG), grant number: DI-18-1429.

References

1. E. Hille, R. S. Phillips, *Functional Analysis and Semi-groups.* rev. ed. American Mathematical Society Colloquium Publications, vol. 31. American Mathematical Society, Providence, R. I., 1957.
2. R. S. Liptser, A. N. Shiriaev, *Statistics of Random Processes.* (Russian) vol. 15. Izdat. Nauka, Moscow, 1974.
3. J. Maas, J. M. A. M. van Neerven, A Clark–Ocone formula in UMD Banach spaces. *Electron. Commun. Probab.* **13** (2008), 151–164.
4. B. Mamporia, A stochastic differential equation for generalized random processes in a Banach space. (Russian) *translated from Teor. Veroyatn. Primen.* **56** (2011), no. 4, 704–725 *Theory Probab. Appl.* **56** (2012), no. 4, 602–620.
5. B. Mamporia, O. Purtukhia, On functionals of the Wiener process in a Banach space. *Trans. A. Razmadze Math. Inst.* **172** (2018), no. 3, part A, 420–428.
6. D. Nualart, *The Malliavin Calculus and Related Topics.* Second edition. Probability and its Applications (New York). Springer-Verlag, Berlin, 2006.
7. N. N. Vakhania, V. I. Tarieladze, S. A. Chobanyan, *Probability Distributions on Banach Spaces.* Translated from the Russian and with a preface by Wojbor A. Woyczynski. Mathematics and its Applications (Soviet Series), 14. D. Reidel Publishing Co., Dordrecht, 1987.
8. J. M. A. M. van Neerven, M. Veraar, L. Weis, Stochastic Integration in Banach Spacesa Survey. In: *Stochastic Analysis: a Series of Lectures*, pp. 297–332, Progr. Probab., 68, Birkhäuser/Springer, Basel, 2015.

(Received 16.12.2019)

[1]Georgian Technical University, Muskhelishvili Institute of Computational Mathematics, Tbilisi, Georgia

E-mail address: badrimamporia@yahoo.com

[2]A. Razmadze Mathematical Institute of I. Javakhishvili Tbilisi State University, Faculty of Exact and Natural Sciences, Department of Mathematics, 6 Tamarashvili Str., Tbilisi 0177, Georgia

E-mail address: o.purtukhia@gmail.com

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 217–233

LOGISTIC REGRESSION WITH TOTAL VARIATION REGULARIZATION

SARA VAN DE GEER

A paper devoted to the 75^{th} *birthday of Estate Khmaladze*

Abstract. We study logistic regression with total variation penalty on the canonical parameter and show that the resulting estimator satisfies a sharp oracle inequality: the excess risk of the estimator is adaptive to the number of jumps of the underlying signal or an approximation thereof. In particular, when there are finitely many jumps, and jumps up are sufficiently separated from jumps down, then the estimator converges with a parametric rate up to a logarithmic term $\log n/n$, provided the tuning parameter is chosen appropriately of order $1/\sqrt{n}$. Our results extend earlier results for quadratic loss to logistic loss. We do not assume any a priori known bounds on the canonical parameter, but instead only make use of the local curvature of the theoretical risk.

1. Introduction

In this paper we consider logistic regression with a total variation penalty on the canonical parameter. Total variation based de-noising was introduced in [15]. Our aim here is to develop theoretical results that show that the estimator adapts to the number of jumps in the signal.

For $i = 1, \ldots, n$, let $Y_i \in \{0, 1\}$ be independent binary observations. Write the unknown probability of success as $\theta_i^0 := P(Y_i = 1)$, and let $f_i^0 := \log(\theta_i^0/(1-\theta_i^0))$ be the log-odds ratio, $i = 1, \ldots, n$. Define the total variation of a vector $f \in \mathbb{R}^n$ as

$$\mathrm{TV}(f) := \sum_{i=2}^{n} |f_i - f_{i-1}|.$$

We propose to estimate the unknown vector f^0 of log-odds ratios applying logistic regression with total variation regularization. The estimator is

$$\hat{f} := \arg \min_{f \in \mathbb{R}^n} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(-Y_i f_i + \log(1 + \mathrm{e}^{f_i}) \right) + \lambda \mathrm{TV}(f) \right\}.$$

Our goal is to derive oracle inequalities for this estimator. The approach we take shares some ideas with [4, 11] and [13]. These papers deal with least squares loss, whereas the current paper studies logistic loss. Moreover, instead of using the projection arguments of the previous mentioned papers, we use entropy bounds. This allows us to remove a redundant logarithmic term: we show that the excess risk of estimator $\hat{f}$ converges under certain conditions with rate $(s + 1) \log n/n$, where s is the number of jumps of f^0 or of an oracle approximation thereof (see Theorem 2.1). This extends the result in [6] to logistic loss and to a sharp oracle inequality.

To arrive at the results of this paper we require that $\|\hat{f}\|_\infty$ stays bounded with high probability. In Theorem 3.1 we show that this requirement holds assuming that both $\|f^0\|_\infty$ and $\mathrm{TV}(f^0)$ remain bounded.

The theory for a total variation regularization with the least squares loss (the fused Lasso) has been developed in a series of papers [4, 8, 14, 16, 17, 21, 22] including higher dimensional extensions [3, 5, 7, 12] and higher order total variation [6, 13, 18, 19].

Logistic regression with ℓ_1-regularization has many applications. When there are co-variables, the penalty is on the total variation of the coefficients. In [25], logistic regression with the fused Lasso is applied to spectral data, and in [9] to gene expression data, whereas [1] applies it to time-varying

2020 *Mathematics Subject Classification.* 62J07, 62G05.
Key words and phrases. Logistic regression; Oracle inequality; Total variation penalty.

networks. In [20] the penalty alternatively takes links between variables into account using a quadratic penalty. The papers [26] and [10] present algorithms for the fused Lasso. In [2], a Bayesian approach with the fused Lasso is presented.

This paper is organized as follows. In Section 2 we state the oracle inequality for $\hat{f}$ (Theorem 2.1). Section 3 derives a bound for $\|\hat{f}\|_\infty$ (Theorem 3.1). The remainder of the paper is devoted to proofs. Section 4 states some standard tools to this end, Section 5 contains a proof of Theorem 2.1 and Section 6 a proof of Theorem 3.1.

2. A Sharp Oracle Inequality

The empirical risk in this paper is given by the normalized minus log-likelihood

$$R_n(f) := \frac{1}{n}\sum_{i=1}^n \biggl(-Y_i f_i + \log(1+\mathrm{e}^{f_i})\biggr),\ f \in \mathbb{R}^n.$$

The theoretical risk is

$$R(f) := \mathbb{E} R_n(f),\ f \in \mathbb{R}^n$$

and $R(f) - R(f^0)$ is called the "excess risk". For $f \in \mathbb{R}^n$, we write $\dot{R}_n(f) := \partial R_n(f)/\partial f$ and $\dot{R}(f) := \mathbb{E}\dot{R}_n(f)$. These are column vectors in $\mathbb{R}^n$. Most of the arguments that follow go through for general convex differentiable loss functions. We do use, however, that or all $f \in \mathbb{R}^n$, $R_n(f) - R(f) = -\epsilon^T f/n$ where $\epsilon = Y - \mathbb{E}Y$ is the noise. In other words, f is the canonical parameter. In the case where the entries of the response vector Y are in $\{0,1\}$, the entries of a noise vector ϵ are bounded by 1. More generally, our theory would need that ϵ has sub-exponential entries. To avoid digressions, we simply restrict ourselves to logistic loss.

Fix a vector $\mathbf{f} \in \mathbb{R}^n$. This vector will play the role of the "oracle" as we will see in Theorem 2.1. We let $S := \{t_1, \ldots, t_s\}$ $(1 < t_1 < \cdots < t_s < n)$ be the location of its jumps:

$$\mathbf{f}_1 = \cdots = \mathbf{f}_{t_1-1} \neq \mathbf{f}_{t_1} = \cdots = \mathbf{f}_{t_2-1} \neq \mathbf{f}_{t_2} \cdots \mathbf{f}_{t_s-1} \neq \mathbf{f}_{t_s} = \cdots = \mathbf{f}_n.$$

Let $d_j := t_j - t_{j-1}$ be the distance between jumps, $j = 1, \ldots, r$, where $r = s+1$, $t_r := n+1$ and $t_0 = 1$. Define $d_{\max} := \max_{1\le j\le r} d_j$.

The quantities Δ_n^2, $\delta_n^2(t)$, $\lambda_n(t)$ and $\Gamma_n^2(t)$ we are about to introduce all depend on $\mathbf{f}$ although we do not express this in our notation. Moreover, being non-asymptotic, these quantities are somewhat involved. After the explicit expressions for Δ_n^2, $\delta_n^2(t)$ and $\lambda_n(t)$ we will give their asymptotic order of magnitude. The asymptotic order of magnitude for $\Gamma_n^2(t)$ depends on the situation. We discuss a special case after the statement of Theorem 2.1.

We let

$$\Delta_n^2 := \frac{4\sum_{j\in[1:r]:\ d_j\ge 1}(\log(d_j-1)+1)}{n} + \frac{s}{n},$$

and define for $t > 0$

$$\begin{aligned}\delta_n^2(t) :=& \biggl(4\nu A_0 \Delta_n + 8\sqrt{\frac{1+t+\log(3+2\log_2 n)}{n}}\biggr)^2 \\ &+ \biggl(\frac{2}{\nu} + 4\sqrt{\frac{A_0\Delta_n}{n}} + \frac{4\sqrt{1+t+\log(3+2\log_2 n)}}{n}\biggr) \\ &\times \biggl(\Delta_n + 2\sqrt{\frac{s}{n}}\biggr)^2,\end{aligned}$$

and

$$\lambda_n(t) := \frac{1}{\sqrt{n}}\biggl(\frac{4}{\nu} + 8\sqrt{\frac{A_0\Delta_n}{n}} + \frac{8\sqrt{1+t+\log(3+2\log_2 n)}}{n}\biggr).$$

One can see that

$$\Delta_n^2 = \mathcal{O}\biggl(\frac{(s+1)\log n}{n}\biggr).$$

Furthermore, for $\nu = 1$ (say) and each fixed t

$$\delta_n^2(t) = \mathcal{O}\bigg(\frac{(s+1)\log n}{n}\bigg),\ \lambda_n(t) = \mathcal{O}\bigg(\frac{1}{\sqrt{n}}\bigg),$$

assuming $n^{-1}\sqrt{(s+1)\log n/n} = \mathcal{O}(1)$ which is certainly true under the standard sparsity assumption $(s+1)\log n/n = o(1)$.

The quantity $\delta_n^2(t)$ will be a part of the bound for the excess risk of $\hat{f}$, and $\lambda_n(t)$ can be thought of as the "noise level" to be overruled by the penalty (see Theorem 2.1). The constant A_0 is the (universal) constant appearing when bounding the entropy of the class of functions with both $\|\cdot\|_\infty$ and $\mathrm{TV}(\cdot)$ bounded by 1 (see Lemma 4.3). The free parameter $t > 0$ determines the confidence level of our statements. Both $\delta_n(t)$ and $\lambda_n(t)$ depend on a further free parameter $\nu > 0$ which we do not express in our notation as one can simply choose $\nu = 1$. It is, however, an option to choose ν larger than 1, possibly growing with n: larger ν relaxes the requirement on the tuning parameter λ, but results in larger bounds for the excess risk.

Finally, we present a bound $\Gamma_n^2(t)$ for the so-called "effective sparsity" as introduced in [13] (see also Definition 5.1). The effective sparsity may be seen as a substitute for the sparsity, which is defined as the number of active parameters of the oracle, which is $s+1$. The effective sparsity will, in general, be larger, than $s+1$. Without going into details, we remark that this is due to correlations in the dictionary X when writing $f = Xb$, with dictionary $X \in \mathbb{R}^{n\times n}$ and coefficients $b_1 := f_1$, $b_k := f_k - f_{k-1}$, $k \in [2:n]$.

Let $q_{t_j} := \mathrm{sign}(\mathbf{f}_{t_j})$, $j = 1,\ldots,s$. We write $J_{\text{monotone}} := \{2 \le j \le s :\ q_{t_{j-1}} = q_{t_j}\}$ and $J_{\text{change}} := [1:r]\backslash J_{\text{monotone}}$. Thus J_{monotone} are jumps with the same sign as the previous one, and J_{change} are jumps that change sign. We count the first jump as well as the endpoint $t_r = n+1$ as a sign change. Our bound for the effective sparsity is now

$$\Gamma_n^2(t) := \frac{\lambda_n^2(t)}{\lambda^2} \sum_{j\in J_{\text{monotone}}} 8(\log(d_j)+1) + \sum_{j\in J_{\text{change}}} \frac{8n(\log(d_j)+2)}{d_j}.$$

The following theorem presents an oracle inequality for $\hat{f}$. Its proof can be found in Section 5.

Theorem 2.1. *Let $\mathcal{F}$ be a convex subset of $\mathbb{R}^n$ (possibly $\mathcal{F} = \mathbb{R}^n$) and*

$$\hat{f} := \arg\min_{f\in\mathcal{F}}\bigg\{R_n(f) + \lambda\mathrm{TV}(f)\bigg\}.$$

Assume $\mathbf{f} \in \mathcal{F}$ satisfies $\|\mathbf{f}\|_\infty \le B$ for some constant B and define

$$\kappa := \frac{(1+\mathrm{e}^B)^2}{\mathrm{e}^B}.$$

Take

$$\lambda \ge \lambda_n(t)\sqrt{\frac{d_{\max}}{2n}}.$$

Then with probability at least $\mathbb{P}(\|\hat{f}\|_\infty \le B) - \exp[-t]$, we have

$$R(\hat{f}) - R(\mathbf{f}) \le 4\kappa\delta_n^2(t) + \frac{\lambda^2}{4}\Gamma_n^2(t).$$

Keeping the constant B fixed, this theorem tells us that

$$R(\hat{f}) - R(\mathbf{f}) = \mathcal{O}_{\mathbf{P}}\bigg(\frac{\sum_{j=1}^r(\log(d_j)+1)}{n} + \lambda^2\Gamma_n^2\bigg),$$

where we recall that $r = s+1$. If the jumps of $\mathbf{f}$ are roughly equidistant we see that $d_j \asymp d_{\max} \sim n/r$. Taking $\lambda \asymp \lambda_n(t)/\sqrt{r} \asymp \sqrt{1/(nr)}$, the bound for the effective sparsity $\Gamma_n^2(t)$ is in the worst case (where the jumps of $\mathbf{f}$ have alternating signs) of order $r^2\log(n/r)$. In other words, in that case the rate is $R(\hat{f}) - R(\mathbf{f}) = \mathcal{O}_{\mathbf{P}}(r\log(n/r)/n)$, which for least squares loss is the minimax rate (see [8]).

If $\mathbf{f}$ is monotone, with $\lambda \asymp \sqrt{d_{\max}}/n$, we get

$$\lambda^2\Gamma_n^2 \asymp \frac{\sum_{j=2}^s \log(d_j)+1}{n} + \frac{1}{n}\bigg(\frac{\log(d_1)d_{\max}}{d_1} + \frac{\log(d_r)d_{\max}}{d_r}\bigg).$$

In other words, the first jump of $\mathbf{f}$ should not occur to early, and the last jump not too late, relative to the distance between the jumps.

We note that the choice $\lambda \asymp \lambda_n(t)\sqrt{d_{\max}/n}$ depends on the oracle $\mathbf{f}$. Thus, if the tuning parameter λ is given, the choice of $\mathbf{f}$ depends on λ.

We assumed that $\|\mathbf{f}\|_\infty \le B$. We do not assume $\|f^0\|_\infty$ to be bounded by the same constant B, but we do hope for a good approximation $\mathbf{f}$ of f^0 with $\|\mathbf{f}\|_\infty \le B$. Nevertheless, Theorem 2.1 presents a sharp oracle inequality directly comparing $R(\hat{f})$ with $R(\mathbf{f})$: it does not require that the excess risk $R(\mathbf{f}) - R(f^0)$ is small in any sense. In the same spirit, the theorem requires that $\|\hat{f}\|_\infty \le B$ with high probability. This can be accomplished by taking $\mathcal{F} := \{f \in \mathbb{R}^n : \ \|f\|_\infty \le B\}$ (or some convex subset thereof). Theorem 2.1 holds for any B, i.e., it is a free parameter. However, one may not want to force $\hat{f}$ to be bounded by a given constant, but let the data decide for a bound on $\hat{f}$. This is a reason why we establish Theorem 3.1 given in the next section.

3. Showing that $\|\hat{f}\|_\infty$ is Bounded (Instead of Assuming this)

Since f^0 minimizes $R(f)$, a two-term Taylor expansion around f^0 gives

$$R(f) - R(f^0) = \frac{1}{2}(f-f^0)^T\ddot{R}(\tilde{f})(f-f^0)$$

where $\tilde{f}_i$ lies between f_i and f_i^0, $i = 1,\ldots,n$. It follows that

$$R(f) - R(f_0) \ge \frac{1}{2K_f^2}\|f-f^0\|_{Q_n}^2,$$

where

$$\|\cdot\|_{Q_n} = \|\cdot\|_2/\sqrt{n}$$

and where (for logistic loss)

$$K_f^2 := \frac{(1+\mathrm{e}^{\|f\|_\infty \vee \|f^0\|_\infty})^2}{\mathrm{e}^{\|f\|_\infty \vee \|f^0\|_\infty}}.$$

Thus, if both $\|f\|_\infty$ and $\|f^0\|_\infty$ stay within the bounds, we have a standard quadratic curvature of $R(\cdot)$ at f^0. Otherwise, the constant K_f grows exponentially fast. We will therefore assume that $\|f^0\|_\infty$ stays bounded and our task is then to show that $\|\hat{f}\|_\infty$ stays bounded, as well. The following theorem (where we have not been very careful with the constants) is derived in Section 6.

Theorem 3.1. *Let* $\mathrm{TV}(f^0) \le M_0$ *for some constant* $M_0 \ge 1$. *Define*

$$K := \frac{(1+\mathrm{e}^{1+2^4M_0+\|f^0\|_\infty})^2}{\mathrm{e}^{1+2^4M_0+\|f^0\|_\infty}}.$$

Suppose

$$\begin{aligned}\lambda &\le \Big(2^4(2K^2)M_0\Big)^{-1},\\ \lambda &\ge 2^8 n^{-2/3}A_0^{2/3}(2K^2)^{1/3},\\ \lambda &\ge 2^8(2K^2)\frac{1+t}{n},\end{aligned}$$

where the last inequality holds for some $t > 0$, *and where in the second last inequality* A_0 *is the constant appearing when bounding the entropy of the class of functions with both* $\|\cdot\|_\infty$ *and* $\mathrm{TV}(\cdot)$ *bounded by* 1 (*see Lemma* 4.3). *Then with probability at least* $1 - \exp[-t]$ *it holds that*

$$\frac{\|\hat{f}-f^0\|_{Q_n}^2}{2K^2} + \lambda\mathrm{TV}(\hat{f}-f^0) \le 4\lambda M_0$$

and

$$\|\hat{f} - f^0\|_\infty \leq \frac{1 + 8M_0}{2}.$$

One may object that the conditions on the tuning parameter λ depend on f^0 via the bounds on $\|f^0\|_\infty$ and $\mathrm{TV}(f^0)$. On the other hand, the choice of λ in Theorem 2.1 will be of larger order than $n^{-2/3}$ if one aims at adaptive results, and it will need to tend to zero. For such λ and for $\|f^0\|_\infty$ and $\mathrm{TV}(f^0)$ remaining bounded, the conditions of Theorem 3.1 will be met for all n sufficiently large.

4. Some Standard Results Useful for Both Theorem 2.1 and Theorem 3.1

Lemma 4.1. *For all vectors $g \in \mathbb{R}^n$, we have*

$$\mathbb{P}(\epsilon^T g \geq \|g\|_2 \sqrt{2t}) \leq \exp[-t], \ \forall\ t > 0.$$

Proof. The entries in ϵ have mean zero, are bounded by 1, and are independent. This means we can apply Hoeffding's inequality to $\epsilon^T g/n$. □

For $\mathbf{Q}$ a probability measure on $\{1, \ldots, n\}$ and a set $\mathcal{G} \subset \mathbb{R}^n$ we let $H(\cdot, \mathcal{G}, \mathbf{Q})$ be the entropy[1] of $\mathcal{G}$ endowed with the metric induced by the $L_2(\mathbf{Q})$-norm

Lemma 4.2. *Let $\mathcal{G} \subset \mathbb{R}^n$ be a set with diameter*

$$R := \sup_{g \in \mathcal{G}} \|g\|_{Q_n}.$$

Suppose

$$J(R) := 2 \int_0^R \sqrt{2H(u, \mathcal{G}, Q_n)} du$$

exists. Then for all $t > 0$, with probability at least $1 - \exp[-t]$ it holds that

$$\sup_{g \in \mathcal{G}} \epsilon^T g/n \leq \frac{J(R)}{\sqrt{n}} + 4R\sqrt{\frac{1+t}{n}}.$$

Proof. We can apply Hoeffding's inequality to $\epsilon^T g/n$ for each g fixed (see Lemma 4.1). The result of the current lemma is thus essentially applying Dudley's entropy integral. The constants are taken from Theorem 17.3 in [23]. □

Lemma 4.3. *Let $\mathcal{G} := \{g \in \mathbb{R}^n :\ \|g\|_\infty \leq 1,\ \mathrm{TV}(g) \leq 1\}$. It holds for any probability measure $\mathbf{Q}$*

$$H(u, \mathcal{G}, \mathbf{Q}) \leq \frac{A_0}{u}\ \forall\ u > 0,$$

where A_0 is a universal constant.

Proof. See [24], Theorem 2.7.5. □

5. Proof of Theorem 2.1.

5.1. **The main body of the proof of Theorem 2.1.**

The following lemma is Lemma 7.1 in [23]. We present a proof for completeness.

Lemma 5.1. *Let $\mathcal{F}$ be a convex subset of $\mathbb{R}^n$ (possibly $\mathcal{F} = \mathbb{R}^n$) and*

$$\hat{f} := \arg\min_{f \in \mathcal{F}} \left\{ R_n(f) + \lambda \mathrm{TV}(f) \right\}.$$

Then for all $f \in \mathcal{F}$,

$$-\dot{R}_n(\hat{f})^T (f - \hat{f}) \leq \lambda \mathrm{TV}(f) - \lambda \mathrm{TV}(\hat{f}).$$

[1] For $u > 0$ the u-covering number $N(u)$ of a metric space $(\mathcal{V}, d)$ is the smallest N such that there exists $\{v_j\}_{j=1}^N \subset \mathcal{V}$ with $\sup_{v \in \mathcal{V}} \min_{1 \leq j \leq N} d(v, v_j) \leq u$. The entropy is $H(\cdot) := \log N(\cdot)$.

Proof of Lemma 5.1. Define, for $0 < \alpha < 1$, $\hat{f}_\alpha := (1-\alpha)\hat{f} + \alpha f$. Then, using the convexity of $\mathcal{F}$

$$\begin{aligned} R_n(\hat{f}) + \lambda \mathrm{TV}(\hat{f}) \leq & R_n(\hat{f}_\alpha) + \lambda \mathrm{TV}(\hat{f}_\alpha) \\ = & R_n(\hat{f}_\alpha) + (1-\alpha)\lambda \mathrm{TV}(\hat{f}) + \alpha \lambda \mathrm{TV}(f). \end{aligned}$$

Thus,

$$\frac{R_n(\hat{f}) - R_n(\hat{f}_\alpha)}{\alpha} \leq \lambda \mathrm{TV}(f) - \lambda \mathrm{TV}(\hat{f}).$$

The result now follows by letting $\alpha \downarrow 0$. □

Lemma 5.2. *Let $\mathcal{F}$ be a convex subset of $\mathbb{R}^n$ and*

$$\hat{f} := \arg\min_{f \in \mathcal{F}} \left\{ R_n(f) + \lambda \mathrm{TV}(f) \right\}.$$

Then for all $f \in \mathcal{F}$,

$$R(\hat{f}) - R(f) + \mathrm{rem}(f, \hat{f}) \leq \epsilon^T (\hat{f} - f)/n + \lambda \mathrm{TV}(f) - \lambda \mathrm{TV}(\hat{f}),$$

where

$$\mathrm{rem}(f, \hat{f}) = R(f) - R(\hat{f}) - \dot{R}(\hat{f})^T (f - \hat{f}).$$

Proof of Lemma 5.2. By Lemma 5.1,

$$-\dot{R}_n(\hat{f})^T (f - \hat{f}) \leq \lambda \mathrm{TV}(f) - \lambda \mathrm{TV}(\hat{f}).$$

So,

$$\begin{aligned} R(\hat{f}) - R(f) + \mathrm{rem}(f, \hat{f}) = & - \dot{R}(\hat{f})^T (f - \hat{f}) \\ = & (\dot{R}_n(\hat{f}) - \dot{R}(\hat{f}))^T (f - \hat{f}) - \dot{R}_n(\hat{f})^T (f - \hat{f}) \\ = & \epsilon^T (\hat{f} - f)/n - \dot{R}_n(\hat{f})^T (f - \hat{f}) \\ \leq & \epsilon^T (\hat{f} - f)/n + \lambda \mathrm{TV}(f) - \lambda \mathrm{TV}(\hat{f}). \end{aligned}$$

□

One sees from Lemma 6.4 that we need appropriate bounds for the empirical process $\{\epsilon^T f/n : \ f \in \mathbb{R}^n\}$. These will be established in the next two subsections, Subsections 5.2 and 5.3. In Subsection 5.2 we announce the final result, and Subsection 5.3 presents the technicalities that lead to this result.

5.2. **The empirical process $\{\epsilon^T f/n : \ f \in \mathbb{R}^n\}$.** We consider the weights[2]

$$w_k^2 := \begin{cases} \left(\frac{k - t_{j-1}}{d_j} \right) \left(\frac{t_j - k}{n} \right), & t_{j-1} + 1 \leq k \leq t_j - 1, \ j \in [1:r] \\ \frac{1}{n}, & k = t_j, \ j \in [1:s] \end{cases}.$$

For a vector $f \in \mathbb{R}^n$ we define $(Df)_k := f_k - f_{k-1}$ $(k = [2:n])$ so that $\|Df\|_1 = \mathrm{TV}(f)$. Let $w = (w_1, \ldots, w_n)^T$ be the vector of weights and $w^{-1} := (1/w_1, \ldots, 1/w_n)$. Write

$$w_{-S}(Df)_{-S} := \{w_k (Df)_k\}_{k \notin S}.$$

[2]These weights are inspired by the following. Let $\mathcal{V}_S$ be the linear space of functions that are piecewise constant with jumps at S and Π_S be the projection operator on the space $\mathcal{V}_S$. Then

$$\epsilon^T f/n = \epsilon^T \Pi_S f/n + \epsilon^T (I - \Pi_S) f/n,$$

and one can verify that

$$\epsilon^T (I - \Pi_S f)/n = \sum_{k \notin S} V_k (f_k - f_{k-1})$$

where $V_{-S} = \{V_k\}_{k \notin S}$ is a vector of random variables with $\mathrm{var}(V_k) = w_k^2$, $k \notin S$.

We use the notation $\|\cdot\|_{Q_n} := \|\cdot\|_2/\sqrt{n}$ for the normalized Euclidean norm. For $t>0$, let

$$\begin{aligned}\delta_n^2(t) \geq & \left(\frac{4\nu A_0\|w^{-1}\|_{Q_n}}{\sqrt{n}} + 8\sqrt{\frac{1+t+\log(3+2\log_2 n)}{n}}\right)^2 \\ & + \left(\frac{1}{2\nu} + 4\sqrt{\frac{A_0\|w^{-1}\|_{Q_n}/\sqrt{n}}{n}} + \frac{4\sqrt{1+t+\log(3+2\log_2 n)}}{n}\right) \\ & \times \left(\|Dw\|_2 + 2\sqrt{\frac{s}{n}}\right)^2,\end{aligned}$$

and

$$\lambda_n(t) \geq \frac{1}{\sqrt{n}}\left(\frac{4}{\nu} + 8\sqrt{\frac{A_0\|w^{-1}\|_{Q_n}/\sqrt{n}}{n}} + \frac{8\sqrt{1+t+\log(3+2\log_2 n)}}{n}\right).$$

After establishing the material of Subsection 5.3 we are able show the following result:

Theorem 5.1. *Let $\mu>0$ and $t>0$ be arbitrary. With probability at least* $1-\exp[-t]$

$$\epsilon^T f/n \leq \mu\delta_n^2(t) + \frac{\|f\|_{Q_n}^2}{\mu} + \lambda_n(t)\|w_{-S}(Df)_{-S}\|_1,$$

uniformly for all $f\in\mathbb{R}^n$.

Proof of Theorem 5.1. This follows from combining Lemma 5.7 with Lemma 5.6 (see Corollary 5.2). □

5.3. **Material for the result for the empirical process $\{\epsilon^T f/n : \ f\in\mathbb{R}^n\}$ in Theorem 5.1.** For all $f\in\mathbb{R}^n$, let

$$\gamma_f := \frac{\sum_{j=1}^n f_j/w_j}{\|w^{-1}\|_2^2}$$

and let

$$f_{\mathrm{P}} := \Pi_{w^{-1}} f := w^{-1}\gamma_f$$

be the projection of f onto the vector w^{-1}. Define the anti-projection $f_{\mathrm{A}} := (I-\Pi_{w^{-1}})f$.

We let

$$wf := \{w_k f_k\}_{k=1}^n.$$

We start with some preliminary bounds.

Lemma 5.3. *For all $f\in\mathbb{R}^n$,*

$$\|wf-\gamma_f\|_\infty \leq \mathrm{TV}(wf)$$

holds, and

$$\frac{\|f_{\mathrm{A}}\|_\infty}{\mathrm{TV}(wf)} \leq \sqrt{n}.$$

Proof of Lemma 5.3. For all $i\in[1:n]$,

$$\begin{aligned}w_i f_i - \gamma_f =& w_i f_i - \frac{\sum_{k=1} f_k/w_k}{\|w^{-1}\|_2^2} \\ =& \frac{\sum_{k=1}^n (w_i f_i - w_k f_k)/w_k^2}{\|w^{-1}\|_2^2} \leq \mathrm{TV}(wf),\end{aligned}$$

or $\|wf-\gamma_f\|_\infty \leq \mathrm{TV}(wf)$. Since, when $g=wf$,

$$f_{\mathrm{A}} = w^{-1}(g-\gamma_f),$$

we see that

$$\|f_{\mathrm{A}}\|_\infty \leq \|w^{-1}\|_\infty \mathrm{TV}(g) = \|w^{-1}\|_\infty \mathrm{TV}(wf).$$

Since $\|w^{-1}\|_\infty = \sqrt{n}$, we conclude that

$$\|f_{\mathrm{A}}\|_\infty \leq \sqrt{n}\mathrm{TV}(wf).$$ □

We use Dudley's entropy integral to bound the empirical process over $\{f: \ \|f_{\rm A}\|_{Q_n} \le R, \ {\rm TV}(wf) \le 1\}$ with the radius R some fixed value.

Lemma 5.4. *Let $R > 0$ be arbitrary. For all $t > 0$, with probability at least* $1 - \exp[-t]$,

$$\sup_{\|f_{\rm A}\|_{Q_n} \le R, \ {\rm TV}(wf) \le 1} \epsilon^T f/n \quad \le \quad 4\sqrt{\frac{2A_0 \|w^{-1}\|_{Q_n} R}{n}} + 4R\sqrt{\frac{1+t}{n}}.$$

Proof of Lemma 5.4. Let $\mathbf{Q}_w$ be the discrete probability measure that puts mass $w_i^{-2}/\|w^{-1}\|_2^2$ on i, ($i \in [1:n]$). Denote the $L_2(\mathbf{Q}_w)$-norm by $\|\cdot\|_{\mathbf{Q}_w}$. For $\mathcal{G} \subset \mathbb{R}^n$, we let $\mathcal{H}(\cdot, \mathcal{G}, \mathbf{Q}_w)$ denote the entropy of $\mathcal{G}$ for the metric induced by $\|\cdot\|_{\mathbf{Q}_w}$. By Lemma 5.3,

$$\|wf - \gamma_f\|_\infty \le {\rm TV}(wf).$$

Thus by Lemma 4.3, with A_0 the constant given there,

$$\mathcal{H}(u, \{wf - \gamma_f: \quad {\rm TV}(wf) \le 1\}, \mathbf{Q}_w) \le \frac{A_0}{u} \ \forall \ u > 0.$$

For $f \in \mathbb{R}^n$, we have

$$\|f_{\rm A}\|_{Q_n}^2 = \frac{1}{n}\sum_{i=1}^n (w_i f_i - \gamma_f)^2 / w_i^2 = \|wf - \gamma_f\|_{\mathbf{Q}_w}^2 \|w^{-1}\|_{Q_n}^2.$$

Therefore,

$$\mathcal{H}(u, \{f_{\rm A}, \ {\rm TV}(wf) \le 1\}, Q_n) \le \frac{A_0 \|w^{-1}\|_{Q_n}}{u} \ \forall \ u > 0.$$

The entropy integral may therefore be bounded as follows:

$$\begin{aligned} 2\int_0^R & \sqrt{2\mathcal{H}(u, \{f_{\rm A}: \|f_{\rm A}\|_{Q_n} \le R, \ {\rm TV}(wf) \le 1\}, Q_n)} du \\ & \le 4\sqrt{2A_0 \|w^{-1}\|_{Q_n} R}. \end{aligned}$$

By Lemma 4.2 the result follows. □

The next lemma invokes Lemma 5.4 and the peeling device to obtain a result for the weighted empirical process.

Lemma 5.5. *For all $t > 0$, with probability at least* $1 - \exp[-t]$,

$$\begin{aligned} \epsilon^T f_{\rm A}/n \le & 8\sqrt{\frac{A_0 \|w^{-1}\|_{Q_n}}{n}} \left(\sqrt{\|f_{\rm A}\|_{Q_n} {\rm TV}(wf)} \vee \frac{{\rm TV}(wf)}{n^{3/4}} \right) \\ & + 8\left(\|f_{\rm A}\|_{Q_n} \vee \frac{{\rm TV}(wf)}{n^{3/2}} \right) \sqrt{\frac{1 + t + \log(2 + 2\log_2 n)}{n}} \end{aligned}$$

holds uniformly over all f.

Proof of Lemma 5.5. Let $t > 0$ and let $\mathcal{A}$ be the event

$$\begin{aligned} \Bigg\{ \epsilon^T f_{\rm A}/n \ge & 8\sqrt{\frac{A_0 \|w^{-1}\|_{Q_n}}{n}} \sqrt{\|f\|_{Q_n} \vee \frac{1}{n^{3/2}}} \\ & + 8\left(\|f\|_{Q_n} \vee \frac{1}{n^{3/2}} \right) \sqrt{\frac{1 + t + \log(2 + 2\log_2 n)}{n}}, \\ & \text{for some } f \text{ with } \|f\|_{Q_n} \le \sqrt{n} \text{ and } {\rm TV}(wf) \le 1 \Bigg\}. \end{aligned}$$

Let $\mathcal{A}_0$ be the event

$$\left\{ \sup_{\|f_{\rm A}\|_{Q_n} \le \frac{1}{n^{3/2}},\ {\rm TV}(wf)\le 1} \epsilon^T f_{\rm A}/n \le 8\sqrt{\frac{A_0\|w^{-1}\|_{Q_n}}{n}}\sqrt{\frac{1}{n^{3/2}}} \right.$$
$$\left. +\frac{8}{n^{3/2}}\sqrt{\frac{1+t+\log(2+2\log_2 n)}{n}} \right\}.$$

Let $N \in \mathbb{N}$ satisfy $2\log_2 n \le N \le 1 + 2\log_2 n$ and for $j \in [1:N]$ let $\mathcal{A}_j$ be the event

$$\left\{ \sup_{\frac{2^{j-1}}{n^{3/2}} < \|f_{\rm A}\|_{Q_n} \le \frac{2^j}{n^{3/2}},\ {\rm TV}(wf)\le 1} \epsilon^T f_{\rm A}/n \le 8\sqrt{\frac{A_0\|w^{-1}\|_{Q_n}}{n}}\sqrt{\frac{2^{j-1}}{n^{3/2}}} \right.$$
$$\left. +\frac{82^{j-1}}{n^{3/2}}\sqrt{\frac{1+t+\log(2+2\log_2 n)}{n}} \right\}.$$

Application of Lemma 5.4 gives that for all $j \ge 0$,

$$\mathbb{P}(\mathcal{A}_j) \le \exp[-(t+\log(2+2\log_2 n)].$$

Since $\mathcal{A} \subset \cup_{j=0}^N \mathcal{A}_j$, it follows that

$$\mathbb{P}(\mathcal{A}) \le \sum_{j=0}^N \mathbb{P}(\mathcal{A}_j) \le (1+N)\exp[-(t+\log(2+2\log_2 n)] \le \exp[-t].$$

The result now follows by replacing $f_{\rm A}$ by $f_{\rm A}/{\rm TV}(wf)$ and noting that

$${\rm TV}\Big(wf_{\rm A}/TV(wf)\Big) = 1,$$

and invoking from Lemma 5.3 the bound

$$\|f_{\rm A}/{\rm TV}(wf)\|_{Q_n} \le \|f_{\rm A}/{\rm TV}(wf)\|_\infty \le \sqrt{n}. \qquad \square$$

We present a corollary that applies the “conjugate inequality” $2ab \le a^2 + b^2$ (with constants a and b in $\mathbb{R}$), then gathers terms and applies the conjugate inequality again.

Corollary 5.1. *Let $\nu > 0$ and $\mu > 0$ be arbitrary. For all $t > 0$ with probability at least* $1 - \exp[-t]$,

$$\begin{aligned}
&\epsilon^T f_{\rm A}/n \\
\le& \left(\frac{4\nu A_0\|w^{-1}\|_{Q_n}}{\sqrt{n}} + 8\sqrt{\frac{1+t+\log(2+2\log_2 n)}{n}} \right)\|f_{\rm A}\|_{Q_n} \\
+& \left(\frac{4}{\nu} + 8\sqrt{\frac{A_0\|w^{-1}\|_{Q_n}/\sqrt{n}}{n}} + \frac{8\sqrt{1+t+\log(2+2\log_2 n)}}{n} \right)\frac{{\rm TV}(wf)}{\sqrt{n}} \\
\le& \frac{\mu}{2}\left(\frac{4\nu A_0\|w^{-1}\|_{Q_n}}{\sqrt{n}} + 8\sqrt{\frac{1+t+\log(2+2\log_2 n)}{n}} \right)^2 \\
+& \frac{\|f_{\rm A}\|_{Q_n}^2}{2\mu} \\
+& \left(\frac{4}{\nu} + 8\sqrt{\frac{A_0\|w^{-1}\|_{Q_n}/\sqrt{n}/}{n}} + \frac{8\sqrt{1+t+\log(2+2\log_2 n)}}{n} \right)\frac{{\rm TV}(wf)}{\sqrt{n}},
\end{aligned}$$

uniformly for all f.

We now add the missing $f_{\rm P} = f - f_{\rm A}$.

Lemma 5.6. *For all $t > 0$ with probability at least $1 - \exp[-t]$,*

$$\begin{aligned}
&\epsilon^T f/n \\
\leq &\frac{\mu}{2}\left(\frac{4\nu A_0 \|w^{-1}\|_{Q_n}}{\sqrt{n}} + 8\sqrt{\frac{1 + t + \log(3 + 2\log_2 n)}{n}}\right)^2 \\
+ &\frac{\|f\|_{Q_n}^2}{2\mu} \\
+ &\left(\frac{4}{\nu} + 8\sqrt{\frac{A_0 \|w^{-1}\|_{Q_n}/\sqrt{n}}{n}} + \frac{8\sqrt{1 + t + \log(3 + 2\log_2 n)}}{n}\right)\frac{\mathrm{TV}(wf)}{\sqrt{n}},
\end{aligned}$$

uniformly for all f.

Proof of Lemma 5.6. By Pythagoras' rule, we have $\|f\|_2^2 = \|f_{\mathrm{P}}\|_2^2 + \|f_{\mathrm{A}}\|_2^2$. Moreover, by Hoeffding's inequality, with probability at least $1 - \exp[-t]$,

$$\epsilon^T f_{\mathrm{P}}/n \leq \|f_{\mathrm{P}}\|_{Q_n}\sqrt{\frac{2t}{n}} \leq \frac{\mu t}{n} + \frac{\|f_{\mathrm{P}}\|_{Q_n}^2}{2\mu}. \qquad \square$$

In Lemma 5.6, the term including $\mathrm{TV}(wf)$ is almost, but not yet quite the one to be dealt with by the penalty. We bound it by $\|w_{-S}(Df)_{-S}\|_1$ with appropriate remaining terms invoking the "chain rule". Here,

$$w_{-S}(Df)_{-S} := \{w_k (Df)_k\}_{k \notin S}.$$

Lemma 5.7. *For all $f \in \mathbb{R}^n$,*

$$\mathrm{TV}(wf) \leq \sqrt{n}\left(\|Dw\|_2 + 2\sqrt{s/n}\right)\|f\|_{Q_n} + \|w_{-S}(Df)_{-S}\|_1.$$

Proof of Lemma 5.7. We use the fact that

$$\begin{aligned}
\mathrm{TV}(wf) \leq &\sum_{i=2}^n |(w_i - w_{i-1}) f_{i-1}| + \sum_{i=2}^n |w_i (f_i - f_{i-1})| \\
\leq &\|Dw\|_2 \|f\|_2 + \|wDf\|_1.
\end{aligned}$$

Moreover,

$$\|wDf\|_1 = \|w_S (Df)_S\|_1 + \|w_{-S}(Df)_{-S}\|_1$$

with

$$w_S (Df)_S := \{w_k (Df)_k\}_{k \in S},$$

satisfying

$$\begin{aligned}
\|w_S (Df)_S\|_1 = &\sum_{j=1}^s |f_{t_j+1} - f_{t_j}|/\sqrt{n} \\
\leq &\sqrt{s}\sqrt{\sum_{j=1}^s |f_{t_j+1} - f_{t_j}|^2}/\sqrt{n} \\
\leq &2\sqrt{s}\|f\|_2/\sqrt{n}.
\end{aligned}$$

Thus,

$$\mathrm{TV}(wf) \leq \left(\|Dw\|_2 + 2\sqrt{s/n}\right)\|f\|_2 + \|w_{-S}(Df)_{-S}\|_1. \qquad \square$$

Corollary 5.2. *The result from Theorem* 5.1 *now follows by using*

$$\left(\|Dw\|_2 + 2\sqrt{s/n}\right)\|f\|_{Q_n} \leq \frac{\mu}{2}\left(\|Dw\|_2 + 2\sqrt{s/n}\right)^2 + \frac{\|f\|_{Q_n}^2}{2\mu}, \ \ f \in \mathbb{R}^n.$$

5.4. **Bounds for the weights and their inverses.** So far we assumed in this section (see Subsection 5.2), that for $t > 0$, the quantities $\delta_n^2(t)$ and $\lambda_n(t)$ involved in the bound for the empirical process in Theorem 5.1 satisfy

$$\delta_n^2(t) \geq \left(\frac{4\nu A_0 \|w^{-1}\|_{Q_n}}{\sqrt{n}} + 8\sqrt{\frac{1+t+\log(3+2\log_2 n)}{n}} \right)^2$$
$$+\left(\frac{1}{2\nu} + 4\sqrt{\frac{A_0\|w^{-1}\|_{Q_n}/\sqrt{n}}{n}} + \frac{4\sqrt{1+t+\log(3+2\log_2 n)}}{n} \right)$$
$$\times \left(\|Dw\|_2 + 2\sqrt{\frac{s}{n}} \right)^2,$$

and

$$\lambda_n(t) \geq \frac{1}{\sqrt{n}} \left(\frac{4}{\nu} + 8\sqrt{\frac{A_0\|w^{-1}\|_{Q_n}/\sqrt{n}}{n}} + \frac{8\sqrt{1+t+\log(3+2\log_2 n)}}{n} \right),$$

involving $\|w^{-1}\|_{Q_n}$ and $\|Dw\|_2$. In this subsection, we present the bounds for these, leading to the values $\delta_n^2(t)$ and $\lambda_n(t)$ presented in Section 2.

Lemma 5.8. *It holds that*

$$\|w^{-1}\|_2^2 \leq 2n \sum_{d_j \geq 2} (\log(d_j - 1) + 1) + ns \leq n^2 \Delta_n^2$$

and

$$\|Dw\|_2^2 \leq 4 \sum_{d_j \geq 2} (\log(d_j - 1) + 1)/n + s/n =: \Delta_n^2.$$

Proof of Lemma 5.8. We have[3]

$$\|w^{-1}\|_2^2 = \sum_{d_j \geq 2} \sum_{k=1}^{d_j - 1} \frac{n d_j}{k(d_j - k)} + ns$$
$$\leq 2n \sum_{j=1}^{r} (\log(d_j - 1) + 1) + ns.$$

Moreover, for $1 \leq k \leq d_j - 1$, $j \in [1 : r]$,

$$|\sqrt{k}\sqrt{d_j - k} - \sqrt{k-1}\sqrt{d_j - (k-1)}| \leq \sqrt{\frac{d_j - k}{k}} + \sqrt{\frac{k-1}{d_j - k}}$$
$$\leq \sqrt{\frac{d_j - 1}{k}} + \sqrt{\frac{d_j - 2}{d_j - k}} \leq \sqrt{\frac{d_j}{k}} + \sqrt{\frac{d_j}{d_j - k}},$$

so that

$$\sum_{k=1}^{d_j - 1} \frac{|\sqrt{k}\sqrt{d_j - k} - \sqrt{k-1}\sqrt{d_j - (k-1)}|^2}{n d_j}$$
$$\leq \frac{2}{n} \sum_{k=1}^{d_j - 1} \left(\frac{1}{k} + \frac{1}{d_j - k} \right)$$
$$\leq \frac{1}{n} \sum_{j=1}^{r} (4\log(d_j - 1) + 2).$$

[3]We use $\sum_{k=1}^{d-1} \frac{d}{k(d-k)} = \sum_{k=1}^{d-1} \left(\frac{1}{k} + \frac{1}{d-k} \right) = 2\sum_{k=1}^{d-1} \frac{1}{k} \leq 2(1 + \log(d-1))$.

Finally, for $j \in [1:s]$,

$$|w_{t_j} - w_{t_j-1}| = \left| \frac{1}{\sqrt{n}} - \sqrt{\frac{d_j - 1}{d_j}} \frac{1}{\sqrt{n}} \right| \le \frac{1}{\sqrt{n}}. \qquad \square$$

5.5. **A bound for the effective sparsity.** For all $f \in \mathbb{R}^n$, we let

$$(Df)_S := \{(Df)_k\}_{k \in S}, \ (Df)_{-S} := \{(Df)_k\}_{k \notin S}.$$

and recall that

$$w_{-S}(Df)_{-S} := \{w_k (Df)_k\}_{k \notin S}.$$

Let $q_{t_j} := \text{sign}(\mathbf{f}_{t_j})$, $j \in [1:s]$. We define $q_S := \{q_{t_j}\}_{j=1}^s$.

Definition 5.1. Let $\lambda \ge \lambda_n(t)\sqrt{d_{\max}/(2n)}$. The effective sparsity at $\mathbf{f}$ is

$$\Gamma^2(\mathbf{f}, t) := \left(\min \left\{ \|f\|_{Q_n} : \ q^T (Df)_S - \|(1 - w_{-S}\lambda(t)/\lambda)(Df)_{-S}\|_1 = 1 \right\} \right)^{-2}.$$

Recall the definitions

$$J_{\text{monotone}} := \{2 \le j \le s : q_{t_j} = q_{t_{j-1}})\}, \ J_{\text{change}} := [1:r] \backslash J_{\text{monotone}}.$$

Lemma 5.9. *For $\lambda \ge \lambda_n(t)\sqrt{d_{\max}/n}$ we have*

$$\Gamma(\mathbf{f}, t) \le \Gamma_n^2(t),$$

where

$$\Gamma_n^2(t) \ := \ \frac{\lambda_n^2(t)}{\lambda^2} \sum_{j \in J_{\text{monotone}}} 8(\log(d_j) + 1) + \sum_{j \in J_{\text{change}}} \frac{8n(\log(d_j) + 2)}{d_j}.$$

Proof of Lemma 5.9. The proof uses interpolating vectors $q \in \mathbb{R}^n$ as in [13], where $q = (q_1, q_{-1})^T$ is given below. We show that

$$q_S^T (Df)_S - \|(1 - w_{-S}\lambda(t)/\lambda)(Df)_{-S}\|_1 \le q_{-1}^T D(\mathbf{f} - \hat{f}).$$

The result then follows from

$$q_{-1}^T D(\mathbf{f} - f) = (D^T q_{-1})^T (\mathbf{f} - f) \le \|D^T q_{-1}\|_2 \|\mathbf{f} - f\|_2.$$

Furthermore, under the boundary conditions $q_1 = q_n = 0$ we see that $\|D^T q_{-1}\|_2 = \|Dq\|_2$. Define

$$\omega_k^2 := \begin{cases} \left(\frac{k - t_{j-1}}{d_j}\right)\left(\frac{t_j - k}{n}\right)\frac{\lambda_n(t)}{\lambda}, & t_{j-1} + 1 \le k \le t_j - 1, \ j \in J_{\text{monotone}}, \ d_j \ge 2 \\ \left(\frac{k - t_{j-1}}{d_j}\right)\left(\frac{t_j - k}{d_j}\right), & t_{j-1} + 1 \le k \le t_j - 1, \ j \in J_{\text{change}} \\ 0, & k = t_j, \ j \in [1:s] \end{cases}.$$

For $j \in [1:r]$, we let $\bar{t}_j = \frac{t_{j-1} + t_j}{2}$ be the midpoints. Moreover, for $k \notin \{t_1, \ldots, t_s\}$, let

$$q_k := \begin{cases} 0 & 1 \le k < \bar{t}_1 \\ \text{sign}(\mathbf{f}_{t_1})(1 - 2\omega_k), & \bar{t}_1 \le k \le t_1 - 1 \\ \text{sign}(\mathbf{f}_{t_{j-1}})(1 - 2\omega_k), & t_{j-1} + 1 \le k < \bar{t}_j, \ j \in [2:s] \\ \text{sign}(\mathbf{f}_{t_j})(1 - 2\omega_k), & \bar{t}_j \le k \le t_j - 1, \ j \in [2:s] \\ \text{sign}(\mathbf{f}_{t_{r-1}})(1 - 2\omega_k), & t_{r-1} \le k < \bar{t}_r \\ 0 & \bar{t}_r \le k \le n \end{cases}.$$

For $\bar{t}_j - 1 \le k < \bar{t}_j$, $j \in J_1$, we get

$$|1 - 2\omega_k| \le \frac{4}{d_j}.$$

For $j \in J_{\text{monotone}}$, we see that

$$\sum_{k=1}^{d_j} |q_{t_{j-1}+k} - q_{t_{j-1}+k-1}|^2 \le \frac{\lambda_n^2(t)}{\lambda^2} \frac{8(\log d_j + 1)}{n},$$

and for $j \in J_{\text{change}}$,

$$\sum_{k=1}^{d_j} |q_{t_{j-1}+k} - q_{t_{j-1}+k-1}|^2 \le \frac{8(\log d_j + 2)}{d_j}.$$

Thus,

$$\|Dq\|_2^2 \le \frac{\lambda_n^2(t)}{\lambda^2} \sum_{j \in J_{\text{monotone}}} \frac{8(\log(d_j) + 1)}{n} + \sum_{j \in J_{\text{change}}} \frac{8(\log(d_j) + 2)}{d_j}.$$

The lemma now follows from $\Gamma^2(\mathbf{f}, t) \le n\|Dq\|_2^2$. $\square$

5.6. **Finalizing the proof of Theorem 2.1.** By Lemma 6.4, we have

$$\begin{aligned}
&R(\hat{f}) - R(\mathbf{f}) + \mathrm{rem}(\mathbf{f}, \hat{f}) \\
\le& \mu\delta_n^2(t) + \frac{\|\hat{f} - \mathbf{f}\|_{Q_n}}{\mu} + \lambda_n(t)\|w_{-S}(D\hat{f})_{-S}\|_1 + \lambda\|D_S\mathbf{f}\|_1 - \lambda\|D\hat{f}\|_1 \\
=& \mu\delta_n^2(t) + \frac{\|\hat{f} - \mathbf{f}\|_{Q_n}^2}{\mu} + \lambda\Big(\|(D\mathbf{f})_S\|_1 - \|(D\hat{f})_S\|_1 - \|(1 - \lambda_n(t)w_{-S}/\lambda)(D\hat{f})_{-S}\|_1\Big) \\
\le& \mu\delta_n^2(t) + \frac{\|\hat{f} - \mathbf{f}\|_{Q_n}^2}{\mu} + \lambda\Gamma_n(t)\|\hat{f} - \mathbf{f}\|_{Q_n} \\
\le& \mu\delta_n^2(t) + \frac{2\|\hat{f} - \mathbf{f}\|_{Q_n}^2}{\mu} + \frac{\lambda^2}{4}\Gamma_n^2(t).
\end{aligned}$$

Choose $\mu = 4\kappa$ to obtain

$$\frac{2\|\hat{f} - \mathbf{f}\|_{Q_n}^2}{\mu} = \frac{\|\hat{f} - \mathbf{f}\|_{Q_n}^2}{2\kappa} \le \mathrm{rem}(\mathbf{f}, \hat{f}),$$

whenever $\|\hat{f}\|_\infty \le B$. $\square$

6. Proof of Theorem 3.1

6.1. **Some lemmas used in the proof of Theorem 3.1.** The proof of Theorem 3.1 applies some auxiliary lemmas which we develop in this subsection. Define

$$\tau(f) := \|f\|_{Q_n}/(\sqrt{2}K) + (\lambda/\delta)\mathrm{TV}(f)$$

with

$$\delta^2 := 2^4\lambda M_0, \; K^2 := \frac{(1 + \mathrm{e}^{1+2^4 M_0 + \|f^0\|_\infty})^2}{\mathrm{e}^{1+2^4 M_0 + \|f^0\|_\infty}},$$

where $M_0 \ge \mathrm{TV}(f^0) \vee 1$. Moreover, we let

$$\hat{f}_\alpha := \alpha\hat{f} + (1 - \alpha)f^0$$

with

$$\alpha := \frac{\delta}{\delta + \tau(f - f_0)}.$$

Let $\mathcal{F}_0 := \{f : \; \tau(f) \le \delta\}$.

Lemma 6.1. *It holds that $\hat{f}_\alpha - f^0 \in \mathcal{F}_0$, i.e., $\tau(\hat{f}_\alpha - f^0) \le \delta$. Moreover, if in fact $\tau(\hat{f}_\alpha - f^0) \le \delta/2$, then $\hat{f} - f^0 \in \mathcal{F}_0$, as well.*

Proof. We have

$$\tau(\hat{f}_\alpha - f^0) = \alpha\tau(\hat{f} - f^0) = \frac{\delta\tau(\hat{f} - f^0)}{\delta + \tau(\hat{f} - f_0)} \leq \delta.$$

If in fact $\tau(\hat{f}_\alpha - f^0) \leq \delta/2$, we have

$$\tau(\hat{f}_\alpha - f^0) = \frac{\delta\tau(\hat{f} - f^0)}{\delta + \tau(\hat{f} - f_0)} \leq \delta/2$$

which gives $\tau(\hat{f} - f^0) \leq \delta/2 + \tau(\hat{f} - f_0)/2$, or $\tau(\hat{f} - f^0) \leq \delta$. □

Lemma 6.2. *For all $f \in \mathbb{R}^n$,*

$$\|f\|_\infty \leq \|f\|_{Q_n} + \mathrm{TV}(f).$$

Moreover,

$$\mathcal{F}_0 \subset \{f : \ \|f\|_\infty \leq \sqrt{2}K\delta + \delta^2/\lambda, \ \mathrm{TV}(f) \leq \delta^2/\lambda\}.$$

Proof. For $f \in \mathbb{R}^n$, we denote its average by

$$\bar{f} := \frac{1}{n}\sum_{i=1}^n f_i.$$

Then

$$\|f\|_{Q_n}^2 = \bar{f}^2 + \|f - \bar{f}\|_{Q_n} \geq \bar{f}^2.$$

Moreover, for all i,

$$f_i - \bar{f} = \frac{1}{n}\sum_{j=1}^n (f_i - f_j) \leq \mathrm{TV}(f).$$

It follows that

$$\|f\|_\infty \leq \bar{f} + \|f - \bar{f}\|_\infty \leq \|f\|_{Q_n} + \mathrm{TV}(f).$$

For $f \in \mathcal{F}_0$, we have $\|f\|_2/\sqrt{n} \leq \sqrt{2}K\delta$ and $\mathrm{TV}(f) \leq \delta^2/\lambda$, so that also $\|f\|_\infty \leq \sqrt{2}K\delta + \delta^2/\lambda$. □

Lemma 6.3. *Let*

$$K^2 := \frac{(1 + \mathrm{e}^{1+2^4M_0+\|f^0\|_\infty})^2}{\mathrm{e}^{1+2^4M_0+\|f^0\|_\infty}}$$

and let $\delta^2 := 2^4\lambda M_0 \leq 1/(2K^2)$. Then for all f with $f - f^0 \in \mathcal{F}_0$, it is true that $K_f \leq K$.

Proof. Since for $f - f^0 \in \mathcal{F}_0$, $\|f - f^0\|_\infty \leq \sqrt{2}K\delta + \delta^2/\lambda \leq 1 + 2^4M_0$, we see that $\|f\|_\infty \leq 1 + 2^4M_0 + \|f^0\|_\infty$. Therefore,

$$K_f^2 = \frac{(1 + \mathrm{e}^{\|f\|_\infty \vee \|f^0\|_\infty})^2}{\mathrm{e}^{\|f\|_\infty \vee \|f^0\|_\infty}} \leq K^2.$$

□

Lemma 6.4. *We have*

$$0 \leq R(\hat{f}) - R(f^0) \leq \epsilon^T(\hat{f} - f^0)/n + \lambda\mathrm{TV}(f^0) - \lambda\mathrm{TV}(\hat{f}).$$

This inequality is also true with $\hat{f}$ replaced by $\hat{f}_\alpha$.

Proof. For any f,

$$\begin{aligned} 0 \leq R(f) - R(f^0) =& - \left[\Big(R_n(f) - R(f)\Big) - \Big(R_n(f^0) - R(f^0)\Big)\right] \\ &+ R_n(f) - R_n(f^0) \\ =& \epsilon^T(f - f^0)/n + R_n(f) - R_n(f^0). \end{aligned}$$

Insert the basic inequality

$$R_n(\hat{f}) + \lambda\mathrm{TV}(\hat{f}) \leq R_n(f^0) + \lambda\mathrm{TV}(f^0),$$

or

$$R_n(\hat{f}) - R_n(f^0) \leq \lambda\mathrm{TV}(f^0) - \lambda\mathrm{TV}(\hat{f}),$$

to arrive at the first statement of the lemma. To obtain the second statement, we note that by the convexity of $f \mapsto R_n(f)$ such basic inequality is also true for $\hat{f}_\alpha$:

$$\begin{aligned}
&R_n(\hat{f}_\alpha) + \lambda \mathrm{TV}(\hat{f}_\alpha) \\
\leq &\alpha R_n(\hat{f}) + \alpha\lambda\mathrm{TV}(\hat{f}) + (1-\alpha)R_n(f^0) + (1-\alpha)\lambda\mathrm{TV}(f^0) \\
\leq &R_n(f^0) + \lambda\mathrm{TV}(f^0).
\end{aligned}$$

□

6.2. **Proof of Theorem 3.1.** We have for $f \in \mathcal{F}_0$, $\|f\|_\infty \leq \sqrt{2}K\delta + \delta^2/\lambda \leq 2\delta^2/\lambda$ as well as $\mathrm{TV}(f) \leq \delta^2/\lambda \leq 2\delta^2/\lambda$. It follows from Lemma 4.3 that

$$H(u, \mathcal{F}_0, Q_n) \leq \frac{2A_0\delta^2}{\lambda u} \ \forall\ u > 0,$$

so that

$$\begin{aligned}
2\int_0^{\sqrt{2}K\delta} \sqrt{2H(u,\mathcal{F}_0,Q_n)}du \leq &4\sqrt{2A_0\sqrt{2}K}\frac{\delta}{\sqrt{\lambda}}\int_0^{\sqrt{2}K\delta}\frac{1}{\sqrt{u}}du \\
= &8\sqrt{\frac{2A_0\sqrt{2}K}{\lambda}}\delta^{3/2}.
\end{aligned}$$

But then, in view of Lemma 4.2, for all $t > 0$ with probability at least $1 - \exp[-t]$,

$$\sup_{f\in\mathcal{F}_0} \epsilon^T f/n \quad \leq \quad 8\sqrt{\frac{2A_0\sqrt{2}K}{n\lambda}}\delta^{3/2} + 4\sqrt{2}K\delta\sqrt{\frac{1+t}{n}}.$$

Since, by Lemma 6.1, $\hat{f}_\alpha - f^0 \in \mathcal{F}_0$ we know from Lemma 6.3 that $K_{\hat{f}_\alpha} \leq K$. Thus, in view of Lemma 6.4 and the bound

$$R(\hat{f}_\alpha) - R(f^0) \geq \frac{\|\hat{f}_\alpha - f^0\|^2_{Q_n}}{2K^2},$$

we have shown that with probability at least $1 - \exp[-t]$,

$$\begin{aligned}
&\frac{\|\hat{f}_\alpha - f^0\|^2_{Q_n}}{2K^2} + \lambda\mathrm{TV}(\hat{f}_\alpha - f^0) \\
\leq &2\lambda\mathrm{TV}(f^0) + 8\sqrt{\frac{2A_0\sqrt{2}K}{n\lambda}}\delta^{3/2} + 4\sqrt{2}K\delta\sqrt{\frac{1+t}{n}} \\
\leq &\ 2\lambda M_0 \quad + \quad 8\sqrt{\frac{2A_0\sqrt{2}K}{n\lambda}}\delta^{3/2} + 4\sqrt{2}K\delta\sqrt{\frac{1+t}{n}}.
\end{aligned}$$

We want the three terms on the right-hand side to add up to at most $\delta^2/4$. We choose

$$\begin{aligned}
\lambda M_0 = &\delta^2/2^3, \\
8\sqrt{\frac{2A_0\sqrt{2}K}{n\lambda}}\delta^{3/2} \leq &\delta^2/2^4, \\
4\sqrt{2}K\delta\sqrt{\frac{1+t}{n}} \leq &\delta^2/2^4.
\end{aligned}$$

or

$$\begin{aligned}
2^4\lambda M_0 = &\delta^2, \\
\left(\frac{2^7\sqrt{2A_0\sqrt{2}K}}{\sqrt{n\lambda}}\right)^4 \leq &\delta^2, \\
\left(2^6\sqrt{2}K\sqrt{\frac{1+t}{n}}\right)^2 \leq &\delta^2.
\end{aligned}$$

The first one is the largest of the three. This leads to the requirement

$$2^4 \lambda M_0 \geq \left(2^7 \sqrt{\frac{2A_0\sqrt{2}K}{n\lambda}} \right)^4,$$

which is true for

$$\lambda \geq 2^8 n^{-2/3} A_0^{2/3} (\sqrt{2}K)^{2/3},$$

and

$$2^4 \lambda M_0 \geq \left(2^6 \sqrt{2} K \sqrt{\frac{1+t}{n}} \right)^2,$$

which holds for

$$\lambda \geq 2^8 (2K^2) \frac{1+t}{n},$$

where we invoked for both requirements that $M_0 \geq 1$. Then with probability at least $1 - \exp[-t]$,

$$\frac{\|\hat{f}_\alpha - f^0\|_{Q_n}^2}{2K^2} + \lambda \mathrm{TV}(\hat{f}_\alpha - f^0) \leq \delta^2/4.$$

For all $f \in \mathbb{R}^n$,

$$\delta\tau(f) = \frac{\delta\|f\|_{Q_n}}{\sqrt{2}K} + \lambda\mathrm{TV}(f) \leq \delta^2/4 + \frac{\|f\|_2^2/n}{2K^2} + \lambda \mathrm{TV}(f).$$

Thus we have shown that

$$\delta\tau(\hat{f}_\alpha - f^0) \leq \delta^2/4 + \delta^2/4 = \delta^2/2$$

or

$$\tau(\hat{f}_\alpha - f^0) \leq \delta/2.$$

By Lemma 6.1, this implies $\hat{f} \in \mathcal{F}_0$. We can now apply the same arguments to $\hat{f}$ as we did for $\hat{f}_\alpha$ to obtain that with probability at least $1 - 2\exp[-t]$,

$$\frac{\|\hat{f} - f^0\|_{Q_n}^2}{2K^2} + \lambda \mathrm{TV}(\hat{f} - f^0) \leq \delta^2/4 = 4\lambda M_0$$

holds. By Lemma 6.2, this implies

$$\|\hat{f} - f^0\|_\infty \leq \frac{\sqrt{2}K\delta}{2} + \frac{\delta^2}{4\lambda} \leq \frac{1 + 8M_0}{2}. \qquad \square$$

Acknowledgement

We are very grateful to the anonymous reviewer for the helpful comments and careful reading of the manuscript.

References

1. A. Ahmed, E. P. Xing, Recovering time-varying networks of dependencies in social and biological studies. *Proceedings of the National Academy of Sciences* **106** (2009), no. 29, 11878–11883.
2. B. Betancourt, A. Rodríguez, N. Boyd, Bayesian fused lasso regression for dynamic binary networks. *J. Comput. Graph. Statist.* **26** (2017), no. 4, 840–850.
3. S. Chatterjee, S. Goswami, New risk bounds for 2d total variation denoising, (2019). arXiv preprint arXiv:1902.01215.
4. A. S. Dalalyan, M. Hebiri, J. Lederer, On the prediction performance of the Lasso. *Bernoulli* **23** (2017), no. 1, 552–581.
5. B. Fang, A. Guntuboyina, B. Sen, Multivariate extensions of isotonic regression and total variation denoising via entire monotonicity and Hardy-Krause variation, 2019. arXiv preprint arXiv:1903.01395.
6. A. Guntuboyina, D. Lieu, S. Chatterjee, B. Sen, Adaptive risk bounds in univariate total variation denoising and trend filtering. *Ann. Statist.* **48** (2020), no. 1, 205–229.
7. J.-C. Hütter, P. Rigollet, Optimal rates for total variation denoising. In: *Conference on Learning Theory*, pp. 1115–1146, 2016.
8. K. Lin, J. L. Sharpnack, A. Rinaldo, R. J. Tibshirani, A sharp error analysis for the fused lasso, with application to approximate changepoint screening. In: *Advances in Neural Information Processing Systems*, pp. 6884–6893, 2017.
9. C. Liu, H. S. Wong, Structured Penalized Logistic Regression for Gene Selection in Gene Expression Data Analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **16** (2017), no. 1, 312–321.

10. J. Liu, L. Yuan, J. Ye, An efficient algorithm for a class of fused lasso problems. In: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 323–332, 2010.
11. F. Ortelli, S. van de Geer, On the total variation regularized estimator over a class of tree graphs. *Electron. J. Stat.* **12** (2018), no. 2, 4517–4570.
12. F. Ortelli, S. van de Geer, Oracle inequalities for image denoising with total variation regularization, 2019. arXiv preprint arXiv:1911.07231.
13. F. Ortelli, S. van de Geer, Prediction bounds for (higher order) total variation regularized least squares, 2019. arXiv preprint arXiv:1904.10871.
14. O. H. M. Padilla, J. Sharpnack, J. G. Scott, The DFS fused lasso: Linear-time denoising over general graphs. *The Journal of Machine Learning Research* **18** (2017), no. 1, 6410–6445.
15. L. I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991). *Phys. D* **60** (1992), no. 1-4, 259–268.
16. V. Sadhanala, R. J. Tibshirani, Additive models with trend filtering. *Ann. Statist.* **47** (2019), no. 6, 3032–3068.
17. V. Sadhanala, Y.-X. Wang, R. J. Tibshirani, Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers. In: *Advances in Neural Information Processing Systems*, pp. 3513–3521, 2016.
18. V. Sadhanala, Y.-X. Wang, J. L. Sharpnack, R. J. Tibshirani, Higher-order total variation classes on grids: Minimax theory and trend filtering methods. In: *Advances in Neural Information Processing Systems*, pp. 5800–5810, 2017.
19. G. Steidl, S. Didas, J. Neumann, Splines in higher order TV regularization. *International Journal of Computer Vision* **70** (2006), no. 3, 241–255.
20. H. Sun, S. Wang, Penalized logistic regression for high-dimensional DNA methylation data with case-control studies. *Bioinformatics* **28** (2012), no. 10, 1368–1375.
21. R. J. Tibshirani, Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.* **42** (2014), no. 1, 285–323.
22. R. Tibshirani, M. Saunders, S. Rosset, Z. Ji, K. Knight, Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67** (2005) no. 1, 91–108.
23. S. van de Geer, *Estimation and Testing Under Sparsity: École d'Eté de Probabilités de Saint Flour XLV-2016*. Springer Science & Business Media, 2016.
24. A. W. van der Vaart, J. A. Wellner, *Weak Convergence and Empirical Processes*. Springer, 1996.
25. D. Yu, S. J. Lee, W. J. Lee, S. C. Kim, J. Lim, S. W. Kwon, Classification of spectral data using fused lasso logistic regression. *Chemometrics and Intelligent Laboratory Systems* **142** (2015), 70–77.
26. D. Yu, J.-H. Won, T. Lee, J. Lim, S. Yoon, High-dimensional fused lasso regression using majorization-minimization and parallel processing. *J. Comput. Graph. Statist.* **24** (2015), no. 1, 121–153.

(Received 06.03.2020)

Seminar for Statistics, ETH Zürich Rämistrasse 101, 8092 Zürich Switzerland
E-mail address: geer@stat.math.ethz.ch

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 235–246

THE METHOD OF PROBABILISTIC SOLUTION FOR DETERMINATION OF ELECTRIC AND THERMAL STATIONARY FIELDS IN CONIC AND PRISMATIC DOMAINS

MAMULI ZAKRADZE*, MURMAN KUBLASHVILI, NANA KOBLISHVILI, AND ALEKSANDRE CHAKHVADZE

Abstract. In this paper, for determination of the electric and thermal stationary fields the Dirichlet ordinary and generalized harmonic problems are considered. The term "generalized" indicates that a boundary function has a finite number of first kind discontinuity curves. For numerical solution of boundary problems the method of probabilistic solution (MPS) is applied, which in its turn is based on a modeling of the Wiener process. The suggested algorithm does not require an approximation of a boundary function, which is main of its important properties. For examining and to illustrate the effectiveness and simplicity of the proposed method four numerical examples are considered on finding the electric and thermal fields. In the role of domains are taken: finite right circular cone and truncated cone; a rectangular parallelepiped. Numerical results are presented.

1. Introduction

Let D be a finite domain in the Euclidian space R^3, bounded by one closed piecewise smooth surface S (i.e., $S = \bigcup_{j=1}^{p} S^j$), where each part S^j is a smooth surface. Besides, we assume: equations of the parts S^j are given; for the surface S it is easy to show that a point $x = (x_1, x_2, x_3) \in R^3$ lies in $\overline{D}$ or not.

It is known (see, e.g., [1,2,6,12,14–17]) that in practical stationary problems (for example, for the determination of the temperature of the thermal field or the potential of the electric field, and so on) there are cases when it is necessary to consider the Dirichlet ordinary (or generalized) harmonic problems: A (or B).

Problem A. Find a function $u(x) \equiv u(x_1, x_2, x_3) \in C^2(D) \bigcap C(\overline{D})$ satisfying the conditions:

$$\Delta u(x) = 0, \quad x \in D,$$
$$u(y) = h(y), \quad y \in S,$$

where $\Delta = \sum_{i=1}^{3} \frac{\partial^2}{\partial x_i^2}$ is the Laplace operator and $h(y) \equiv h(y_1, y_2, y_3)$ is a continuous function on S.

It is known (see, e.g., [12, 16, 17]) that Problem A is correct, i.e., its solution exists, is unique and depends on data continuously. It should be noted that in general the difficulties and respectively the laboriousness of solving problems sharply increase along with the dimension of the problems considered. Therefore, as a rule, one fails to develop standard methods for solving a wide class of multidimensional problems with the same high accuracy as in the one-dimensional case. For example, the exact solution of Problem A for a circle is written by one-dimensional Poisson's integral and in the case of kernel by two-dimensional Poisson's integral. The given simple example shows the difficulty in determining of the solution with the high accuracy of the Dirichlet ordinary harmonic problem when the dimension increases. In this paper, besides the fact that numerical solution of problems of type A by MPS is interesting and important (see, e.g., [3, 4, 18]), it has an additional role in this paper (see section 3).

2020 *Mathematics Subject Classification.* 34J25, 35J05, 65C30, 65N75.
Key words and phrases. Electric and thermal fields; Dirichlet ordinary and generalized harmonic problems; Method of probabilistic solution; Wiener process.

*Corresponding author.

Problem B. Function $g(y)$ is given on the boundary S of the domain D and is continuous everywhere, except a finite number of curves $l_1, l_2, \dots, l_n$ which represent discontinuity curves of the first kind for the function $g(y)$. It is required to find a function $u(x) \equiv u(x_1, x_2, x_3) \in C^2(D) \bigcap C(\overline{D} \setminus \bigcup_{k=1}^{n} l_k)$ satisfying the conditions:

$$\Delta u(x) = 0, \quad x \in D, \tag{1.1}$$

$$u(y) = g(y), \quad y \in S, \quad y \overline{\in} l_k \subset S \ \ (k = \overline{1,n}), \tag{1.2}$$

$$|u(y)| < c, \qquad y \in \overline{D}, \tag{1.3}$$

where c is a real constant.

It is shown (see [5,20]) that Problem (1.1), (1.2), (1.3) has a unique solution depending continuously on the data, and for a generalized solution $u(x)$ the generalized extremum principal is valid:

$$\min_{x \in S} u(x) < \underset{x \in D}{u(x)} < \max_{x \in S} u(x), \tag{1.4}$$

where for $x \in S$ it is assumed that $x \overline{\in} l_k$ $(k = \overline{1,n})$.

It is evident that actually, the surface S is divided into open parts S_i $(i = \overline{1,m})$ by curves l_k $(k = \overline{1,n})$ or $S = (\bigcup_{i=1}^{m} S_i) \bigcup (\bigcup_{k=1}^{n} l_k)$, where for the concrete case, between m and n from the following conditions: $n = m$, $n < m$, $n > m$ take place one of. On the basis of noted, the boundary function $g(y)$ has the following form

$$g(y) = \begin{cases} g_1(y), & y \in S_1, \\ g_2(y), & y \in S_2, \\ \dots\dots\dots\dots & \\ g_m(y), & y \in S_m, \end{cases} \tag{1.5}$$

where the functions $g_i(y), y \in S_i$ are continuous on the parts S_i of S, respectively.

Note (see [20]) that the additional requirement (1.3) of boundedness concerns actually only the neighborhoods of discontinuity curves of the function $g(y)$ and it plays an important role in the extremum principle (1.4).

On the basis of (1.3), in general, the values of $u(y)$ are not defined on the curves l_k. For example, if Problem B concerns the determination of the thermal (or the electric) field, then $u(y) = 0$ when $y \in l_k$, respectively, in this case, in physical sense the curves l_k are non-conductors (or dielectrics).

Remark 1. If inside the surface S there is a vacuum then we have the ordinary and generalized problems with respect to closed shells.

In general, it is known (see [6,7,20]) that the methods used to obtain an approximate solution to ordinary boundary problems are less suitable (or not suitable at all) for solving boundary problems of type B. In particular, the convergence of the approximate process is very slow in the neighborhood of boundary singularities and, consequently, the accuracy of the approximate solution of the generalized problem is very low.

The choice and construction of computational schemes (algorithms) mainly depend on problem class, its dimension, geometry and location of singularities on the boundary, e.g., Dirichlet generalized plane problems for harmonic functions with concrete location of discontinuity points in the cases of simply connected domains are considered in [1,2,6,8,15], and general cases for finite and infinite domains are studied in [9–11,13,14,19].

In the case of 3D harmonic generalized problems, due to their higher dimension, the difficulties become more significant. In particular, there does not exist a standard scheme which can be applied to a wide class of domains. In the classical literature, simplified, or so called "solvable" generalized problems (problems whose "exact" solutions can be constructed by series, whose terms are represented by special functions) are considered, and for their solution the classical method of separation of variables is mainly applied and therefore the accuracy of the solution is rather low. In the mentioned problems, the boundary functions (conditions) are mainly constants, and in the general case, the analytic form of the "exact" solution is so difficult in the sense of numerical implementation, that it only has theoretical significance (see, e.g., [1, 2, 6, 12, 15]).

As a consequence of the above, from our viewpoint, the construction of high accuracy and effectively realizable computational schemes for approximate solution of 3D Dirichlet generalized harmonic problems (whose application is possible to a wide class of domains) have both theoretical and practical importance.

It should be noted that in literature (see, e.g., [1, 2, 6, 12, 15]), while solving Dirichlet generalized harmonic problems, the existence of discontinuity curves often is neglected. This fact and application of classical methods to solving problems of type B are reasons of the inaccuracies. Therefore, for numerical solution of generalized harmonic problems we should apply such methods which do not require approximation of a boundary function and in which the existence of discontinuity curves is not ignored. The suggested algorithm is one of such methods.

2. The Method of Probabilistic Solution

In this section the essence of the suggested algorithm for numerical solving problems of type A and B is given, and its detail description is in [21]. The main theorem in realization of the MPS is the following one (see, e.g., [5])

Theorem 1. *If a finite domain $D \in R^3$ is bounded by piecewise smooth surface S and $g(y)$ is continuous* (*or discontinuous*) *bounded function on S, then the solution of the Dirichlet ordinary* (*or generalized*) *boundary problem for the Laplace equation at the fixed point $x \in D$ has the form*

$$u(x) = E_x g(x(\tau)). \tag{2.1}$$

In (2.1): $E_x g(x(\tau))$ is the mathematical expectation of the values of the boundary function $g(y)$ at the random intersection points of the trajectory of the Wiener process and the boundary S; τ is the random moment of first exit of the Wiener process $x(t) = (x_1(t), x_2(t), x_3(t))$ from the domain D. It is assumed that the starting point of the Wiener process is always $x(t_0) = (x_1(t_0), x_2(t_0), x_3(t_0)) \in D$, where the value of the desired function is being determined. If the number N of the random intersection points $y^i = (y_1^i, y_2^i, y_3^i) \in S$ $(i = 1, 2, \dots, N)$ is sufficiently large, then according to the law of large numbers, from (2.1) we have

$$u(x) \approx u_N(x) = \frac{1}{N} \sum_{i=1}^{N} g(y^i) \tag{2.2}$$

or $u(x) = \lim u_N(x)$ for $N \to \infty$, in probability. Thus, in the presence of the Wiener process the approximate value of the probabilistic solution to Problems A and B at a point $x \in D$ are calculated by formula (2.2).

In order, to simulate of the Wiener process we use the following recursion relations (see, e.g., [21]):

$$\begin{gathered} x_1(t_k) = x_1(t_{k-1}) + \gamma_1(t_k)/nq, \\ x_2(t_k) = x_2(t_{k-1}) + \gamma_2(t_k)/nq, \\ x_3(t_k) = x_3(t_{k-1}) + \gamma_3(t_k)/nq, \\ (k = 1, 2, \dots), \quad x(t_0) = x, \end{gathered} \tag{2.3}$$

according of which the coordinates of the point $x(t_k) = (x_1(t_k),\ x_2(t_k),\ x_3(t_k))$ are being determined. In (2.3): $\gamma_1(t_k), \gamma_2(t_k), \gamma_3(t_k)$ are three normally distributed independent random numbers for the k-th step, with zero means and variances one; nq is a number of quantification (nq) such that $1/nq = \sqrt{t_k - t_{k-1}}$ and when $nq \to \infty$, then the discrete process approaches to the continuous Wiener process. In the implementation, the random process is simulated at each step of the walk and continues until it crosses the boundary.

In the considered case computations and generation of random numbers are done in MATLAB.

3. Numerical Examples

In this section, problems of type A and B are solved for one and the same domain. The reason of this is the following: since there exist exact test problems for type A, and there are none for type B, therefore, Problem A has an additional role in this paper. Namely, verification of a scheme needed for numerical solution of Problem B and corresponding calculating program is carried out with the help of Problem A, which consists in following.

Function

$$u(x^0, x) = \frac{1}{|x - x^0|}, \quad x \in D, \quad x^0 = (x_1^0, x_2^0, x_3^0) \overline{\in} \overline{D}, \tag{3.1}$$

is taken in the role of the exact test solution for Problem A under boundary condition $h(y) = \frac{1}{|y-x^0|}$, $y \in S$, where $|x - x^0|$ denotes the distance between the points x and x^0. After this, function $h(y)$ is taken in the role of functions $g_i(y)$ $(i = \overline{1,m})$ in Problem B and consequently in calculating program. Evidently, in this case curves l_k represent removable discontinuity curves for function $g_(y)$, therefore instead of problem of type B we have problem of type A. For the obtained problem, verification of the scheme needed for numerical solution of Problem B and corresponding calculating program (comparison of the obtained results with exact solution) is carried out first of all, and then Problem B is being solved under boundary conditions (1.5).

In the case when Problems A and B concern electrostatic field, for full investigation of the field it is necessary to find both potential and strength of the field. It is known [6, 15] that the strength $E(x) = (E_1(x), E_2(x), E_3(x))$ of electrostatic field is defined as follows:

$$E(x) = -\left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \frac{\partial u}{\partial x_3}\right), \quad x \in D, \tag{3.2}$$

where $u(x)$ is potential of electrostatic field. It is known that the vector $E(x)$ is directed where the potential of the electric field is less.

Since in our case Problems A and B are solved by a numerical method, therefore, for the test problem, coordinates of vector $E(x)$ are defined by formula (3.2), and in the case of numerical solution by the central difference formula

$$f'(t) \approx \frac{f(t+h) - f(t-h)}{2h} \tag{3.3}$$

is used, whose accuracy is $O(h^2)$

Thus on the basis of (3.2) and (3.3) for definition components of the vectors $E(x)$ and $E^N(x)$ we have:

$$E_k(x) = -\frac{\partial u(x)}{\partial x_k} = \frac{x_k - x_k^0}{|x - x^0|^3}, \quad (k = 1, 2, 3); \tag{3.4}$$

$$\begin{aligned} E_k^N(x) = -\frac{\partial u_N(x)}{\partial x_k} &\approx -[u_N((x_1 + h)\delta_{1k}, (x_2 + h)\delta_{2k}, (x_3 + h)\delta_{3k}) \\ &- u_N((x_1 - h)\delta_{1k}, (x_2 - h)\delta_{2k}, (x_3 - h)\delta_{3k})]/(2h), \end{aligned} \tag{3.5}$$

where δ_{ik} is Kronecker symbols.

In the present paper the MPS is applied to four examples. In tables, N is the number of the implementation of the Wiener process for the given points $x^i = (x_1^i, x_2^i, x_3^i) \in D$, and nq is the number of the quantification. For simplicity, in the considered examples the values of nq and N are the same. In tables for problems of type A we present the maximum absolute errors Δ^i at the points $x^i \in D$ of $u_N(x)$, in the MPS approximation, for $nq = 200$ and various values of N, and under notations of type $(E \pm k)$, $10^{\pm k}$ are meant. In particular, $\Delta^i = |u_N(x^i) - u(x^0, x^i)|$, where $u_N(x^i)$ is the approximate solution of Problem A at the point x^i, which is defined by formula (2.2), and the exact solution $u(x^0, x^i)$ of the test problem is given by (3.1). In tables, for problems of type B, the probabilistic solution $u_N(x)$ is presented at the points x^i, defined by (2.2).

Remark 2. The Problems A and B for ellipsoidal, spherical, cylindrical domains and for the kernel layer are considered in [21].

Example 3.1. *In the first example it is required to determine the electrostatic field in the domain D. In the role of D is taken interior of the finite right circular cone S_c :*

$$(x_1)^2 + (x_2)^2 - \left(\frac{r}{h}\right)^2 (h - x_3)^2 = 0, \qquad 0 \le x_3 \le h, \tag{3.6}$$

where h is a height of the cone, r is a radius of its base S_1, and $x(x_1, x_2, x_3)$ is a current point of the conic surface S_c (the full surface S of D is $S = S_c \bigcup S_1$).

In numerical experiments for the considered example, is taken: 1) $h = 2$, $r = 1$; 2) in the test Problem A the boundary function $h(y) = 1/|y - x^0|$, $y \in S$, $x^0 = (0, 0, -5)$; 3) in Problem B the boundary function $g(y) \equiv g(y_1, y_2, y_3)$ has the form

$$g(y) = \begin{cases} 2, & y \in S_1 = \{y \in S \; |0 \le (y_1)^2 + (y_2)^2 < 1, y_3 = 0\}, \\ 1.5, & y \in S_2 = \{y \in S_c \; |0 < y_3 < 0.5\}, \\ 1, & y \in S_3 = \{y \in S_c \; |0.5 < y_3 < 1\}, \\ 0.5, & y \in S_4 = \{y \in S_c \; |1 < y_3 \le 2\}, \\ 0, & y \in l_k \; (k = 1, 2, 3). \end{cases} \tag{3.7}$$

It is evident that in the considered case l_1 is the circle of the base S_1; l_2 and l_3 are the circles, which are obtained by intersection of the planes $x_3 = 0.5$, $x_3 = 1$ and the surface S_c. Besides, in the physical sense the circles l_k are non-conductors.

In order to determine the intersection points $y^i = (y_1^i, y_2^i, y_3^i)$ $(i = \overline{1, N})$ of the trajectory of the Wiener process and the surface S, we operate in the following way. During the implementation of the Wiener process, for each current point $x(t_k)$, defined from (2.3), its location with respect to S is checked, i.e., for the point $x(t_k)$ the value

$$d = (x_1(t_k))^2 + (x_2(t_k))^2 - \left(\frac{r}{h}\right)^2 (h - x_3(t_k))^2$$

is calculated and the following conditions: 1) $d = 0$ and $0 < x_3(t_k) < h$; 2) $d < 0$ and $0 < x_3(t_k) < h$; 3) $d < 0$ or $d > 0$ and $x_3(t_k) < 0$; 4) $d > 0$ and $0 < x_3(t_k) < h$ are checked. In the first case $x(t_k) \in S_c$ and $y^i = x(t_k)$. In the second case $x(t_k) \in D$ and the process continuous until it crosses the boundary of D. In the cases (3) and (4) $x(t_k) \overline{\in} \overline{D}$.

Let $x(t_{k-1}) \in D$ for the moment $t = t_{k-1}$ and $x(t_k) \overline{\in} \overline{D}$ for the moment $t = t_k$. In the mentioned case we have only two variants: 3) or 4). In the case 3) we find the intersection point $y = (y_1, y_2, 0)$ of the plane $x_3 = 0$ and a line l passing through the points $x(t_{k-1})$ and $x(t_k)$. If $0 \le (y_1)^2 + (y_2)^2 < r^2$ then $y^i = (y_1, y_2, 0)$. In the case 4), for approximate determination of the point y^i, a parametric equation of a line L passing through the points $x(t_{k-1})$ and $x(t_k)$ is firstly obtained, which has the following form

$$\begin{cases} x_1 = x_1^{k-1} + (x_1^k - x_1^{k-1})\theta, \\ x_2 = x_2^{k-1} + (x_2^k - x_2^{k-1})\theta, \\ x_3 = x_3^{k-1} + (x_3^k - x_3^{k-1})\theta, \end{cases} \tag{3.8}$$

where (x_1, x_2, x_3) is the current point of L and θ is a parameter $(-\infty < \theta < \infty)$, and $x_i^{k-1} \equiv x_i(t_{k-1})$, $x_i^k \equiv x_i(t_k)$ $(i = 1, 2, 3)$. After this, for definition of the intersection points x^* and x^{**} of the line L and the surface S equation (3.6) is solved with respect to θ.

It is easy to see that for parameter θ we obtain an equation

$$A\theta^2 + 2B\theta + C = 0 \tag{3.9}$$

whose discriminant $d^* = B^2 - AC > 0$.

Since the discriminant of (3.9) is positive, the points x^* and x^{**} are defined respectively on the basis of (3.8) for solutions of (3.9) θ_1 and θ_2. In the role of the points y^i the one (from x^* and x^{**}) for which $|x(t_k) - x|$ is minimal is chosen.

In Table 3.1A the errors Δ^i of the approximate solution $u_N(x)$ of the test problem at the points $x^i \in D$ $(i = \overline{1, 5})$ are presented. On the basis of (3.4) and (3.5) we calculated exact and approximate strengths of the electric field (or $E_3(x)$ and $E_3^N(x)$) on the axis Ox_3 at the points x^i $(i = 1, 2, 3)$

Table 3.1A. Results for Problem A (in Example 3.1)

x^i	$(0,0,0.5)$	$(0,0,1)$	$(0,0,1.8)$	$(0.2,0.2,0.5)$	$(-0.2,-0.2,0.5)$
N	Δ^1	Δ^2	Δ^3	Δ^4	Δ^5
$5E+3$	$0.30E-3$	$0.39E-3$	$0.74E-5$	$0.52E-4$	$0.17E-4$
$1E+4$	$0.63E-4$	$0.14E-3$	$0.13E-4$	$0.11E-3$	$0.15E-3$
$5E+4$	$0.98E-4$	$0.72E-4$	$0.40E-4$	$0.25E-4$	$0.72E-4$
$1E+5$	$0.24E-4$	$0.49E-4$	$0.18E-4$	$0.49E-4$	$0.21E-4$
$5E+5$	$0.66E-5$	$0.26E-4$	$0.31E-4$	$0.18E-4$	$0.36E-4$
$1E+6$	$0.32E-5$	$0.42E-4$	$0.31E-4$	$0.15E-4$	$0.27E-4$

for $N = 10^6, nq = 200, h = 0.03$. We obtained the following results: $E_3(0,0,0.5) = 0.033111$; $E_3(0,0,1) = 0.027778$; $E_3(0,0,1.8) = 0.021626$; $E_3^N(0,0,0.5) = 0.033061$; $E_3^N(0,0,1) = 0.027781$; $E_3^N(0,0,1.8) = 0.021651$;

It is evident that the results obtained for $E_3^N(x^i)$ are in good agreement with the values of $E_3(x^i)$ $(i = 1,2,3)$.

Table 3.1B. Results for Problem B (in Example 3.1)

x^i	$(0,0,0.5)$	$(0,0,1)$	$(0,0,1.8)$	$(0.2,0.2,0.5)$	$(-0.2,-0.2,0.5)$
N	$u_N(x^1)$	$u_N(x^2)$	$u_N(x^3)$	$u_N(x^4)$	$u_N(x^5)$
$5E+3$	1.32910	0.77620	0.50010	1.31410	1.31470
$1E+4$	1.33335	0.77370	0.50010	1.32060	1.31690
$5E+4$	1.33374	0.77282	0.50016	1.31599	1.32765
$1E+5$	1.33447	0.77314	0.50035	1.32166	1.32319
$5E+5$	1.33318	0.77234	0.50023	1.32131	1.32318
$1E+6$	1.33356	0.77211	0.50023	1.32223	1.32185

In Table 3.1B the values of the approximate solution $u_N(x)$ to Problem B at the same points $x^i(i = \overline{1,5})$ are given. The boundary function (3.7) is symmetric with respect to the axis Ox_3, respectively, the obtained results for x^4 and x^5 are symmetric with respect to the axis Ox_3 and have sufficient accuracy for many practical problems.

For illustration, we calculated the electrostatic field strength by (3.5) on the axis Ox_3 at the same points $x^i(i = 1,2,3)$ for $N = 10^6, nq = 200, h = 0.03$. We obtained the following results: $E_3^N(0,0,0.5) = 1.234714$; $E_3^N(0,0,1) = 0.989775$; $E_3^N(0,0,1.8) = 0.00426$. The obtained results are in good agreement with the real physical picture.

Example 3.2. *In this example, the problem on the temperature distribution is considered. In the role of domain D the interior of a truncated right circular cone S_c is taken:*

$$(x_1)^2 + (x_2)^2 - \Big(\frac{R-r}{h}\Big)^2\Big(\frac{Rh}{R-r} - x_3\Big)^2 = 0,\ 0 \le x_3 \le h,$$

where h is the height, R the radius of the lower base, r is the radius of the upper base, and $x(x_1,x_2,x_3)$ is a current point of the conic surface S_c. The boundary of D is $S = S_1 \bigcup S_c \bigcup S_2$, where $S_1 = \{y \in S\ |0 \le d < R, y_3 = 0\}$ and $S_2 = \{y \in S\ |0 \le d < r, y_3 = h\}$, and $d = sqrt((y_1)^2 + (y_2)^2)$.

The problems A and B are solved when $h = 2$, $R = 1$, $r = 0.5$, $x^0 = (0, 0, -5)$, and the boundary function $g(y)$ has the form

$$g(y) = \begin{cases} 2, & y \in S_1, \\ 0, & y \in S_2, \\ 1.5, & y \in S_3, \\ 1, & y \in S_4, \\ 1.5, & y \in S_5, \\ 1, & y \in S_6, \\ 0, & y \in l_k \ (k = \overline{1,6}). \end{cases} \tag{3.10}$$

In (3.10): l_1, l_2 are the circles of the bases S_1 and S_2; l_3, l_4, l_5, l_6 are the generatices of the conic surface S_c, which pass through the points $(R, 0)$, $(0, R)$, $(-R, 0)$, $(0, -R)$, respectively; $S_3 = \{y \in S_c \, |r < d < R, y_1 > 0, y_2 > 0, 0 < y_3 < h\}$; $S_4 = \{y \in S_c \, |r < d < R, y_1 < 0, y_2 > 0, 0 < y_3 < h\}$; $S_5 = \{y \in S_c \, |r < d < R, y_1 < 0, y_2 < 0, 0 < y_3 < h\}$; $S_6 = \{y \in S_c \, |r < d < R, y_1 > 0, y_2 < 0, 0 < y_3 < h\}$. Besides, in this case the curves l_k and S_2 are non-conductors.

In the considered case, for determination of the intersection points y^i $(i = \overline{1, N})$ of the trajectory of the Wiener process and the surface S the same algorithm, described in Example 3.1 is applied.

In Table 3.2A the errors Δ^i of the approximate solution $u_N(x)$ of the test problem are presented at the points $x^i \in D$ $(i = \overline{1, 5})$.

Table 3.2A. Results for Problem A (in Example 3.2)

x^i	$(0,0,0.5)$	$(0,0,1)$	$(0,0,1.8)$	$(0.5,0.5,1)$	$(-0.5,-0.5,1)$
N	Δ^1	Δ^2	Δ^3	Δ^4	Δ^5
$5E+3$	$0.99E-4$	$0.27E-3$	$0.99E-4$	$0.58E-4$	$0.10E-3$
$1E+4$	$0.66E-4$	$0.11E-3$	$0.88E-4$	$0.48E-4$	$0.58E-4$
$5E+4$	$0.65E-4$	$0.52E-4$	$0.90E-4$	$0.27E-4$	$0.30E-4$
$1E+5$	$0.40E-4$	$0.26E-4$	$0.25E-4$	$0.17E-4$	$0.48E-4$
$5E+5$	$0.16E-4$	$0.19E-4$	$0.55E-4$	$0.24E-4$	$0.23E-4$
$1E+6$	$0.81E-5$	$0.28E-4$	$0.53E-4$	$0.27E-4$	$0.25E-4$

The values of the approximate solution $u_N(x)$ of Problem B at the same points x^i are given in Table 3.2B. Since the boundary function (3.10) is symmetric with respect to the axis Ox_3, therefore, for control in the role of x^i $(i = 4, 5)$, the points which are symmetric with respect to the axis Ox_3 are taken. The obtained results have sufficient accuracy for many practical problems and are in good agreement with the real physical picture.

Table 3.2B. Results for Problem B (in Example 3.2)

x^i	$(0,0,0.5)$	$(0,0,1)$	$(0,0,1.8)$	$(0.5,0.5,1)$	$(-0.5,-0.5,1)$
N	$u_N(x^1)$	$u_N(x^2)$	$u_N(x^3)$	$u_N(x^4)$	$u_N(x^5)$
$5E+3$	1.53710	1.29085	0.56980	1.48650	1.48660
$1E+4$	1.52895	1.29140	0.56875	1.48575	1.48580
$5E+4$	1.52950	1.28705	0.57235	1.48548	1.48538
$1E+5$	1.52615	1.28884	0.57426	1.48584	1.48558
$5E+5$	1.52680	1.28701	0.57756	1.48578	1.48579
$1E+6$	1.52670	1.28710	0.57615	1.48558	1.48566

Example 3.3. *Here in the role of domain D the interior of rectangular parallelepiped $MNKOM_1N_1K_1O_1$ is taken with the vertex at the origin $O(0, 0, 0)$ of Cartesian coordinate right-handed system and measurements a, b and c. It is evident that the boundary S of D is $S =$*

$(\bigcup\limits_{j=1}^{6} S_j)\bigcup(\bigcup\limits_{k=1}^{12} l_k)$, *where S_j are open faces and l_k are edges. In this example, the problem on the temperature distribution is considered.*

In order to determine the intersection points $y^i = (y_1^i, y_2^i, y_3^i)$ $(i = \overline{1,N})$ of the trajectory of the Wiener process and the surface S of mentioned parallelepiped the following way is used. During the implementation of the Wiener process, for each current point $x(t_k)$, defined by (2.3), its location with respect to S is checked, i.e., for the point $x(t_k)$ the following conditions

$$0 < x_1(t_k) < a, \quad 0 < x_2(t_k) < b, \quad 0 < x_3(t_k) < c$$

are checked. If the mentioned conditions are fulfilled then the process (2.3) continuous. If $x(t_k) \in S$ then $y^i = x(t_k)$.

Let $x(t) \in D$ for the moment $t = t_{k-1}$ and $x(t) \overline{\in} \overline{D}$ for the moment $t = t_k$. In this case, for approximate determination of the point y^i, a parametric equation of a line L passing through the points $x(t_{k-1})$ and $x(t_k)$ is firstly obtained in the form (3.8). After this, the intersection point x^* of the line L and that face, which is intersected by the trajectory of wiener process is found and respectively, in this case $y^i = x^*$.

In numerical experiments, we took: 1) $a = 1$, $b = 2$, $c = 3$; 2) in the test Problem A, $x^0 = (0.5, 1, -5)$; 3) in Problem B the boundary function $g(y)$ has the following form

$$g(y) = \begin{cases} 3, & y \in S_1 = \{y \in S \mid y_1 = 0, 0 < y_2 < b, 0 < y_3 < c\}, \\ 1, & y \in S_2 = \{y \in S \mid y_1 = a, 0 < y_2 < b, 0 < y_3 < c\}, \\ 0.5, & y \in S_3 = \{y \in S \mid 0 < y_1 < a, y_2 = 0, 0 < y_3 < c\}, \\ 0.5, & y \in S_4 = \{y \in S \mid 0 < y_1 < a, y_2 = b, 0 < y_3 < c\}, \\ 0, & y \in S_5 = \{y \in S \mid 0 < y_1 < a, 0 < y_2 < b, y_3 = 0\}, \\ 2, & y \in S_6 = \{y \in S \mid 0 < y_1 < a, 0 < y_2 < b, y_3 = c\}, \\ 0, & y \in l_k \ (k = \overline{1,12}), \end{cases} \tag{3.11}$$

where l_k and S_5 are dielectrics.

The errors Δ^i of the approximate solution $u_N(x)$ to test Problem A at the points $x^i \in D$ $(i = \overline{1,5})$ are given in Table 3.3A.

TABLE 3.3A. Results for Problem A (in Example 3.3)

x^i	$(0.5, 1, 0.5)$	$(0.5, 1, 1)$	$(0.5, 1, 1.5)$	$(0.5, 1, 2)$	$(0.5, 1, 2.5)$
N	Δ^1	Δ^2	Δ^3	Δ^4	Δ^5
$5E+3$	$0.11E-3$	$0.28E-3$	$0.17E-3$	$0.14E-3$	$0.14E-3$
$1E+4$	$0.73E-4$	$0.28E-4$	$0.25E-3$	$0.77E-4$	$0.13E-4$
$5E+4$	$0.57E-4$	$0.66E-4$	$0.40E-4$	$0.48E-4$	$0.12E-4$
$1E+5$	$0.17E-4$	$0.20E-4$	$0.61E-4$	$0.29E-4$	$0.15E-4$
$5E+5$	$0.35E-4$	$0.32E-4$	$0.25E-4$	$0.15E-4$	$0.21E-4$
$1E+6$	$0.65E-5$	$0.43E-5$	$0.49E-5$	$0.19E-4$	$0.19E-4$

The values of the approximate solution $u_N(x)$ of Problem B at the points $x^i \in D$ $(i = 1, 2, 3)$ are given in Table 3.3B. Since the boundary function (3.11) is symmetric with respect to the plane $x_2 = 1$, therefore, for control in the role of x^i $(i = 4, 5)$, the points which are symmetric with respect to the plane $x_2 = 1$ are taken. The obtained results have sufficient accuracy for many practical problems and are in good agreement with the real physical picture (see Table 3.3B).

Example 3.4. *In this example, the problems A and B on the temperature distributionn are considered. In the role of D we took the same rectangular parallelepiped as in Example* 3.3. *In this case, Problem*

B under the boundary function $g(y)$ with specific form

$$g(y) = \begin{cases} v_1, & y \in S_1 = \{y \in S \mid y_1 = 0, 0 < y_2 < b, 0 < y_3 < c\}, \\ v_2, & y \in S_2 = \{y \in S \mid y_1 = a, 0 < y_2 < b, 0 < y_3 < c\}, \\ 0, & y \in (\bigcup_{j=3}^{6} S_j) \bigcup (\bigcup_{k=1}^{12} l_k) \end{cases} \tag{3.12}$$

is solved, where $S_j (j = \overline{1,6}), l_k (k = \overline{1,12})$ are the same as in Example 3.3, v_1 and v_2 are constants. It is evident that S_j $(j = \overline{3,6})$ and l_k $(k = \overline{1,12})$ are non-conductors.

Table 3.3B. Results for Problem B (in Example 3.3)

x^i	$(0.5, 1, 0.5)$	$(0.5, 1, 1.5)$	$(0.5, 1, 2.5)$	$(0.5, 0.5, 1.5)$	$(0.5, 1.5, 1.5)$
N	$u_N(x^1)$	$u_N(x^2)$	$u_N(x^3)$	$u_N(x^4)$	$u_N(x^5)$
$5E+3$	1.36592	1.77016	1.87272	1.51024	1.50580
$1E+4$	1.37656	1.79118	1.87680	1.48596	1.49162
$5E+4$	1.38256	1.79529	1.87079	1.49502	1.50417
$1E+5$	1.37245	1.79176	1.87154	1.51296	1.49943
$5E+5$	1.36930	1.78991	1.86844	1.50475	1.50233
$1E+6$	1.37228	1.78882	1.86735	1.50309	1.50362

In numerical experiments we took: $a = 1$, $b = 2$, $c = 3$, $v_1 = 3$, $v_2 = 1$ and $x^0 = (0.5, 1, -5)$. For determination of the intersection points y^i $(i = \overline{1,N})$ the same algorithm is applied, which is described in Example 3.3.

In Table 3.4A the errors Δ^i of the approximate solution $u_N(x)$ of the test problem A are presented at the points $x^i \in D$ $(i = \overline{1,5})$. The obtained results have sufficient accuracy for many practical problems.

Table 3.4A. Results for Problem A (in Example 3.4)

x^i	$(0.9, 1, 1.5)$	$(0.8, 1, 1.5)$	$(0.5, 1, 1.5)$	$(0.2, 1, 1.5)$	$(0.1, 1, 1.5)$
N	Δ^1	Δ^2	Δ^3	Δ^4	Δ^5
$5E+3$	$0.17E-3$	$0.48E-4$	$0.19E-3$	$0.17E-3$	$0.53E-4$
$1E+4$	$0.40E-4$	$0.14E-4$	$0.69E-4$	$0.23E-3$	$0.69E-5$
$5E+4$	$0.52E-4$	$0.46E-4$	$0.14E-6$	$0.39E-4$	$0.25E-4$
$1E+5$	$0.23E-4$	$0.73E-5$	$0.18E-4$	$0.51E-4$	$0.51E-5$
$5E+5$	$0.72E-5$	$0.27E-4$	$0.19E-4$	$0.19E-5$	$0.27E-5$
$1E+6$	$0.69E-5$	$0.56E-5$	$0.19E-4$	$0.14E-4$	$0.38E-5$

The values of the approximate solution $u_N(x)$ of Problem B at the same points $x^i \in D$ are given in Table 3.4B. The obtained results have sufficient accuracy for many practical problems and are in good agreement with the real physical picture.

It should be noted that Example 3.4 is considered in [2], where it is solved by the method of separation of variables. It is shown that in conditions (3.12) the analytical solution to Problem B has the following form

$$u(x) \equiv u(x_1, x_2, x_3) = \frac{16}{\pi^2} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \frac{f_1(x_1) f_2(x_2, x_3)}{(2p+1)(2q+1)}, \tag{3.13}$$

where

$$f_1(x_1) = \frac{v_1 sh(l(a - x_1)) + v_2 sh(l x_1)}{sh(la)},$$

$$f_2(x_2,x_3) = sin\frac{(2p+1)\pi x_2}{b} sin\frac{(2q+1)\pi x_3}{c},$$
$$l = \frac{\pi}{bc}\sqrt{(c(2p+1))^2+(b(2q+1))^2},$$

and $sh(t)$ is hyperbolic sine.

Table 3.4B. Results for Problem B (in Example 3.4)

x^i	$(0.9, 1, 1.5)$	$(0.8, 1, 1.5)$	$(0.5, 1, , 1.5)$	$(0.2, 1, 1.5)$	$(0.1, 1, 1.5)$
N	$u_N(x^1)$	$u_N(x^2)$	$u_N(x^3)$	$u_N(x^4)$	$nu_N(x^5)$
$5E+3$	1.13260	1.26040	1.75080	2.46040	2.73280
$1E+4$	1.12920	1.27440	1.74840	2.44380	2.70770
$5E+4$	1.12640	1.25490	1.74864	2.44730	2.71620
$1E+5$	1.12798	1.25773	1.74749	2.45006	2.71432
$4E+5$	1.12858	1.26055	1.75092	2.44749	2.71528
$1E+6$	1.12824	1.25857	1.75128	2.44603	2.71731

It is easy to see that the series (3.13) converges rapidly for all points $x = (x_1, x_2, x_3) \in D$, when $p, q \to \infty$. In order to compare the results obtained by the MPS and the (3.13), the partial sum $u_m(x)$ of the series (3.13) for $p = \overline{0, m}$ and $q = \overline{0, m}$ at the points x^i $(i = \overline{1, 5})$ were calculadet (see Table 3.4B). Because of rapid convergence of the series (3.13) when $x \in D$, the calculations have shown that practically $u_m(0.9, 1, 1.5) = 1.12524$, $u_m(0.8, 1, 1.5) = 1.25747$, $u_m(0.5, 1, 1.5) = 1.75388$, $u_m(0.2, 1, 1.5) = 2.45277$, $u_m(0.1, 1, 1.5) = 2.72234$, when $m = 50, 100, 150$. These results are sufficiently close to results which are presented in Table 3.4B.

It is evident that for the solution $u(x)$ the boundary condition (3.12) is satisfied on $(\bigcup_{j=3}^{6} S_j)\bigcup(\bigcup_{k=1}^{12} l_k)$. If $x \in S_1 \bigcup S_2$, then the rate of convergence of (3.13) becomes worse, especially in the neighborhood of the discontinuity curves. In particular, the convergence is very slow and consequently, the accuracy in the satisfaction of boundary condition on $S_1 \bigcup S_2$ is very low (see Section 1). This is caused by the fact that, when $x \in S_1 \bigcup S_2$ and tends to the discontinuity curves (edges), then all the terms of the series (3.13) tend to zero.

Table 3.4C. Results for partial sum $u_m(x)$

i	x^i	$u_m(x), m = 50$	$u_m(x), m = 100$	$u_m(x), m = 150$
1	$(1, 1, 1.5)$	0.987309	0.993644	0.995760
2	$(1, 1.8, 1.5)$	0.973200	0.986554	0.991000
3	$(1, 1.9, 1.5)$	1.033759	0.976574	1.011400
4	$(1, 1.99, 1.5)$	0.867110	1.175235	1.021747
5	$(1, 1.999, 1.5)$	0.099228	0.198273	0.295694
6	$(1, 1, 2.99)$	0.623372	1.044791	1.176482
7	$(1, 1, 2.999)$	0.662021	0.132586	0.198485
8	$(1, 1.99, 2.99$	0.547480	1.235730)	1.207186
9	$(1, 1.999, 2.999)$	0.006653	0.026456	0.058941
10	$(1, 0.001, 0.5)$	0.100479	0.199549	0.295065

From Table 3.4C it is clear that accuracy of the solution $u(x)$ is very low in the neighborhood of the discontinuity curves, as expected.

Remark 3. if V_1 or V_2 is not constant then the analytic form of the solution is so difficult in the sense of numerical implementation, that it has only theoretical significance (see [1]).

In this work the problems of type B are specially solved when boundary functions $g_i(y)$ $(i=\overline{1,m})$ are constants. This was caused by our interest to find out how much the obtained results were in agreement with real physical picture. It is evident that solving Problem B under condition (1.5) is as easy as Problem A. In general, Problem B can be solved for all such locations of discontinuity curves, which give the possibility to establish the part of surface S where the intersection point is located.

The analysis of the results of numerical experiments show that the results obtained by the suggested algorithm are reliable and it is effective for numerical solution of problems of type A and B. In particular, the algorithm is sufficiently simple for numerical implementation.

Besides, it should be noted that the accuracy of probabilistic solution of problems A and B is not significantly increasing (except some cases, see tables) when $N \to \infty$. It is caused by the fact that nq (the number of the quantification) is fixed. If more accuracy is needed then calculations for sufficiently large values of nq and N (see [20]) must be realized. In this case, numerical realization on a PC takes much time. This difficulty can be avoided by applying the method of parallel calculation. For this suitable computing technique is needed. Respectively, significantly less time will be needed for numerical realization and besides the accuracy of the obtained results will improve.

4. Concluding Remarks

1. In this work have demonstrated that the method of probabilistic solution(MPS) is ideally suited for numerical solving of both ordinary and generalized(2D and 3D) Dirichlet problems for rather a wide class of domains, in the case of Laplace equation.

2. The MPS does not require an approximation of a boundary function, which is one of its important properties.

3. The MPS is a fast solver for the above noted problems. Besides, it is easy to programme, its computational cost is low, it characterized by an accuracy which is sufficient for many problems.

References

1. B. M. Budak, A. A. Samarski, A. N. Tikhonov, *A Collection of Problems in Mathematical Physics.* (Russian) Third edition. Nauka, Moscow, 1980.
2. H. S. Carslaw, J. C. Jaeger, *Conduction of Heat in Solids.* Oxford University Press, London, 1959.
3. A. Sh. Chaduneli, Z. A. Tabagari, M. Zakradze, On solving the internal three-dimensional Dirichlet problem for a harmonic function by the method of probabilistic solution. *Bull. Georgian Natl. Acad. Sci. (N.S.)* **2** (2008), no. 1, 25–28.
4. A. Sh. Chaduneli, M. V. Zakradze, Z. A. Tabagari, A method of probabilistic solution to the ordinary and generalized plane Dirichlet problem for the Laplace equation. Science and Computing. In: *Proc. Sixth TSTC Scientific Advisory Committee Seminar.* vol. 2, pp. 361–366, 2003.
5. E. B. Dynkin, A. A. Juškevič, *Theorems and Problems on Markov Processes.* (Russian) Nauka, Moscow, 1967.
6. G. A. Grinberg, *Selected Problems of the Mathematical Theory of Electric and Magnetic Phenomena.* (Russian) Izd. Akad. Nauk SSSR, 1948.
7. L. V. Kantorovich, V. I. Krylov, *Approximate Methods of Higher Analysis.* (Russian) Fifth corrected edition Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow-Leningrad 1962.
8. A. Karageorghis, Modified methods of fundamental solutions for harmonic and biharmonic problems with boundary singularities. *Numer. Methods Partial Differential Equations* **8** (1992), no. 1, 1–19.
9. N. Koblishvili, M. Zakradze, On solving the Dirichlet generalized problem for a harmonic function in the case of infinite plane with holes. *Proc. A. Razmadze Math. Inst.* **164** (2014), 71–82.
10. N. Koblishvili, Z. Tabagari, M. Zakradze, On reduction of the Dirichlet generalized boundary value problem to an ordinary problem for harmonic function. *Proc. A. Razmadze Math. Inst.* **132** (2003), 93–106.
11. N. Koblishvili, M. Kublashvili, Z. Sanikidze, M. Zakradze, On solving the Dirichlet generalized problem for a harmonic function in the case of an infinite plane with a crack-type cut. *Proc. A. Razmadze Math. Inst.* **168** (2015), 53–62.
12. N. S. Koshlyakov, E. B. Gliner, M. M. Smirnov, *Equations in Partial Derivatives of Mathematical Physics.* (Russian) Vysshaya Shkola, Moscow, 1970,
13. M. Kublashvili, Z. Sanikidze, M. Zakradze, A method of conformal mapping for solving the generalized Dirichlet problem of Laplace's equation. *Proc. A. Razmadze Math. Inst.* **160** (2012), 71–89.
14. M. A. Lavrentjev, B. V. Shabat, *Methods of the Theory of Functions of a Complex Variable.* (Russian) Nauka, Moscow, 1973.
15. W. R. Smythe, *Static and Dynamic Electricity.* second edition. New York, Toronto, London, 1950.
16. A. N. Tikhonov, A. A. Samarskii, *The Equations of Mathematical Physics.* (Russian) Fourth edition, corrected. Nauka, Moscow, 1972.

17. V. S. Vladimirov, *Equations of Mathematical Physics.* (Russian) Second edition, revised and augmented. Nauka, Moscow, 1971.
18. M. Zakradze, Z. Sanikidze, Z. Tabagari, On solving the external three-dimensional Dirichlet problem for a harmonic function by the probabilistic method. *Bull. Georgian Natl. Acad. Sci. (N.S.)* **4** (2010), no. 3, 19–23.
19. M. Zakradze, N. Koblishvili, A. Karageorghis, Y. Smyrlis, On solving the Dirichlet generalized problem for harmonic function by the method of fundamental solutions. *Semin. I. Vekua Inst. Appl. Math. Rep.* **34** (2008), 24–32, 124.
20. M. Zakradze, M. Kublashvili, Z. Sanikidze, N. Koblishvili, Investigation and numerical solution of some 3D internal Dirichlet generalized harmonic problems in finite domains. *Trans. A. Razmadze Math. Inst.* **171** (2017), no. 1, 103–110.
21. M. Zakradze, B. Mamporia, M. Kublashvili, N. Koblishvili, The method of probabilistic solution for 3D Dirichlet ordinary and generalized harmonic problems in finite domains bounded with one surface. *Trans. A. Razmadze Math. Inst.* **172** (2018), no. 3, part A, 453–465.

(Received 22.07.2019)

Georgian Technical University, N. Muskhelishvili Institute of Computational Mathematics, Tbilisi, Georgia

E-mail address: mamuliz@yahoo.com
E-mail address: mkublashvili@mail.ru
E-mail address: nanakoblishvili@yahoo.com
E-mail address: aleqsandre92@gmail.com

Transactions of A. Razmadze Mathematical Institute
Volume 174, 2020, issue 2, 249–259

PRELIMINARY COMMUNICATIONS

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 249–250

ON DOUBLE FOURIER SERIES WITH RESPECT TO THE CLASSICAL REARRANGEMENTS OF THE WALSH–PALEY SYSTEM

ROSTOM GETSADZE

Abstract. The following theorem is established: there exists a continuous function on $[0,1]^2$ with a certain smoothness, whose double Fourier-Walsh series diverges by rectangles on a set of positive measure. Similar theorem is true also for the double Walsh–Kaczmarz system.

1. Introduction

There are two classical rearrangements of the Walsh–Paley system: (a) the Walsh system and (b) the Walsh–Kaczmarz system. It is well-known (see [3, 4]) these systems are systems of convergence. The system of Rademacher functions $\{r_n(x)\}_{n=0}^{\infty}$ on $[0,1)$ is defined as follows. Set

$$r_0(x) = \begin{cases} 1 & \text{for } 0 \le x < \frac{1}{2}, \\ -1 & \text{for } \frac{1}{2} \le x < 1. \end{cases}$$

We extend the function $r_0(x)$ on $(-\infty, \infty)$ with period 1. For $n \ge 1$, we set

$$r_n(x) = r_0(2^n x).$$

For each $k \in N = \{0, 1, 2, \dots\}$, we introduce a function $\alpha_k :\ [0,1) \to \{0,1\}$ defined by the dyadic expansion of x

$$x = \sum_{k=0}^{\infty} \frac{\alpha_k(x)}{2^{k+1}}.$$

If x is a dyadic rational, then we suppose that its dyadic expansion contains infinitely many zeros.

The Walsh–Paley system of functions $\{W_n(x)\}_{n=0}^{\infty}$ on $[0,1)$ is defined as follows. Set $W_0(x) = 1$ for all $x \in [0,1)$. For $n \ge 1$, we consider the dyadic representation $n = 2^{m_1} + 2^{m_2} + \dots + 2^{m_q}$, $(n \ge 1$, $m_1 > m_2 \dots > m_q \ge 0)$ and set

$$W_n(x) = r_{m_1}(x) r_{m_2}(x) \dots r_{m_q}(x) \qquad x \in [0,1).$$

The modulus of continuity $\omega(F;\delta)$ of a continuous function F on $[0,1]^2$ is defined by

$$\omega(F;\delta) = \sup_{\sqrt{(x_1-x_2)^2+(y_1-y_2)^2} \le \delta} \{|F(x_1,y_1) - F(x_2,y_2)|,\ (x_1,y_1),\ (x_2,y_2) \in [0,1]^2\}.$$

Recently [2], we have proved the following

Theorem 1. *There exists a continuous function F on $[0,1]^2$ such that*

$$\omega(F;\delta) = O\left(\frac{1}{\sqrt{\log_2 \frac{1}{\delta}}}\right), \qquad \delta \to 0+,$$

and the Fourier series of F with respect to the double Walsh–Paley system $\{W_m(x)W_n(y)\}_{m,n=0}^{\infty}$ diverges on a set of positive measure by rectangles.

2020 *Mathematics Subject Classification.* 42C10.

Key words and phrases. Walsh system; Rearrangements; Double Fourier series.

The Walsh system $\{\varphi_m(x)\}_{m=0}^{\infty}$ was introduced by Walsh (see, e.g., [4]) and defined as follows:

$$\varphi_0(x) = 1, \quad \varphi_1(x) = (-1)^{\alpha_0(x)}, \quad \varphi_{2^n}(x) = (-1)^{\alpha_{n-1}(x)+\alpha_n(x)},$$
$$\varphi_{2^n+k}(x) = \varphi_{2^n}(x)\varphi_k(x), \quad k = 0, 1, \dots, 2^n - 1; \quad n = 0, 1, \dots,$$

To define the Walsh–Kaczmarz system $\{h_m(x)\}_{m=1}^{\infty}$, we first introduce an auxiliary system of functions

$$\psi_{n,i}(x) = r_{n-j_1-1}(x) r_{n-j_2-1}(x) \dots r_{n-j_p-1}(x), \qquad x \in [0,1),$$

where $n, i \in N$, $2 \le i \le 2^n$, $n \ge 1$ and

$$i - 1 = 2^{j_1} + 2^{j_2} + \dots + 2^{j_p},$$

with $j_1 > j_2 > \dots > j_p \ge 0$, is the dyadic expansion of the integer $i-1$.

For $i = 1$ and $n \ge 1$, we set

$$\psi_{n,1}(x) = 1, \qquad x \in [0,1).$$

The Walsh–Kaczmarz system $\{h_m(x)\}_{m=1}^{\infty}$ on $[0,1)$ is defined as follows:

$$h_1(x) = 1 \quad \text{and} \quad h_2(x) = r_0(x), \quad x \in [0,1).$$

For $m = 2^n + i$, $n \ge 1$, $1 \le i \le 2^n$, we set

$$h_m(x) = h_{2^n+i}(x) = \psi_{n,i}(x) r_n(x), \qquad x \in [0,1).$$

We establish the following two theorems.

Theorem 2. *There exists a continuous function G on $[0,1]^2$ such that*

$$\omega(G;\delta) = O\left(\frac{1}{\sqrt{\log_2 \frac{1}{\delta}}}\right), \quad \delta \to 0+,$$

and the Fourier series of G with respect to the double Walsh system $\{\varphi_m(x)\varphi_n(y)\}_{m,n=0}^{\infty}$ diverges on a set of positive measure by rectangles.

Theorem 3. *There exists a continuous function H on $[0,1]^2$ such that*

$$\omega(H;\delta) = O\left(\frac{1}{\sqrt{\log_2 \frac{1}{\delta}}}\right), \quad \delta \to 0+,$$

and the Fourier series of H with respect to the double Walsh–Kaczmarz system $\{h_m(x)h_n(y)\}_{m,n=1}^{\infty}$ diverges on a set of positive measure by rectangles.

A weaker result than Theorem 3 has been proved by us in [1].

Acknowledgement

The presented work was supported by grant No. DI-18-118 of the Shota Rustaveli National Science Foundation of Georgia.

References

1. R. Getsadze, On the divergence of multiple Fourier series with respect to the Walsh–Kaczmarcz system. (Russian) *Problems of Theory of Functions and Functional Analysis* **4** (1988), 36–42.
2. R. Getsadze, On the Divergence of Double Fourier–Walsh–Paley Series of Continuous Functions. *Acta Sci. Math. (Szeged)* **86** (2020), 287–302.
3. W. Sang Young, On rearrangements of Walsh–Fourier series and Hardy–Littlewood type maximal inequalities. *Bulletin of the American Mathematical Society* **80** (1974), no. 3, 490–494.
4. F. Šipp, Certain rearrangements of series in the Walsh system. (Russian) *Mat. Zametki* **18** (1975), no. 2, 193–201.

(Received 11.12.2019)

Department of Mathematics, Uppsala University, Box 480, 751 06 UPPSALA, Sweden
E-mail address: rostom.getsadze@math.uu.se; rostom.getsadze@telia.com

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 251–255

ON THE BOUNDEDNESS OF MULTIPLE CAUCHY SINGULAR AND FRACTIONAL INTEGRALS DEFINED ON THE PRODUCT OF RECTIFIABLE CURVES

VAKHTANG KOKILASHVILI

Abstract. The present paper deals with the boundedness criteria of multiple Cauchy singular integrals and multiple fractional integrals defined on the product of rectifiable curves in weighted Lebesgue spaces.

1. Introduction

This research is stimulated by the R. Coifman and Y. Meyer's well-known lectures [5] and the paper by St. Semme [18] discussing various problems of non-harmonic Fourier Analysis, among them the problem of boundedness of integral operators generated by the Cauchy singular integrals defined on the sets of intricate geometry. A special interest to this problem is shown by its wide possible applications to the boundary value problems of analytic and harmonic functions, boundary integral equations, PDEs of Mathematical Physics, Mechanics of continuum media, etc.

A complete description of those rectifiable curves, for which Cauchy singular integral operator is bounded in $L^p(\Gamma)$ $(1 < p < \infty)$ has been done by G. David [6]. A modern weight theory for the Cauchy singular integrals in the framework of Muckenhoupt weights is constructed in [4] and [12]. In [13], the boundedness criteria in weighted L^p $(1 < p < \infty)$ spaces was established for multiple Cauchy singular integrals defined on the product of two smooth curves. The necessary and sufficient condition both for curves and weights ensuring the boundedness of the Cauchy singular integral operator in some non-standard Banach function spaces, namely, in weighted grand Lebesgue spaces, can be found in [16] (see also [11, 16, 17]).

The mapping properties of a conjugate function of several variables and the related problems of Fourier trigonometric series were investigated by K. Sokol–Sokolovskii [19]. Further exploration of the problems of multi-dimensional Fourier series and conjugate functions is developed in the papers due to A. Zygmund [23], L. V. Zhizhiashvili (see, e. g., [20–22]), C. L. Fefferman [8], P. L. Lizorkin [18]. To the comprehensive study of multiple singular integrals on the product spaces in weighted setting are devoted the papers by R. Fefferman [10] and E. M. Stein [9]. For the surveys of multiple Fourier series and related integral operators we refer to [1, 2] and [23].

2. Function Spaces. Integral Operators

Let $\Gamma = \{t \in \mathbb{C} : \ t = t(s), \ 0 \le s \le l < \infty\}$ be a simple rectifiable curve with the arc-length measure ν. In the sequel, we set

$$D(t, r) = \Gamma \cap B(t, r), \quad 0 < r < d, \quad d = \operatorname{diam} \Gamma,$$

where

$$B(t, r) = \{z \in \mathbb{C} : \ |z - t| < r\}, \quad t \in \Gamma.$$

Γ is called Carleson (regular) curve if

$$\sup_{\substack{t\in\Gamma \\ 0<r<\operatorname{diam}\Gamma}} \frac{\nu D(t, r)}{r} < \infty.$$

2020 *Mathematics Subject Classification.* Primary: 42A50, 44A30, 42B28, 42B35.

Key words and phrases. Singular integrals; Carleson curve; Weights; Multiple Cauchy singular integrals; Multiple fractional integrals.

An almost everywhere positive integrable on Γ function is called a weight. By $L^p_w(\Gamma)$, $(1<p<\infty)$ we denote the set of all measurable functions $f:\Gamma\longrightarrow\mathbb{C}$ for which the norm

$$\|f\|_{L^p_w(\Gamma)}=\Big(\int\limits_{\Gamma}|f(t)|^p w(t)d\nu\Big)^{\frac{1}{p}}$$

is finite.

A weight function w is said to be of Muckenhoupt type class if

$$\sup_{\substack{t\in\Gamma\\0<r<d}}\bigg(\frac{1}{r}\int\limits_{D(t,r)}w(\tau)d\nu\bigg)\bigg(\frac{1}{r}\int\limits_{D(t,r)}w^{1-p'}(\tau)d\nu\bigg)^{p-1}<\infty,\quad p'=\frac{p}{p-1}.$$

Now let $\Gamma=\Gamma_1\times\Gamma_2$ be the product of two simple rectifiable curves of finite lengths, endowed by a product measure $\nu=\nu_1\times\nu_2$.

Let $1<p<\infty$ and a weight function w be given on Γ. By $L^p_w(\Gamma)$ we denote the set of all ν-measurable functions $f:\Gamma\longrightarrow\mathbb{C}$ for which the norm

$$\|f\|_{L^p_w(\Gamma)}=\Big(\int\limits_{\Gamma}|f(t,\tau)|^p w(t,\tau)d\nu\Big)^{1/p}$$

is finite.

Along with $L^p(\Gamma)$ we will treat also the weighted Lebesgue spaces with a mixed norm. For the simplicity of our presentation we consider a 2-multiple case.

Let $1<p_1<p_2<\infty$, $\overrightarrow{p}=(p_1,p_2)$, $\overrightarrow{\nu}=(\nu_1,\nu_2)$. By $L^{\overrightarrow{p}}_{\overrightarrow{\nu}}(\Gamma)$ we denote the set of all ν-measurable functions $f:\Gamma\longrightarrow\mathbb{C}$ for which the norm

$$\|f\|_{L^{\overrightarrow{p}}_{\overrightarrow{\nu}}(\Gamma)}=\bigg(\int\limits_{\Gamma_1}\Big(\int\limits_{\Gamma_2}|f(t,\tau)|^{p_2}d\nu_2\Big)^{\frac{p_1}{p_2}}d\nu_1\bigg)^{1/p_1}$$

is finite.

The notion and properties of the mixed-norm Lebesgue spaces were introduced in [3].

The goal of our paper is to discuss the boundedness problem for multiple maximal Cauchy singular integrals and multiple fractional integrals defined on the product $\Gamma=\Gamma_1\times\Gamma_2$.

Let $\Gamma_{\varepsilon\eta}(t,\tau)=(\Gamma_1\backslash D_1(t,\varepsilon))\times((\Gamma_2\backslash D_2(\tau,\eta))$.

The double maximal Cauchy singular integral is defined as

$$S^*_{\Gamma}f(t,\tau)=\sup_{\substack{0<\varepsilon_1<d_1\\0<\eta<d_2}}\bigg|\int\limits_{\Gamma_{\varepsilon\eta}}\frac{f(t_0,\tau_0)d\nu_1d\nu_2}{(t-t_0)(\tau-\tau_0)}\bigg|,\;\; d_i=\operatorname{diam}\Gamma_i.$$

The double fractional integral defined on Γ looks as

$$\mathbb{I}^{\gamma}f(t,\tau)=\int\limits_{\Gamma}\frac{f(t_0,\tau_0)d\nu_1d\nu_2}{|t-t_0|^{1-\gamma_1}|\tau-\tau_0|^{1-\gamma_2}},$$

$\gamma=(\gamma_1,\gamma_2)$, $0<\gamma_i<1$, $j=1,2$.

In the sequel, we will employ the following class of weight functions:

$$A_p(\Gamma)=\bigg\{w:\;\sup\frac{1}{r\varrho}\int\limits_{V_{r\rho}(t,\tau)}w(t_0,\tau_0)d\nu\bigg(\frac{1}{r\rho}\int\limits_{V_{r\rho}(t,\tau)}w^{1-p'}(t_0,\tau_0)d\nu\bigg)^{p-1}\bigg\}<\infty$$

where $V_{r\rho}(t,\tau)=D_1(t,\tau)\times D_2(\tau,\rho)$ and the supremum being taken over all $t\in\Gamma_1$, $\tau\in\Gamma_2$ and r, ρ, $0<r<d_1$ and $0<\rho<d_2$.

3. Main Results

Theorem 1. *Let $1<p<\infty$. The operator S_Γ^* is bounded in $L^p_w(\Gamma)$ if and only if Γ_i are the Carleson curves and $w\in A_p(\Gamma)$.*

The proof of Theorem 1 is based on the following lemmas.

Lemma 2. *Let $1<p<\infty$ and $w\in A_p(\Gamma)$. Let Γ_i $(i=1,2)$ be Carleson curves. Then for arbitrary $f\in L^p_w(\Gamma)$, almost all $(t,\tau)\in\Gamma$ and arbitrary ε and η, the following equality*

$$\int\limits_{\Gamma_1}\frac{d\nu_1}{t(s)-t(s_0)}\Bigg(\int\limits_{\Gamma_2\backslash D(\tau,\eta)}\frac{f(s_0,\sigma_0)d\nu_2}{\tau(\sigma)-\tau(\sigma_0)}\Bigg)$$
$$=\int\limits_{\Gamma_2\backslash D(\tau,\eta)}\frac{1}{\tau(\sigma)-\tau(\sigma_0)}\Bigg(\int\limits_{\Gamma_1}\frac{f(s_0,\sigma_0)d\nu_1}{t(s)-t(s_0)}\Bigg)d\nu_2$$

holds.

Lemma 3. *Let Γ be a Carleson curve, $1<p<\infty$ and $w\in A_p(\Gamma)$. Then there exists a positive constant b such that for arbitrary $f\in L^p_w(\Gamma)$, arbitrary $\varepsilon>0$, $t\in\Gamma$ and $\bar t\in D\left(t,\frac{\varepsilon}{4}\right)$ the inequality*

$$\left|\frac{f(s_0)d\nu}{t(s)-t(s_0)}\right|\le b\Bigg(\Bigg|\int\limits_{\Gamma}\frac{f(s_0)d\nu}{\bar t(s)-t(s_0)}\Bigg|$$
$$+\Bigg|\int\limits_{\Gamma}\frac{f(s_0)\chi_{D(t,\varepsilon)}(s_0)d\nu}{\bar t(s)-t(s_0)}\Bigg|+M_\Gamma f(t)\Bigg)$$

holds.

Here

$$M_\Gamma f(t)=\sup_{0<r<d(\Gamma)}\frac{1}{r}\int\limits_{D(t,r)}|f(\tau)|d\nu.$$

Lemma 4. *Let Γ be a Carleson curve of finite length and let $\varphi\in L(\Gamma)$. Then for arbitrary $\varepsilon>0$, for all $\bar t\in D\left(t,\frac{\varepsilon}{4}\right)$ and almost all $t\in\Gamma$ the inequality*

$$\Bigg|\int\limits_{\Gamma\backslash D(t,\varepsilon)}\left(\frac{1}{t(s)-t(s_0)}-\frac{1}{\bar t-t(s_0)}\right)\varphi(s_0)d\nu\Bigg|\le cM_\Gamma\varphi(t)$$

holds with a positive constant c, independent of φ and t.

Lemma 5. *Let $1<p<\theta<\infty$ and let Γ be a simple Carleson curve. Suppose $w\in A_p(\Gamma)$. Assume that (Y,μ) is some measure space. Then for arbitrary measurable $f:\Gamma\times Y\longrightarrow\mathbb{C}$ the following inequality*

$$\Bigg(\int\limits_{\Gamma}\Bigg(\int\limits_{Y}M^\theta_\Gamma(f)(t,y)d\mu_y\Bigg)^{p/\theta}w(t)d\nu\Bigg)^{1/p}$$
$$\le c\Bigg(\int\limits_{\Gamma}\Bigg(\int\limits_{Y}|f(t,y)|^\theta d\mu_y\Bigg)^{p/\theta}w(t)d\nu\Bigg)^{\frac{1}{p}}$$

holds with a constant $c>0$, independent of f.

Theorem 6. *Let $1<p_i<\infty$ $(i=1,2)$. The operator S_Γ is bounded in $L^{\overrightarrow{p}}_{\overrightarrow{w}}(\Gamma)$ if and only if Γ_i $(i=1,2)$ are Carleson curves and $w_i\in A_{p_i}(\Gamma)$.*

In the sequel, we will discuss a description of those rectifiable curves for which the operator $\mathbb{I}^\gamma_\Gamma$ is bounded from $L^{\overrightarrow{p}}(\Gamma)$ to $L^{\overrightarrow{q}}(\Gamma)$, $1<p_i<q_i<\infty$, $(i=1,2)$.

Theorem 7. *Let $1 < p_j < q_j < \infty$, $0 < \gamma_j < 1$ $(j = 1, 2)$. The operator $\mathbb{I}_\Gamma^\gamma$ is bounded from $L^{\overrightarrow{p}}_{\overrightarrow{\nu}}(\Gamma)$ to $L^{\overrightarrow{q}}_{\overrightarrow{\nu}}(\Gamma)$ if and only if*

$$\sup_{\substack{t_j r_j \\ j=1,2}} \nu_j \left(D_j(t_j, r_j)\right) r_j^{-\frac{p_j q_j (1-\gamma_j)}{p_j q_j + p_j - q_j}} < \infty.$$

From Theorem 7 follows

Theorem 8 (Sobolev type statement)**.** *Let $1 < p_j < \frac{1}{\gamma_j}$ and let $\frac{1}{q_j} = \frac{1}{p_j} - \gamma_j$. Then the operator $\mathbb{I}_\Gamma^\gamma$ is bounded from $L^{\overrightarrow{p}}(\Gamma)$ to $L^{\overrightarrow{q}}(\Gamma)$ if and only if Γ_j $(j = 1, 2)$ are Carleson curves.*

Acknowledgement

This work was supported by the Shota Rustaveli National Scientific Foundation grant (contract number: FR-18-2499).

References

1. Sh. A. Alimov, V. A. Ilyn, E. M. Nikishin, Convergence problems of multiple Fourier series and spectral decomposition: II (Rusiian) *Uspekhi Math. Nauk* **31/32**(1966/67), 28–83, 107–130.
2. Sh. A. Alimov, R. R. Ashurov, A. K. Pulatov, *Multiple Fourier Series and Fourier Integrals.* Commutative harmonic analysis, IV, 1–95, Encyclopaedia Math. Sci., 42, Springer, Berlin, 1992.
3. A. Benedek, R. Panzone, The spaces L^P, with mixed norm. *Duke Math. J.* **28**(1961), 301–324.
4. A. Böttcher, Y. I. Karlovich, *Carleson Curves, Muckenhoupt Weights, and Toeplitz Operators.* Progress in Mathematics, 154. Birkhäuser Verlag, Basel, 1997.
5. R. Coifman, Y. Meyer, Nonlinear harmonic analysis, operator theory and P.D.E. In: *Beijing Lectures in Harmonic Analysis* (Beijing, 1984), 3–45, Ann. of Math. Stud., 112, Princeton Univ. Press, Princeton, NJ, 1986.
6. G. David, Opérateurs intgraux singuliers sur certaines courbes du plan complexe. (French) *Ann. Sci. École Norm. Sup. (4)* **17** (1984), no. 1, 157–189.
7. D. E. Edmunds, V. Kokilashvili, A. Meskhi, *Bounded and Compact Integral Operators.* Mathematics and its Applications, 543. Kluwer Academic Publishers, Dordrecht, 2002.
8. C. L. Fefferman, Estimates for double Hilbert transforms. *Studia Math.* **44** (1972), 1–15.
9. R. Fefferman, E. M. Stein, Singular integrals on product spaces. *Adv. in Math.* **45** (1982), no. 2, 117–143.
10. R. Fefferman, Harmonic analysis on product spaces. *Ann. of Math. (2)* **126** (1987), no. 1, 109–130.
11. A. Fiorenza, V. Kokilashvili, Nonlinear harmonic analysis of integral operators in weighted grand Lebesgue spaces and applications. *Ann. Funct. Anal.* **9** (2018), no. 3, 413–425.
12. G. Khuskivadze, V. Kokilashvili, V. Paatashvili, Boundary value problems for analytic and harmonic functions in domains with nonsmooth boundaries. Applications to conformal mappings. *Mem. Differential Equations Math. Phys.* **14** (1998), 195 pp.
13. V. Kokilashvili, Singular and bisingular integral operators in weight spaces. (Russian) *Trudy Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. SSR* **69** (1982), 51–79.
14. V. Kokilashvili, On weighted Lizorkin-Triebel spaces. Singular integrals, multipliers, imbedding theorems. (Russian) *Trudy Mat. Inst. Steklov.* **161** (1983), 125–149. English transl.: *Proc. Steklov Inst. Mat.* **3** (1984), 135–162.
15. V. Kokilashvili, A. Meskhi, Fractional integrals on measure spaces. *Fract. Calc. Appl. Anal.* **4** (2001), no. 1, 1–24.
16. V. Kokilashvili, Boundedness criteria for singular integrals in weighted grand Lebesgue spaces. Problems in mathematical analysis. no. 49. *J. Math. Sci. (N.Y.)* **170** (2010), no. 1, 20–33.
17. V. Kokilashvili, Weighted grand Lebesgue space with mixed norm and integral operators. English transl.: *Dokl. Math.* **100** (2019), no. 3, 549–550.
18. St. Semmes, Nonlinear Fourier analysis. *Bull. Amer. Math. Soc. (N.S.)* **20** (1989), no. 1, 1–18.
19. K. Sokół-Sokołowski, K. On trigonometric series conjugate to Fourier series of two variables. *Fund. Math.* **34** (1947), 166–182.
20. L. V. Zhizhiashvili, Conjugate functions of n variables. (Russian) *Dokl. Akad. Nauk SSSR* **218** (1974), 517–518.
21. L. V. Zhizhiashvili, Some problems in the theory of simple and multiple trigonometric and orthogonal series. (Russian) *Uspekhi Math. Nauk* **28** (1973), no. 2 (170), 5–119.
22. L. V. Zhizhiashvili, *Trigonometric Fourier Series and their Conjugates.* Revised and updated translation of Some problems of the theory of trigonometric Fourier series and their conjugate series. (Russian) Kluwer Academic Publishers Group, Dordrecht, 1996.

23. A. Zygmund, On the boundary values of functions of several complex variables. I. *Fund. Math.* **36** (1949), 207–235.

(Received 18.03.2020)

A. Razmadze Mathematical Institute of I. Javakhishvili Tbilisi State University, 6 Tamarashvili Str., Tbilisi 0177, Georgia

E-mail address: vakhtang.kokilashvili@tsu.ge

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 257–259

A NOTE ON THE MULTIPLE FRACTIONAL INTEGRALS DEFINED ON THE PRODUCT OF NONHOMOGENEOUS MEASURE SPACES

VAKHTANG KOKILASHVILI[1] AND TSIRA TSANAVA[1,2]

Abstract. In this note we present a trace type inequality in the mixed-norm Lebesgue spaces for multiple fractional integrals defined on an arbitrary measure quasi-metric space.

1. Introduction

Let (X, d, μ) be of nonhomogeneous type, i.e., a topological space endowed with a locally finite complete measure μ and quasi-metric $d : X \longrightarrow R$ satisfying the following conditions:

(i) $d(x,x) = 0$, for all $x \in X$;

(ii) $d(x,y) > 0$, for all $x \neq y$, $x, y \in X$;

(iii) there exists a positive constant a_0 such that $d(x,y) \le a_0 d(y,x)$ for every $x, y \in X$;

(iv) there exists a positive constant a_0 such that

$$d(x,y) \le a_1 \left(d(x,z) + d(z,y) \right) \quad \text{for every} \quad x, y, z \in X;$$

(v) for every neighbourhood V of the point $x \in X$ there exists $r > 0$ such that the ball $B(x,r) = \{y \in X : \ d(x,y) < r\}$ is contained in V;

(vi) the ball $B(x,r)$ is measurable for every $x \in X$ and for arbitrary $r > 0$.

Let

$$I^\gamma f(x) = \int\limits_X (d(x,y))^{\gamma - 1} f(y) d\mu, \quad 0 < \gamma < 1.$$

In [3] (see also [2, Chapter 6]), the following statement is proved.

Theorem A. *Let $1 < p < q < \infty$ and $0 < \gamma < 1$. The operator I^γ acts boundedly from $L^p_\mu(X)$ to $L^q_\mu(X)$ if and only if there exists a constant $c > 0$ such that*

$$\mu B(x,r) \le c r^\beta, \quad \beta = \frac{pq(1-\gamma)}{pq + p - q}$$

for an arbitrary ball $B(x,r)$.

Let now X_j, d_j, μ_j $(j = 1, 2, \dots, n)$ be the measure quasi-metric spaces. Assume that $\overrightarrow{p} = (p_1, \dots, p_n)$, $1 < p_j < \infty$ $(j = 1, 2, \dots, n)$ and $\overrightarrow{\mu} = (\mu_1, \mu_2, \mu \dots, \mu_n)$. For the measure $f : \prod\limits_{j=1}^n X_j \longrightarrow R^1$ we set the mixed-norm Lebesgue spaces $L^{\overrightarrow{p}}_{\overrightarrow{\mu}}\Big(\prod\limits_{j=1}^n X_j, \prod\limits_{j=1}^n \mu_j \Big)$ with the norm

$$\|f\|_{L^{\overrightarrow{p}}_{\overrightarrow{\mu}}} = \Bigg(\int\limits_{X_1} \cdots \Bigg(\int\limits_{X_n - 1} \Bigg(\int\limits_{X_n} |f(x_1, \dots, x_n)|^{p_n} d\mu_n \Bigg)^{\frac{p_{n-1}}{p_n}} d\mu_{n-1} \Bigg)^{\frac{p_{n-2}}{p_{n-1}}} \dots d\mu_1 \Bigg)^{\frac{1}{p_1}}.$$

The mixed-norm Lebesgue spaces were introduced and studied in [1].

2020 *Mathematics Subject Classification.* 26A33, 43A15, 42B35.

Key words and phrases. Multiple fractional integrals; Mixed-norm Lebesgue space; Trace inequality; Nonhomogeneous measure space.

Consider the multiple fractional integral defined on the product space $X = X_1 \times \cdots \times X_n$:

$$I^\gamma f(x) = \int\limits_X \frac{f(y_1,\dots,y_n)d\mu_1\cdots d\mu_n}{\prod\limits_{j=1}^n (d_j(x_j,y_j))^{1-\gamma j}}, \quad \gamma = (\gamma_1,\dots,\gamma_n).$$

The following statement is true.

Theorem 1. *Let $1 < p_j < q_j < \infty$ $(j = 1,2,\dots,n)$. The operator I^γ is bounded from $L^{\overrightarrow{p}}_{\overrightarrow{\mu}}$ to $L^{\overrightarrow{q}}_{\overrightarrow{\mu}}$ if and only if there exists a positive constant c such that*

$$\mu_j B_j(x_j,r_j) \le c r_j^{\frac{p_j q_j(1-\gamma_j)}{p_j q_j + p_j - q_j}}, \quad j = 1,2,\dots,n \tag{1}$$

for arbitrary balls B_j from X_j.

Theorem 1 says that if the condition (1) fails, then I^γ is unbounded from $L^{\overrightarrow{p}}_{\overrightarrow{\mu}}$ to $L^{\overrightarrow{q}}_{\overrightarrow{\mu}}$. Nevertheless, there exists a weight $\overrightarrow{v} : X \longrightarrow R^1$ such that I^γ is bounded from $L^{\overrightarrow{p}}_{\overrightarrow{\mu}}$ to $L^{\overrightarrow{q}}_{\overrightarrow{\mu}}(\overrightarrow{v})$.

Let us introduce the functions

$$\Omega(x_j) = \sup_{r_j>0} \frac{\mu_j B(x_j,r_j)}{r_j^{\beta_j}},$$

where

$$\beta_j = \frac{p_j q_j(1-\gamma_j)}{p_j q_j + p_j - q_j}. \tag{2}$$

The following statement holds.

Theorem 2. *Let $1 < p_j < q_j < \infty$ $(j = 1,2,\dots,n)$. Then there exists a positive constant $c > 0$ such that for an arbitrary $f \in L^{\overrightarrow{p}}_{\mu}(X)$ we have*

$$\left\| I^\gamma f(x_1,\dots,x_n) \prod_{j=1}^n \Omega_j^{\frac{\gamma_j-1}{p_j}}(x_j) \right\|_{L^{\overrightarrow{q}}_{\overrightarrow{\mu}}} \le c\|f\|_{L^{\overrightarrow{p}}_{\overrightarrow{\mu}}}.$$

Let now $\Gamma_i = \{t \in \mathbb{C} : \ t = t(s), 0 \le s \le l\}$ be arbitrary rectifiable simple curves with arc-length measures ν_i $(i = 1,2,\dots,n)$.

Suppose

$$D_j(t_j,r_j) = \Gamma_j \bigcap B_i(t_j,r_j),$$

where

$$B_i(t_i,r_i) = \{z_i \in \mathbb{C} : \ |z_j - t_j| < r_j\}, \quad t_j \in \Gamma_j.$$

Let

$$\Omega_j(t_j) = \sup_{r_j>0} \frac{\nu_j D(t_j,r_j)}{r_j^{\beta_j}},$$

where β_j are defined by (2).

Then for the operator

$$I^\gamma_\Gamma f(t_1,t_2,\dots,t_n) = \int\limits_\Gamma \frac{f(\tau_1,\tau_2,\dots,\tau_n)d\nu_1\dots d\nu_n}{\prod\limits_{j=1}^n |t_j-\tau_j|^{1-\gamma_j}}, \quad \Gamma = \Gamma_1 \times \cdots \times \Gamma_n$$

we have the following assertion.

Theorem 3. *Let $1 < p_j < q_j < \infty$. Then there exists a positive constant c such that for an arbitrary $f \in L^{\overrightarrow{p}}_{\overrightarrow{\nu}}(\Gamma)$ we have*

$$\|I^\gamma_\Gamma f(t_1,t_2,\dots,t_n) \cdot \Omega_j^{\frac{\gamma_j-1}{p_j}}(t_j)\|_{L^{\overrightarrow{q}}_{\overrightarrow{\nu}}(\Gamma)} \le c\|f\|_{L^{\overrightarrow{q}}_{\overrightarrow{\nu}}}.$$

Acknowledgement

This work was supported by the Shota Rustaveli National Scientific Foundation grant (contract number: FR-18-2499).

References

1. A. Benedek, R. Panzone, The space L^P, with mixed norm. *Duke Math. J.* **28** (1961), 301–324.
2. D. E. Edmunds, V. Kokilashvili, A. Meskhi, *Bounded and Compact Integral Operators.* Mathematics and its Applications, 543. Kluwer Academic Publishers, Dordrecht, 2002.
3. V. Kokilashvili, A. Meskhi, Fractional integrals on measure spaces. *Fract. Calc. Appl. Anal.* **4** (2001), no. 1, 1–24.

(Received 08.04.2020)

[1]A. Razmadze Mathematical Institute of I. Javakhishvili of Tbilisi State University, 6 Tamarashvili Str., Tbilisi 0177, Georgia

[2]Department of Mathematics, Georgian Technical University, 77 Kostava Str., Tbilisi 0171, Georgia

E-mail address: `vakhtang.kokilashvili@tsu.ge`
E-mail address: `ts.tsanava@gtu.ge`

Submission Guidelines

Manuscripts can be submitted in English.

The first page should include abstracts. The first footnote of this page should include 2010 Mathematics Subject Classification numbers, key words and phrases. Abbreviations of the names of journals and references to books should follow the standard form established by Mathematical Reviews. After the references please give the author's address.

Authors should strive for expository clarity and good literary style. The manuscript lacking in these respects will not be published.

It is recommended that each submitted article be prepared in camera-ready form using TEX(plain, LATEX, AMSLATEX) macro package. The typefont for the text is ten point roman with the baselinkship of twelve point. The text area is 155×230 mm excluding page number. The final pagination will be done by the publisher.

The submission of a paper implies the author's assurance that it has not been copyrighted, published or submitted for publication elsewhere.

Transactions of A. Razmadze Mathematical Institute
I. Javakhishvili Tbilisi State University
6 Tamarashvili Str., Tbilisi 0177
Georgia

Tel.: (995 32) 239 78 30, (995 32) 239 18 05

E-mail: kokil@tsu.ge, vakhtangkokilashvili@yahoo.com
maia.svanadze@gmail.com
luiza.shapakidze@tsu.ge

Further information about the journal can be found at:
http://www.rmi.ge/transactions/