Approximations of empirical processes

Ph.D. Thesis

By Gábor Szűcs

Supervisor: Professor Sándor Csörgő

Doctoral School in Mathematics and Computer Science
Bolyai Institute, University of Szeged
Szeged, 2011

Contents

Chapter 1

Introduction

1.1 Approximations of empirical processes

Empirical processes play a fundamental role in statistics. The theoretical background of many methods is the fact that an empirical process converges in distribution in some function space. The best known and most important example is the uniform empirical process of independent and identically distributed uniform sample variables. The classical goodness of fit tests, the Kolmogorov–Smirnov and the Cramér–von Mises test, are based on the convergence of this process to the Brownian bridge.

In general, an applied statistician does not need any further properties of the uniform empirical process, because the weak convergence implies the validity of the standard methods. However, in theoretical research sometimes it can be very useful if we have a stronger type of convergence. It is well-known that the uniform empirical process does not converges almost surely, because the process has infinitely many accumulation points with probability one. The solution of the problem is the approximation method, which means that on a suitable probability space we can represent the process in such a way, that the distance between the process and the elements of a suitable sequence of Brownian bridges converges to zero almost surely. In the '60's and the '70's several appropriate representations were constructed. The most remarkable one, which provides the best rate of convergence, is due to Hungarian mathematicians, to János Komlós, Péter Major and Gábor Tusnády. This is the Hungarian construction, also known as the KMT approximation.

By applying the Hungarian construction in the '70's and the '80's many new results were achieved in the theory of empirical processes, which were unavailable by using only the earlier techniques. One of the motivations of our research in the subject is the fact that many of these improvements are due to Hungarian mathematicians, and some of them were achieved in Szeged, at the university of the author. Endre Csáky, Miklós Csörgő, Pál Révész, and the author's supervisor, Sándor Csörgő were among the first who understood and applied the method. Among many others, a few years later two of their students, Edit Gombay and Lajos Horváth joined them in the research. In the viewpoint of our research we must mention that Sándor Csörgő's co-author and friend,

David M. Mason also has important results in the theory.

Consider independent and identically distributed real valued variables $X_1, \dots, X_n$ and a measurable function $f(t,x)$ defined on a Borel set $\mathbb{T} \times \mathbb{X} \subseteq \mathbb{R}^2$. Assume the X_i's lie in $\mathbb{X}$ with probability one and the mean function $h(t) = Ef(t,X)$ is finite for every $t \in \mathbb{T}$. Also, consider the empirical counterpart of h defined by

$$h_n(t) = \frac{1}{n} \sum_{i=1}^{n} f(t, X_i)\,, \qquad t \in \mathbb{T}\,.$$

The classical definition of the empirical process corresponding to the function h is

$$H_n(t) = n^{\alpha}\big[h_n(t) - h(t)\big]\,, \qquad t \in \mathbb{T}\,, \tag{1.1}$$

where α is a suitable real constant. The phrase "suitable" means that using the scaling factor n^{α} the process $H_n(t)$, $t \in \mathbb{T}$, converges in distribution as n goes to infinity. Note that in most cases $\alpha = 1/2$ by the central limit theorem. Observe that the classical empirical process can be represented by this form, and also, the empirical characteristic and moment generating process are considered in this way, as well.

It is important to note that the process H_n defined in (1.1) is not optimal in every applications, in many problems it is worth to modify the formula. For example, in some cases the function h is unknown, and we must estimate it with some $\hat{h}_n$. In such a situation we can obtain an empirical process by replacing h with $\hat{h}_n$ in (1.1). Also, working in a regression or a bootstrap model we apply the conditional mean of $f(t,X)$ with respect to some background variable(s) instead of $h(t)$, $t \in \mathbb{T}$.

1.2 Aims of the thesis

In the thesis we investigate the asymptotic behavior of some empirical processes based on independent and identically distributed random variables. In most cases we apply the approximation technique, that is, on a suitable probability space we construct a representation of the underlying process and copies of an appropriate Gaussian process such that the distance between the empirical process and the Gaussian processes converges to zero in almost sure or stochastic sense as the sample size goes to infinity.

We study two types of empirical processes. In Chapter 3 we investigate the parametric and the non-parametric bootstrap versions of the parameter estimated empirical process defined on a parametric family of distributions. The main goal of the chapter is to show the convergence of the processes by proving weak approximation theorems for them. We present a bootstrap algorithm for testing goodness of fit, and we demonstrate the bootstrap method with a simulation study.

In Chapter 4 we investigate empirical processes based on the probability generating function of non-negative valued random variables, and we work out an effective and flexible background for the study of the subject. Using this framework we prove a strong approximation result and a law of the iterated logarithm for the generating process and its derivatives. Also, we define the bootstrapped and the parameter estimated version

of the probability generating process, and we apply them to construct confidence bands for the probability generating function and to test goodness of fit.

Chapter 2 is a technical one, we introduce some basic concepts which will be applied in our research. We present the main results of the Hungarian construction method, we prove a background theorem for the bootstrap technique and we extend the definition of stochastic integration on a finite interval to stochastic integral on the real line.

The author has written three papers on the subject of the thesis. The convergence of the parametric bootstrap version of the estimated empirical process is published in Szűcs (2008). The related theorem for the non-parametric bootstrap process and the simulation study in Chapter 3 are the subjects of an accepted paper, see Szűcs (20??) for reference. Finally, Szűcs (2005) contains the statements on the probability generating process for non-negative integer valued variables. The generalization of the results on generating processes for arbitrary non-negative valued variables is new and unpublished.

Chapter 2

Some basic concepts

2.1 The Hungarian construction

In this section we provide an overview on the so-called Hungarian construction and the related strong approximation method of empirical processes. Let $U_1, U_2, \ldots$ be independent random variables distributed uniformly on the interval $[0,1]$, and consider the empirical distribution function based on the sample variables $U_1, \ldots, U_n$ defined by the form

$$E_n(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{U_i \le u\}}\,, \qquad 0 \le u \le 1\,.$$

(Here and throughout our work the notation $\mathbb{1}$ stands for the indicator of an event or a Borel set on the real line. We mark the indicator variable of an event A by $\mathbb{1}_A$, and $\mathbb{1}_A(u)$, $u \in \mathbb{R}$, denotes the indicator function of the set $A \subseteq \mathbb{R}$, that is, $\mathbb{1}_A(u) = \mathbb{1}_{\{u \in A\}}$.) Then, the uniform empirical process based on the sample is

$$\beta_n(u) = n^{1/2}\big[E_n(u) - u\big]\,, \qquad 0 \le u \le 1\,. \tag{2.1}$$

The process β_n is a random element of the Skorohod space $D[0,1]$, that is, the space of all càdlàg functions defined on the interval $[0,1]$. (We say that a real valued function is càdlàg if it is right-continuous and has left-side limit at every point where it is defined.) The space $D[0,1]$ is endowed with the Skorohod metric, under which $D[0,1]$ is complete and separable. By the well-known result of Donsker the uniform empirical process converges in distribution in $D[0,1]$ as $n \to \infty$ to the Brownian bridge, which is a Gaussian process defined on the interval $[0,1]$. For the detailed properties of the Skorohod space and a proof of Donsker's theorem see Chapter 3 in Billingsley (1968).

The Brownian bridge noted by $B(u)$, $0 \le u \le 1$, in our work is a sample-continuous Gaussian process having pointwise mean 0 and covariance function

$$\mathrm{Cov}\big(B(u), B(v)\big) = \min(u,v) - uv\,, \qquad 0 \le u, v \le 1\,.$$

It is obvious that B is a random element of $C[0,1]$, the space of all continuous functions considered on the interval $[0,1]$. The distribution of the Brownian bridge can also be

represented based on the standard Wiener process $W(u)$, $u \geq 0$, since the process defined by the form

$$B^*(u) = W(u) - uW(1)\,, \qquad 0 \leq u \leq 1\,,$$

has the same distribution as B has. In the later chapters we will work with a special sequence of Brownian bridges $B_1, B_2, \ldots$, which are not independent, and we desire to describe the dependence of the processes. It turns out that a biparametric Gaussian process is the suitable tool to achieve our wish. The Kiefer process $K(u,x)$, $0 \leq u \leq 1$, $0 \leq x$, is a centered sample-continuous Gaussian process with covariance function

$$\mathrm{Cov}\big(K(u,x), K(v,y)\big) = \min(x,y)\big(\min(u,v) - uv\big)\,, \qquad 0 \leq u,v \leq 1\,,\ x,y \geq 0\,.$$

The covariance structure implies that in the variable x the process $K(u,x)$ acts as a Wiener process, and in the other variable u it is a Brownian bridge. More precisely, the standardized process

$$B^\star(u) = x^{-1/2}K(u,x)\,, \qquad 0 \leq u \leq 1\,,$$

is a Brownian bridge for every fixed $x > 0$.

The weak convergence of the uniform empirical process is a crucial tool in statistics, and an applied statistician does not need any further nice properties of β_n. However, in theoretical research sometimes it would be very useful if we have a stronger type of convergence. By the famous result of Finkelstein (1971) the uniform empirical process is relative compact in the space $D[0,1]$, and hence, we can not obtain convergence with probability 1 as n goes to infinity. Fortunately, it turns out that on suitable probability spaces one can construct a representation of the variables $U_1, U_2, \ldots$, such that the related uniform empirical process can be approximated strongly by a sequence of Brownian bridges. We do not detail the construction method of the random variables and the Brownian bridges, because in our work we will only need the following theorem and its corollaries provided by Komlós, Major and Tusnády (1975, 1976). (Recent details on the theory and on the proof of Theorem 2.1 are given by Bretagnolle and Massart (1989) and Castelle and Laurent-Bonvalot (1998).)

Theorem 2.1 (Komlós, Major and Tusnády, 1975). *On a sufficiently rich probability space $(\Omega, \mathcal{A}, P)$ one can construct a sequence of Brownian bridges $B_1, B_2, \ldots$ and independent random variables $U_1, U_2, \ldots$ having uniform distribution on the interval $[0,1]$, such that the related uniform empirical process β_n satisfies*

$$P\Big(\sup_{0\leq u\leq 1}\big|\beta_n(u) - B_n(u)\big| > n^{-1/2}(x + c_1\log n)\Big) \leq c_2\exp(-c_3 x)$$

for all $x \geq 0$ and $n = 1, 2, \ldots$, where c_1, c_2 and c_3 are positive universal constants. Also, one can construct a Kiefer process K and an other sequence of variables $U_1', U_2', \ldots$, such that for the related uniform empirical process β_n' we have

$$P\Big(\sup_{1\leq m\leq n}\sup_{0\leq u\leq 1}\big|m^{1/2}\beta_m'(u) - K(u,m)\big| > (x + c_1'\log n)\log n\Big) \leq c_2'\exp(-c_3' x)$$

for all $x \geq 0$ and $m = 1, 2, \ldots$, with positive universal constants c_1', c_2' and c_3'.

The result of Komlós, Major and Tusnády is usually called the Hungarian construction or the KMT approximation of the empirical process. During our work we refer the probability space and the variables defined in Theorem 2.1 as the KMT space. The Hungarian construction provides a very powerful method to handle empirical processes, but mostly we will need only the following consequence in our applications. The next statement points out also that the representation is optimal in some sense. Here and throughout, for a sequence of random variables $V_1, V_2, \dots$ and a sequence of positive constants $a_1, a_2, \dots$ we write $V_n = \mathcal{O}(a_n)$, $n \to \infty$, if there exists a universal real constant C, not depending on the underlying distributions, such that

$$\limsup_{n\to\infty} \bigl|V_n/a_n\bigr| \le C\,, \qquad n \to \infty\,, \qquad \text{a.s.}$$

Theorem 2.2 (Komlós, Major and Tusnády, 1975)**.** *On the KMT space we have*

$$\sup_{0\le u\le 1} \bigl|\beta_n(u) - B_n(u)\bigr| = \mathcal{O}\biggl(\frac{\log n}{n^{1/2}}\biggr)\,, \qquad n \to \infty\,,$$

and also,

$$\sup_{0\le u\le 1} \bigl|\beta_n'(u) - n^{-1/2}K(u,n)\bigr| = \mathcal{O}\biggl(\frac{\log^2 n}{n^{1/2}}\biggr)\,, \qquad n \to \infty\,.$$

Furthermore, the rate of the approximation can not be sharpened, there does not exist construction which provides $o(n^{-1/2}\log n)$ or $o(n^{-1/2}\log^2 n)$, respectively.

It is important to note that there were approximations for the empirical process β_n before the Hungarian construction, but they provided slower rates of convergence. The first representation using a sequence of Brownian bridges was constructed by Brillinger (1969), and Kiefer (1972) was the first who applied the biparametric process named after him. These constructions provided the rates

$$\mathcal{O}\bigl(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}\bigr) \qquad \text{and} \qquad \mathcal{O}\bigl(n^{1/2}(\log n)^{2/3}\bigr)\,.$$

At this point someone may say that the whole theory is not interesting at all, if we are able to state something only for the uniform distribution. However, it turns out that using a simple and well-known technique we can transfer the statements for an arbitrary distribution. We will refer this method in the following as the quantile transformation. Let $F(x)$, $x \in \mathbb{R}$, be an arbitrary distribution function, and consider the related quantile function

$$F^{-1}(u) = \inf\bigl\{x : F(x) \le u\bigr\}\,, \qquad 0 < u < 1\,. \tag{2.2}$$

Also, consider a sequence of independent random variables $U_1, U_2, \dots$ having uniform distribution on the interval $[0,1]$ and being defined on an arbitrary probability space, and introduce the transformed variables

$$X_i = F^{-1}(U_i)\,, \qquad i = 1, 2, \dots \tag{2.3}$$

Clearly, the X_i's are independent and have common distribution function

$$P(X_i \le x) = P\big(U_i \le F(x)\big) = F(x)\,, \qquad x \in \mathbb{R}\,.$$

Also, the empirical distribution function of the variables $X_1, \dots, X_n$ is

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i \le F(x)\}} = E_n\big(F(x)\big)\,, \qquad x \in \mathbb{R}\,.$$

That is, the empirical process corresponding to the sample can be written in the form

$$\alpha_n(x) = n^{1/2}\big[F_n(x) - F(x)\big] = n^{1/2}\big[E_n\big(F(x)\big) - F(x)\big] = \beta_n\big(F(x)\big)\,, \quad x \in \mathbb{R}\,. \tag{2.4}$$

By considering the uniformly distributed variables provided by Theorem 2.1, the first approximation of Theorem 2.2 immediately implies that for the transformed variables defined on the KMT space by (2.3) we have

$$\sup_{x\in\mathbb{R}} \big|\alpha_n(x) - B_n\big(F(x)\big)\big| \le \sup_{0\le u\le 1} \big|\beta_n(u) - B_n(u)\big| = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right), \qquad n \to \infty\,.$$

Note that one can get a corresponding result using the Kiefer process K, as well.

Theorem 2.3 (Komlós, Major and Tusnády, 1975). *Consider the variables $X_1, X_2, \dots$ defined by (2.3) based on the uniform variables of the KMT space. The corresponding empirical process α_n satisfies*

$$\sup_{x\in\mathbb{R}} \big|\alpha_n(x) - B_n\big(F(x)\big)\big| = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right), \qquad n \to \infty\,.$$

As a result of Theorem 2.3 we get weak convergence of the general empirical process α_n to the Gaussian process $B(F)$ in $D[-\infty,\infty]$, the space of càdlàg functions defined on the real line. (For details on the topology of $D[-\infty,\infty]$ see Chapter VI of Pollard (1984).)

There is an other important consequence of Theorem 2.1, which can not be achieved by using only the weak convergence of the uniform empirical process. Let ψ be a real valued measurable functional defined on the space $D[0,1]$, and assume that ψ is continuous on the subspace $C[0,1]$. Then, since the Brownian bridge lies in $C[0,1]$ with probability 1, it follows from Corollary 1 to Theorem 1.5.1 in Billingsley (1968), that $\psi(\beta_n)$ converges in distribution to $\psi(B)$. The most important application of this result is the case when ψ is the supremum of the functions in $D[0,1]$. Unfortunately, this technique does not provide a rate of convergence, but using Theorem 2.1 one can improve the result. The following statement holds for any representation of the uniform empirical process not only for the one provided by the KMT space.

Theorem 2.4 (Komlós, Major and Tusnády, 1975). *Consider a real valued measurable functional ψ defined on $D[0,1]$ satisfying the Lipschitz condition*

$$\big|\psi(f) - \psi(g)\big| \le M \sup_{0\le u\le 1} \big|f(u) - g(u)\big|$$

with some finite $M > 0$, and suppose that $\psi(B)$ has bounded density. Then, we have

$$\sup_{x\in\mathbb{R}} \left|P\big(\psi(\beta_n) \le x\big) - P\big(\psi(B) \le x\big)\right| = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right), \qquad n \to \infty\,.$$

Note that we apply similar method to prove our Theorem 4.20 as required for the justification of Theorem 2.4. With some improvements on the technique one can obtain similar results for certain continuous functionals which are not Lipschitzian. Perhaps the most important example is the functional

$$\psi : D[0,1] \to \mathbb{R}\,, \qquad \psi(f) = \int_0^1 f^2(u)\,du\,,$$

that is, when $\psi(\beta_n)$ is the Cramér–von Mises type statistics based on the sample. Csörgő (1976) showed that the distribution function of $\psi(\beta_n)$ converges to that of $\psi(B)$ with the same rate as Lipschitzian functionals do in Theorem 2.4.

Motivated by the results of Komlós, Major and Tusnády (1975, 1976) the Hungarian construction became a widely used tool to handle both unweighted and weighted empirical processes. Note that Csörgő and Révész (1975) and Philipp and Pinzur (1980) provide similar uniform strong approximations also for the multivariate empirical process. We do not present these statements, because in our research we investigate only one-dimensional problems. For a survey on the most important results obtained by the KMT approximation see Csörgő and Révész (1981), and Shorack and Wellner (1986) provides a comprehensive overview on the whole theory of empirical processes.

2.2 The bootstrap method

Consider a typical statistical problem. There is a sample of independent and identically distributed variables $X_1, \dots, X_n$ having unknown distribution function $F(x)$, $x \in \mathbb{R}$, and the goal is to estimate some parameter $\tau = \tau(F)$ of the background distribution. In a favorable case there is a statistic

$$\tau_n = \tau_n(X_1, \dots, X_n)$$

which serves as an estimator of the parameter τ. For example, if we want to determine the expected value or the variance of the distribution then the sample mean and the empirical variance provides a good estimation. In some other applications we want to specify not a single parameter but a function $\tau(x, F)$, $x \in \mathbb{R}$, of the distribution. The most important case is that when the desired function is the theoretical distribution function $F(x)$, $x \in \mathbb{R}$, and the empirical distribution function $F_n(x)$, $x \in \mathbb{R}$, of the sample is a suitable tool. But, unfortunately, based on a single collection of observations we can not determine the variance or the error of the statistics τ_n. This difficulty is not surprising in the sense that to get an estimation for the variance of τ_n we need several observations of the statistic, but having only one sample $X_1, \dots, X_n$ we can write it up only once. That is, we have no tool the determine the variance by using

only the standard statistical methods. Similarly, the empirical distribution function is a pointwise estimate of its theoretical counterpart, but we can not construct a confidence band for F. A possible solution for these difficulties is the application of some resampling technique, that is, the usage of the jackknife or the bootstrap. In our work we study the latter one.

The bootstrap method was introduced by Efron (1979) and the basic idea is very simple. It seems that all of the presented difficulties are derived from the fact that we have only a single collection of observations $X_1, \dots, X_n$. Since it is not possible to obtain further samples from the "real life" let us create them by ourselves. Of course, we have no information on the original distribution of the sample variables, and hence, we can not construct additional samples having the unknown distribution function F. The solution of the problem is to obtain samples by choosing values with replacement from the observations $X_1, \dots, X_n$. That is, consider bootstrapped variables $X^*_{1,n}, \dots, X^*_{m_n,n}$ being conditionally independent with respect to the X_i's and having distribution

$$P\big(X^*_{i,n} = X_j \mid X_1, \dots, X_n\big) = 1/n\,, \qquad j = 1, \dots, n\,, \quad i = 1, \dots, m_n\,.$$

By this way we get the bootstrapped sample $X^*_{1,n}, \dots, X^*_{m_n,n}$ with size m_n, where m_n is an arbitrary positive integer. Now, let us apply the statistics τ_{m_n} on the bootstrapped variables to obtain a bootstrap estimation

$$\tau^*_{m_n,n} = \tau_{m_n}\big(X^*_{1,n}, \dots, X^*_{m_n,n}\big)$$

for the unknown parameter τ. Heuristically, since the bootstrapped variables comes from the original sample they "behave similarly" as the X_i's, and hence, the bootstrapped statistics $\tau^*_{m_n,n}$ has a distribution being close to the law of τ_n in some sense. That is, the distribution of $\tau^*_{m_n,n}$ may be a good estimation for the law of τ_n, and we can estimate the error or the variance of the statistics τ_n based on the corresponding attributes of $\tau^*_{m_n,n}$. Because the bootstrapped observations are obtained by a simple replacement from the original sample, the conditional attributes of the statistics $\tau^*_{m_n,n}$ can be determined in theoretic way by direct calculations or by applying the Edgeworth expansion method. However, thanks to the high performance of computers, nowadays we can apply a much simpler method. By using a computer we can generate hundreds of copies of the bootstrapped sample $X^*_{1,n}, \dots, X^*_{m_n,n}$ in a few seconds, and by calculating the statistics $\tau^*_{m_n,n}$ for each sample we can obtain a good empirical estimation for the law of $\tau^*_{m_n,n}$.

In his fundamental paper Efron (1979) recommended also a variant of the presented method. Suppose that we know it a priori that the sample comes from a member of a distribution family

$$\mathcal{F} = \big\{F(x,\theta) : x \in \mathbb{R}, \theta \in \Theta\big\}\,,$$

that is, $F(x) = F(x,\theta_0)$, $x \in \mathbb{R}$, with a fixed but unknown $\theta_0 \in \Theta$. Again, the goal is to estimate some parameter $\tau_0 = \tau(\theta_0)$ of the background distribution using the sample $X_1, \dots, X_n$. In a favorable case the desired parameter is a smooth function $\tau(\theta)$ on

the family, and we can obtain a good statistics in the form $\tau_n = \tau(\theta_n)$ based on an estimator

$$\theta_n = \theta_n(X_1, \dots, X_n)$$

of the unknown parameter θ_0. Unfortunately, in this way we get only a point estimator, but in most cases we have no information on the variance and the error of τ_n. Again, the basic idea is to consider additional samples, but in this case the bootstrapped variables are not obtained by replacement from the original sample. If the function $F(x, \theta)$ is smooth enough in the parameter θ, and θ_n is a good estimator of θ_0, then the laws corresponding to the functions $F(x, \theta_0)$ and $F(x, \theta_n)$ are close to each other in some sense. That is, if we consider independent bootstrapped variables $X^*_{1,n}, \dots, X^*_{m_n,n}$ having conditional distribution function $F(x, \theta_n)$, $x \in \mathbb{R}$, then the bootstrapped sample "behaves similarly" as the original one. Hence, we can estimate the distribution of τ_n with the statistics $\tau^*_{m_n,n} = \tau(\theta^*_{m_n,n})$, where

$$\theta^*_{m_n,n} = \theta_{m_n}\big(X^*_{1,n}, \dots, X^*_{m_n,n}\big)$$

is the estimation of θ_n based on the bootstrap sample. Again, the attributes of $\tau^*_{m_n,n}$ can be determined by direct calculation, but this method can be very difficult if the function $F(x, \theta)$ of the family is provided only in a complicated formula. In this case the usage of computer programs to determine the distribution of $\tau^*_{m_n,n}$ by generating large number of bootstrap samples is preferred.

The first method, when we obtain the bootstrap sample by replacement from the original observations is the so-called "Efron bootstrap", and the second one, when we generate the variables by using an estimation θ_n is the "parametric bootstrap". The common point of the two methods is that in both cases there is an estimator $\hat{F}_n(x)$ of the unknown distribution function $F(x)$, $x \in \mathbb{R}$, based on the sample $X_1, \dots, X_n$, and the bootstrapped observations $X^*_{1,n}, \dots, X^*_{m_n,n}$ are conditionally independent and have conditional distribution function $\hat{F}_n$ with respect to the X_i's. This remark highlights that the bootstrap method might be applied in a much general framework than the Efron and the parametric bootstrap. In some applications it might be preferable to obtain bootstrapped variables based on some other estimation $\hat{F}_n(x)$ of $F(x)$, and not by applying the distribution functions $F_n(x)$ or $F(x, \theta_n)$, $x \in \mathbb{R}$.

It is important to note that the presented idea is only heuristics and not a proved fact, but under some conditions we can verify the validity of the bootstrap technique. The theoretical background of many standard statistical methods is that some functional $\psi_n = \psi_n(\tau_n, \tau)$ of the estimator τ_n converges weakly to a variable φ. In many applications we can determine the distribution of φ which provides a better or wrong estimation for the law of τ_n. The difficulty arises in those models when we can not specify the distribution of the limit variable φ. In a typical case these statistics can be written in the forms

$$\psi_n = \psi(\alpha_n) \qquad \text{and} \qquad \varphi = \psi\big(B(F)\big)\,,$$

where $\alpha_n(x)$, $x \in \mathbb{R}$, is the empirical process based on the sample $X_1, \dots, X_n$, the random function $B(u)$, $0 \le u \le 1$, is the Brownian bridge, and ψ is a functional on

the space $D[-\infty,\infty]$. In many applications the convergence of the functional ψ_n follows from the weak convergence of the empirical process α_n to $B(F)$. Now, consider a bootstrapped sample based on the original variables, that is, choose conditionally independent values $X^*_{1,n},\dots,X^*_{m_n,n}$ from $X_1,\dots,X_n$ with replacement. Then, the bootstrapped observations have empirical distribution function

$$F^*_{m_n,n}(x)=\frac{1}{m_n}\sum_{i=1}^{m_n}\mathbb{1}_{\{X_{i,n}\le x\}}$$

and their conditional theoretical distribution function with respect to the X_i's is the empirical distribution function $F_n(x)$, $x\in\mathbb{R}$, of the original sample. The heuristics presented in the previous paragraph advices us to consider a bootstrapped sample and apply the same statistical function on the generated variables as we use for the original sample. That is, it seems to be a good start to apply the functional $\psi^*_{m_n,n}=\psi(\alpha^*_{m_n,n})$, where

$$\alpha^*_{m_n,n}(x)=n^{1/2}\big[F^*_{m_n,n}-F_n(x)\big]\,,\qquad x\in\mathbb{R}\,.$$

is the so-called bootstrap empirical process, which is conditionally the empirical process of the bootstrap sample. It is important to note that we can determine the conditional distribution of $\psi^*_{m_n,n}$ by applying theoretical calculations, or at least, we can obtain an empirical estimation for it by large number of bootstrap sample generations. Now, if we can show that $\psi^*_{m_n,n}$ converges weakly to φ, then we have a better or worse estimate of the distribution of φ, and hence, we have an approximation for the law of ψ_n, and we are done. Similarly to the non-bootstrapped setup in many cases the convergence of $\psi^*_{m_n,n}$ can be obtained by showing that the bootstrap empirical process $\alpha^*_{m_n,n}$ converges in distribution to $B(F)$. This asymptotic behavior would not be surprising in the sense that the basis of the bootstrap heuristics is the idea that the objects based on the generated sample behave just the same way as the corresponding objects based on the original sample.

The paper of Efron (1979) was much rather a receipt book than a theoretical work. He introduced the basic concepts of the bootstrap on some interesting examples, but he did not prove the validity of the technique. The theoretic background of the method was demonstrated firstly by Bickel and Freedman (1981) who provided a law of large numbers and a central limit theorem for the Efron bootstrap, and they showed also the convergence of $\alpha^*_{m_n,n}(x)$ in distribution to $B(F(x))$, $x\in\mathbb{R}$, in the space $D[-\infty,\infty]$. The next decade was the renaissance era of the bootstrap, and many applications of the method was found both in applied statistics and in theoretic fields. For an overview on the most important developments see Efron and Tibshirani (1993) or Hall (1992). Also, for a theoretic survey on the asymptotic properties of bootstrapped sums see the paper of Csörgő and Rosalsky (2003). In the viewpoint of our work the most interesting result is a uniform strong approximation for the bootstrap empirical process provided by Csörgő and Mason (1989).

Theorem 2.5 (Csörgő and Mason, 1989)**.** *Consider an arbitrary distribution function $F(x)$, $x\in\mathbb{R}$, and assume that there exist positive constants C_1 and C_2 such that*

$$C_1<m_n/n<C_2\,,\qquad n=1,2,\dots$$

On a sufficiently rich probability space one can define independent random variables $X_1, X_2, \ldots$ *having common distribution function* $F(x)$, $x \in \mathbb{R}$, *and bootstrapped sample variables* $X^*_{1,n}, \ldots, X^*_{m_n,n}$, $n = 1, 2, \ldots$ *based on the* X_i*'s, and a sequence of Brownian bridges* $B^*_1, B^*_2, \ldots$, *such that for the corresponding bootstrap empirical process we have*

$$\sup_{x\in\mathbb{R}} \left|\alpha^*_{m_n,n}(x) - B^*_{m_n}\big(F(x)\big)\right| = \mathcal{O}\big(\max\{l(m_n), l(n)\}\big)\,,$$

with the function $l(n) = n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}$. *Also, the Brownian bridges are independent from the* X_i*'s.*

We must note that the last statement of the theorem about the independence of the Brownian bridges is not stated in the referred result of Csörgő and Mason (1989), but it also holds. The proof of the theorem applies the KMT approximation presented in Theorem 2.1. The basic idea of the construction is that we consider the product of two KMT spaces, and define the variables $X_1, X_2, \ldots$ on the first space, and construct the bootstrapped samples on the other one. In this way the empirical process and the Brownian bridges of the second KMT space serve as $\alpha^*_{m_n,n}$ and B^*_n, and they provides the desired approximation. Since the original sample variables $X_1, X_2, \ldots$ are considered on the first KMT space, they are independent from the Brownian bridges. This interesting idea is adapted also by us in the proofs of our Theorems 3.2 and 3.3.

Unfortunately, this approximation has a weakness. In a typical application we have a significance level $0 < \alpha < 1$, and the goal is to obtain a critical values $c_n(\alpha)$ satisfying the equation

$$P\big(\psi_n \le c_n(\alpha)\big) = 1 - \alpha\,,$$

or at least the convergence

$$P\big(\psi_n \le c_n(\alpha)\big) \to 1 - \alpha\,, \qquad n \to \infty\,. \tag{2.5}$$

Since the statistics ψ_n converges weakly to φ, the critical value $c_n(\alpha)$ can be estimated by using one any value from the interval $[c(\alpha), d(\alpha)]$, where

$$c(\alpha) = \inf\big\{x \in \mathbb{R} : F_\varphi(x) \ge 1 - \alpha\big\}\,, \quad d(\alpha) = \sup\big\{x \in \mathbb{R} : F_\varphi(x) \le 1 - \alpha\big\}\,, \tag{2.6}$$

and $F_\varphi(x)$, $x \in \mathbb{R}$, is the distribution function of φ. By the bootstrap heuristics the quantile $c(\alpha)$ may be approximated by

$$c^*_n(\alpha) = \inf\Big\{x \in \mathbb{R} : P\big(\psi^*_{m_n,n} \le x \mid X_1, \ldots, X_n\big) \ge 1 - \alpha\Big\}\,, \tag{2.7}$$

and $c^*_n(\alpha)$ can be estimated by the empirical quantiles of $\psi^*_{m_n,n}$ after a large number of bootstrap sample generations. That is, working with a given sample $X_1, \ldots, X_n$ we need the conditional weak convergence of the bootstrap empirical process with respect to the observations. The difficulty derives from the fact that Theorem 2.5 implies only unconditional convergence. Fortunately, the approximation technique provides also a solution for this problem.

Consider the observations $X_1, X_2, \dots$, the related bootstrapped variables, and an arbitrary statistics with the form

$$\psi^*_{m_n,n} = \psi^*_{m_n,n}(X_1, \dots, X_n, X^*_{1,n}, \dots, X^*_{m_n,n}) .$$

Since for any fixed n the bootstrapped variables $X^*_{1,n}, \dots, X^*_{m_n,n}$ come from the original sample $X_1, \dots, X_n$, the statistics $\psi^*_{m_n,n}$ is independent from the remaining variables $X_{n+1}, X_{n+2}, \dots$ Assume that the background probability space is rich enough to provide a sequence of variables $\varphi_1, \varphi_2, \dots$ having the same distribution function $F_\varphi(x)$, $x \in \mathbb{R}$, such that

$$\left|\psi^*_{m_n,n} - \varphi_{m_n}\right| \to 0 , \qquad n \to \infty , \tag{2.8}$$

unconditionally in almost sure or stochastic sense. Based on the KMT approximation such a representation may exist. Additionally, suppose that the sequence $\varphi_1, \varphi_2, \dots$ is independent from the observations $X_1, X_2, \dots$ The goal is to obtain some conditions under which the quantile $c^*_n(\alpha)$ introduced in (2.7) converges to $c(\alpha)$ of (2.6).

First, investigate the case when we have almost sure convergence, that is, the event

$$A = \left\{|\psi^*_{m_n,n} - \varphi_{m_n}| \to 0, n \to \infty\right\}$$

has probability 1. Since the variables $X_1, X_2, \dots$ are independent the conditional probability of the convergence in (2.8) can be written as

$$P\big(|\psi^*_{m_n,n} - \varphi_{m_n}| \to 0 \mid X_1, \dots, X_n\big) = P\big(A \mid X_1, X_2, \dots\big) . \tag{2.9}$$

By using Lemma 1.1 of Csörgő and Rosalsky (2003) it follows that the event A has probability 1 if and only if the conditional probability (2.9) is equal to 1, and hence, the unconditional strong approximation in (2.8) is equivalent with the related conditional strong convergence. Since the variables $\varphi_1, \varphi_2, \dots$ have the same distribution as φ and they are independent from the X_i's, we also obtain the conditional weak convergence of $\psi^*_{m_n,n}$ to φ, that is, we have

$$P\big(\psi^*_{m_n,n} \le c \mid X_1, \dots, X_n\big) \to P\big(\varphi \le c \mid X_1, \dots, X_n\big) = F_\varphi(c) , \tag{2.10}$$

as $n \to \infty$ for every continuity points $c \in \mathbb{R}$ of the function F_φ.

In our next proposition we show that the conditional weak convergence (2.10) holds also in the case when we have the convergence in (2.8) only in stochastic sense, and in Theorem 2.7 we provide a theoretical background for our applications in the following chapters. The proofs of the statements adapt the ideas presented in the Appendix of Csörgő and Mason (1989). Note that the results can be applied both for the parametric and for the non-parametric case. It is possible that these results or similar ones are already published by someone, but the we have not find any reference for them.

Proposition 2.6. *If the convergence (2.8) holds in probability then for any continuity point $c \in \mathbb{R}$ of the distribution function $F_\varphi(x)$, $x \in \mathbb{R}$, we have*

$$P\big(\psi^*_{m_n,n} \le c \mid X_1, \dots, X_n\big) \xrightarrow{P} F_\varphi(c) , \qquad n \to \infty .$$

Proof. Fix any values $\varepsilon, \delta > 0$ and consider the events

$$A_n(\varepsilon) = \Big\{ \big|\psi^*_{m_n,n} - \varphi_{m_n}\big| > \varepsilon \Big\},$$

and

$$B_n(\varepsilon,\delta) = \Big\{ P\big(A_{m_n,n}(\varepsilon) \mid X_1,\dots,X_n\big) > \delta \Big\}.$$

Note that the stochastic convergence in (2.8) implies that the probability of the event $A_n(\varepsilon)$ converges to 0. By restricting the domain of integration from the whole universe Ω to $B_n(\varepsilon,\delta)$ we get

$$P\big(A_n(\varepsilon)\big) = E\Big(P\big(A_n(\varepsilon) \mid X_1,\dots,X_n\big)\Big) \ge \delta P\big(B_n(\varepsilon,\delta)\big).$$

Since the left side converges to 0 it follows that

$$P\Big(P\big(A_n(\varepsilon) \mid X_1,\dots,X_n\big) > \delta\Big) = P\big(B_n(\varepsilon,\delta)\big) \to 0,$$

for any positive δ, and hence,

$$P\big(|\psi^*_{m_n,n} - \varphi_{m_n}| > \varepsilon \mid X_1,\dots,X_n\big) = P\big(A_n(\varepsilon) \mid X_1,\dots,X_n\big) \xrightarrow{P} 0.$$

Using the fact that

$$\big\{\psi^*_{m_n,n} \le c\big\} \subseteq \big\{\varphi_{m_n} \le c+\varepsilon\big\} \cup \big\{|\psi^*_{m_n,n} - \varphi_{m_n}| > \varepsilon\big\}$$

and the independence of the variables $\varphi_1, \varphi_2, \dots$ from the X_i's, we obtain

$$\begin{aligned} P\big(\psi^*_{m_n,n} \le c \mid X_1,\dots,X_n\big) &\le P\big(\varphi_{m_n} \le c+\varepsilon \mid X_1,\dots,X_n\big) \\ &\quad + P\big(|\psi^*_{m_n,n} - \varphi_{m_n}| > \varepsilon \mid X_1,\dots,X_n\big) \xrightarrow{P} F_\varphi(x+\varepsilon) + 0. \end{aligned} \tag{2.11}$$

Similarly, from

$$\big\{\varphi_{m_n} \le c-\varepsilon\big\} \subseteq \big\{\psi^*_{m_n,n} \le c\big\} \cup \big\{|\psi^*_{m_n,n} - \varphi_{m_n}| > \varepsilon\big\}$$

we have also the inequality

$$\begin{aligned} &P\big(\varphi_{m_n} \le c-\varepsilon \mid X_1,\dots,X_n\big) \\ &\quad \le P\big(\psi^*_{m_n,n} \le c \mid X_1,\dots,X_n\big) + P\big(|\psi^*_{m_n,n} - \varphi_{m_n}| > \varepsilon \mid X_1,\dots,X_n\big) \end{aligned}$$

which implies that

$$\begin{aligned} P\big(\psi^*_{m_n,n} \le c \mid X_1,\dots,X_n\big) &\ge P\big(\varphi_{m_n} \le c-\varepsilon \mid X_1,\dots,X_n\big) \\ &\quad - P\big(|\psi^*_{m_n,n} - \varphi_{m_n}| > \varepsilon \mid X_1,\dots,X_n\big) \xrightarrow{P} F_\varphi(x-\varepsilon) - 0. \end{aligned} \tag{2.12}$$

Since $\varepsilon > 0$ was arbitrary and the distribution function F_φ is continuous at the point c, the statement follows from formulas (2.11) and (2.12). □

Theorem 2.7. *Consider independent and identically distributed variables $X_1, X_2, \dots$, the related bootstrapped samples $X^*_{1,n}, \dots, X^*_{m_n,n}$, $n = 1, 2, \dots$, and a statistic*

$$\psi^*_{m_n,n} = \psi^*_{m_n,n}\big(X_1, \dots, X_n, X^*_{1,n}, \dots, X^*_{m_n,n}\big)\,.$$

Suppose that the underlying probability space provides identically distributed variables $\varphi_1, \varphi_2, \dots$ which are independent from the sequence $X_1, X_2, \dots$ and

$$\big|\psi^*_{m_n,n} - \varphi_{m_n}\big| \xrightarrow{P} 0\,, \qquad n \to \infty\,.$$

Fix an arbitrary value $0 < \alpha < 1$, and assume that the distribution function $F_\varphi(x)$, $x \in \mathbb{R}$, of the φ_i's is continuous at the points $c(\alpha)$ and $d(\alpha)$ defined in (2.6) Then, for the quantile $c^*_n(\alpha)$ of (2.7) we have

$$F_\varphi\big(c^*_n(\alpha)\big) \xrightarrow{P} 1 - \alpha\,, \qquad n \to \infty\,.$$

Furthermore, if $c(\alpha) = d(\alpha)$ then $c^*_n(\alpha) \to c(\alpha)$ in probability.

Proof. Choose any $\varepsilon > 0$ such that $c(\alpha) - \varepsilon$ and $d(\alpha) + \varepsilon$ are also continuity points of F_φ. Using Proposition 2.6 and the definition of $c(\alpha)$ in (2.6) we have

$$P\big(\varphi_{m_n} \le c(\alpha) - \varepsilon \mid X_1, \dots, X_n\big) \xrightarrow{P} F_\varphi\big(c(\alpha) - \varepsilon\big) < 1 - \alpha\,,$$

and from (2.7) it follows that

$$P\big(c^*_n(\alpha) < c(\alpha) - \varepsilon\big) \le P\Big(P\big(\varphi_{m_n} \le c(\alpha) - \varepsilon \mid X_1, \dots, X_n\big) \ge 1 - \alpha\Big) \xrightarrow{P} 0\,.$$

Similarly, by Proposition 2.6 we obtain the convergence

$$P\big(\varphi_{m_n} \le d(\alpha) + \varepsilon \mid X_1, \dots, X_n\big) \xrightarrow{P} F_\varphi\big(d(\alpha) + \varepsilon\big) > 1 - \alpha\,,$$

and we have

$$P\big(c^*_n(\alpha) > d(\alpha) + \varepsilon\big) \le P\Big(P\big(\varphi_{m_n} \le d(\alpha) + \varepsilon \mid X_1, \dots, X_n\big) \le 1 - \alpha\Big) \xrightarrow{P} 0\,.$$

That is, we obtain that

$$P\big(c(\alpha) - \varepsilon \le c^*_n(\alpha) \le d(\alpha) + \varepsilon\big) \to 1\,, \qquad n \to \infty\,, \tag{2.13}$$

and the monotony of the distribution function F_φ implies

$$P\Big(F_\varphi\big(c(\alpha) - \varepsilon\big) \le F_\varphi\big(c^*_n(\alpha)\big) \le F_\varphi\big(d(\alpha) + \varepsilon\big)\Big) \to 1\,.$$

Note that $c(\alpha)$ and $d(\alpha)$ are continuity points of F_φ by assumption, which implies that the distribution function is constant $1 - \alpha$ on the interval $[c(\alpha), d(\alpha)]$ by (2.6). Since ε is an arbitrary positive value we immediately get that

$$F_\varphi\big(c^*_n(\alpha)\big) \xrightarrow{P} 1 - \alpha\,, \qquad n \to \infty\,.$$

Similarly, if $c(\alpha) = d(\alpha)$ then formula (2.13) implies the convergence $c^*_n(\alpha) \to c(\alpha)$ in probability. □

To summarize what we found consider the significance level α and the quantiles defined in (2.5)–(2.7). In a favorable case $c(\alpha) = d(\alpha)$ and $c(\alpha)$ is a continuity point of the distribution function F_φ. Since ψ_n converges in distribution to φ by assumption, the theory of statistics implies that for any sequence $c_n(\alpha)$ satisfying (2.5) we have $c_n(\alpha) \to c(\alpha)$ in probability. Also, in this case we have $c_n^*(\alpha) \to c(\alpha)$ in the same sense by Theorem 2.7, and hence, the quantile $c_n^*(\alpha)$ can be applied as an estimator of $c(\alpha)$ and $c_n(\alpha)$.

In the less favorable case, when F_φ is continuous at $c(\alpha)$ and $d(\alpha)$, but it is not sure, that these values are equal, then for any x in the interval $[c(\alpha), d(\alpha)]$ we have

$$P(\psi_n \le x) \to P(\varphi \le x) = 1 - \alpha$$

in probability. That is, any element of the interval can serves as $c_n(\alpha)$. By Theorem 2.7 the bootstrap quantile $c_n^*(\alpha)$ asymptotically lies in $[c(\alpha), d(\alpha)]$, and it can play the role of $c_n(\alpha)$, again.

Based on this concept in the following chapters we will construct such a representations of various bootstrapped empirical processes which provide approximations for the processes, and using Theorem 2.7 we will apply the bootstrap quantiles $c_n^*(\alpha)$ to estimate the unknown critical values $c(\alpha)$ of some related statistics ψ_n.

2.3 Stochastic integral processes

Stochastic integration with respect to a local martingale on a finite interval is an important and well investigated field of probability, but in our work we will face stochastic integrals on the whole real line. In this section we provide an extension of the basic theory. We prove a condition for the existence of the integral, and also, we investigate processes defined as the integral of a bivariate deterministic function. (For a simple introduction to stochastic calculus we recommend the book of Karatzas and Shreve (1988). For a more general and deeper framework see Jacod and Shiryaev (2003).)

Let $M(x)$, $x \in \mathbb{R}$, be a local martingale with respect to a filtration $\{\mathcal{F}_x : x \in \mathbb{R}\}$ on the underlying probability space. Suppose that M has pointwise mean 0 and càdlàg trajectories, that is, with probability 1 the process is right-continuous and has left-side limit at every point $x \in \mathbb{R}$. Additionally, assume that M is locally square integrable, which means that it has finite second moment $EM^2(x)$ for any real x, and the second moment is bounded on each finite interval. The Doob–Meyer decomposition provides an adapted process $\langle M \rangle_x$, $x \in \mathbb{R}$, which has non-decreasing and càdlàg trajectories, such that $M^2(x) - \langle M \rangle_x$, $x \in \mathbb{R}$, is a local martingale. (See Theorem 4.2 in Chapter I of Jacod and Shiryaev (2003).) The process $\langle M \rangle$ is called the quadratic characteristic of the local martingale M, and its mean function $E\langle M \rangle_x$, $x \in \mathbb{R}$, is also non-decreasing and càdlàg.

Consider a real valued deterministic function $f(x)$, $x \in \mathbb{R}$, such that f is locally bounded, that is, bounded on every finite interval. Additionally, assume that f is left-continuous and has right-side limit at every point $x \in \mathbb{R}$. Then, for any fixed values

$T_1 < 0 < T_2$ the stochastic integral

$$\int_{T_1}^{T_2} f(x)\, dM(x) \tag{2.14}$$

is a well-defined random variable having mean 0 and variance

$$\|f\|^2_{[T_1,T_2]} = E \int_{[-T_1,T_2]} f^2(x)\, d\langle M\rangle_x \,.$$

Also, the covariance of the stochastic integrals of the functions $f(x)$ and $g(x)$ is

$$\mathrm{Cov}\left(\int_{T_1}^{T_2} f(x)\, dM(x), \int_{T_1}^{T_2} g(x)\, dM(x) \right) = E \int_{[T_1,T_2]} f(x)g(x)\, d\langle M\rangle_x \,.$$

We say that the function f can be integrated on the real line with respect to the local martingale M if the variable in (2.14) has limit in L^2 sense as $T_1 \to -\infty$ and $T_2 \to \infty$. In this case the integral is defined as

$$\int_{-\infty}^{\infty} f(x)\, dM(x) = \lim_{\substack{T_1\to-\infty \\ T_2\to\infty}} \int_{T_1}^{T_2} f(x)\, dM(x) \qquad \text{in } L^2 \text{ sense.}$$

In our first statement we provide a condition for the existence of the integral.

Lemma 2.8. *The integral $\int_{\mathbb{R}} f(x)dM(x)$ is well-defined if and only if*

$$\|f\|^2_{\mathbb{R}} = E \int_{\mathbb{R}} f^2(x)\, d\langle M\rangle_x < \infty \,.$$

In this case $\int_{\mathbb{R}} f(x)dM(x)$ has mean 0 and finite variance $\|f\|^2_\infty$.

Proof. If the integral $\int_{\mathbb{R}} f(x)dM(x)$ exists then by definition it is the L^2 limit of the variables $\int_{T_1}^{T_2} f(x)dM(x)$, which implies that $\int_{\mathbb{R}} f(x)dM(x)$ has mean 0 and variance

$$\lim_{\substack{T_1\to-\infty \\ T_2\to\infty}} \|f\|^2_{[T_1,T_2]} = \|f\|^2_{\mathbb{R}} < \infty \,.$$

Contrary, let $S_1, T_1 \to -\infty$ and $S_2, T_2 \to \infty$ such that $T_1 \le S_1 < 0 < S_2 \le T_2$, and let $\mathbb{1}_B(x)$, $x \in \mathbb{R}$, denote the indicator function of the Borel set B. Then we have

$$\begin{aligned} E\bigg[\int_{T_1}^{T_2} f(x)\, dM(x) - \int_{S_1}^{S_2} f(x)\, dM(x) \bigg]^2 &= E\bigg[\int_{T_1}^{T_2} f(x)\big(1 - \mathbb{1}_{[S_1,S_2]}(x)\big)\, dM(x) \bigg]^2 \\ &= E \int_{[T_1,T_2]} f^2(x)\big(1^2 - \mathbb{1}^2_{[S_1,S_2]}(x)\big)\, d\langle M\rangle_x \\ &= \|f\|^2_{[T_1,T_2]} - \|f\|^2_{[S_1,S_2]} \le \|f\|^2_{\mathbb{R}} - \|f\|^2_{[S_1,S_2]} \to 0 \,, \end{aligned}$$

and hence, the integrals $\int_{T_1}^{T_2} f(x)dM(x)$, $T_1 < 0 < T_2$, form a Cauchy system in the Hilbert space of all square integrable random variables being defined on the underlying probability space and having mean 0. This implies that they have finite L^2 limit as $T_1 \to -\infty$ and $T_2 \to \infty$. □

In many applications the local martingale $M(x)$ can be represented by the form $W(F(x))$, $x \in \mathbb{R}$, where $W(u)$, $u \geq 0$, is a standard Wiener process on the underlying probability space and $F(x)$, $x \in \mathbb{R}$, is the right-continuous distribution function of an arbitrary random variable X. In our next result we investigate the stochastic integrals of functions $f(x,t)$, $x \in \mathbb{R}$, parametrized by the variable t.

Lemma 2.9. *Consider a family of deterministic functions $f(x,t)$, $x \in \mathbb{R}$, parametrized by $t \in \mathbb{T} \subseteq \mathbb{R}$, and assume that the integral*

$$I(t) = \int_{\mathbb{R}} f(x,t)\, dW\big(F(x)\big)$$

exists for every $t \in \mathbb{T}$. Then, $I(t)$ is a Gaussian process with pointwise mean 0 and covariance function

$$\operatorname{Cov}\big(I(s), I(t)\big) = \int_{\mathbb{R}} f(x,s) f(x,t)\, dF(x)\,, \qquad s,t \in \mathbb{T}\,.$$

Proof. First, note that $W(F(x))$, $x \in \mathbb{R}$, is a centered and càdlàg martingale for any distribution function $F(x)$, and it has the pointwise second moment and the quadratic characteristic

$$EW^2\big(F(x)\big) = \big\langle W(F)\big\rangle_x = F(x)\,, \qquad x \in \mathbb{R}\,.$$

To show that the process $I(t)$ is Gaussian on $\mathbb{T}$ fix an arbitrary integer $d = 1, 2, \ldots$ and parameters $t_1, \ldots, t_d \in \mathbb{T}$, and consider the d-dimensional vector valued function

$$f_d(x; t_1, \ldots, t_d) = \big(f(x,t_1), \ldots, f(x,t_d)\big)\,, \qquad x \in \mathbb{R}\,.$$

By the standard definition of the stochastic integral the integral of f_d on $[-T,T]$ is

$$\begin{aligned} &\int_{-T}^{T} f_d(x; t_1, \ldots, t_d)\, dW\big(F(x)\big) \\ &\qquad = \bigg(\int_{-T}^{T} f(x,t_1)\, dW\big(F(x)\big), \ldots, \int_{-T}^{T} f(x,t_d)\, dW\big(F(x)\big) \bigg)\,, \end{aligned} \tag{2.15}$$

and the properties of the Wiener process imply that (2.15) has d-dimensional normal distribution. Since each component of (2.15) converges in L^2 sense to the corresponding integral considered on the whole real line, the vector variable has the L^2 limit

$$\bigg(\int_{-\infty}^{\infty} f(x,t_1)\, dW\big(F(x)\big), \ldots, \int_{-\infty}^{\infty} f(x,t_d)\, dW\big(F(x)\big) \bigg)\,, \tag{2.16}$$

as $T \to \infty$. Because the L^2 convergence implies the convergence in distribution, and the family of normal distributions is closed under the weak convergence, we immediately obtain that (2.16) has normal law. Then, the process $I(t)$, $t \in \mathbb{T}$, has normal finite dimensional distributions, and hence, it is Gaussian.

Using Lemma 2.8 it is clear that $I(t)$ has mean 0 at every $t \in \mathbb{T}$. Since the covariance matrix of (2.15) converges to that of (2.16) it follows that

$$\begin{aligned}\operatorname{Cov}\big(I(s), I(t)\big) &= \lim_{T\to\infty} \operatorname{Cov}\left(\int_{-T}^{T} f(x,s)\, dW\big(F(x)\big), \int_{-T}^{T} f(x,t)\, dW\big(F(x)\big) \right) \\ &= \lim_{T\to\infty} \int_{[-T,T]} f(x,s) f(x,t)\, d\big\langle W(F)\big\rangle_x = \int_{\mathbb{R}} f(x,s) f(x,t)\, dF(x)\,,\end{aligned}$$

with any $s, t \in \mathbb{T}$. This completes the proof. □

In our work a Gaussian process presented by the form $I(t)$, $t \in \mathbb{T}$, will acts as the weak limit of some empirical process based on some sample variables. The main goal will be to approximate some functional of the corresponding empirical process by the same functional of the limiting process. In such an applications we will sometimes face the problem that this functional of the Gaussian processes must have bounded density function, or at least continuous distribution function. The following result is a simplified version of Theorem 1 of Tsirel'son (1975), and it will serve us well when we will investigate the supremum functionals of Gaussian processes.

Theorem 2.10 (Tsirel'son, 1975)**.** *Consider variables $X_1, X_2, \ldots$ defined on the same probability space having normal distribution with mean $EX_i \geq 0$, $i = 1, 2, \ldots$. Assume that not all of the X_i's are degenerate and they have supremum*

$$S = \sup_{i\in\mathbb{Z}^+} X_i < \infty \qquad a.s.$$

Let $F_S(x) = P(S \leq s)$, $s \in \mathbb{R}$, denote the distribution function of S, and let

$$s_0 = \inf\big\{ s \in \mathbb{R} : F_S(s) > 0 \big\} \in [-\infty, \infty)$$

stand for the left endpoint of the support of F_S. Then, the distribution function F_S is absolute continuous on the interval (s_0, ∞), and its derivative is bounded on $[s, \infty)$ with any $s > s_0$.

Chapter 3

Bootstrapped parameter estimated empirical processes

3.1 Introduction and preliminary results

Consider a family $\mathcal{F} = \{F(x,\theta) : x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^d\}$ of univariate distributions, and let $X_1, X_2, \ldots$ be independent random variables having common distribution function $F(x,\theta_0)$, $x \in \mathbb{R}$, where $\theta_0 \in \Theta$ is a fixed parameter. Let $F_n(x)$, $x \in \mathbb{R}$, stand for the empirical distribution function of the sample variables $X_1, \ldots, X_n$, and applying some estimation method let $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ be an estimator of θ_0. The parameter estimated empirical process based on the sample $X_1, \ldots, X_n$ can be defined as

$$\hat{\alpha}_n(x) = n^{1/2}\big[F_n(x) - F(x,\hat{\theta}_n)\big]\,, \qquad x \in \mathbb{R}\,. \tag{3.1}$$

The weak convergence of $\hat{\alpha}_n$ to a Gaussian process $G(x)$, $x \in \mathbb{R}$, in the space $D[-\infty,\infty]$ was showed by Durbin (1973), and a few years later Burke et al. (1979) proved a weak approximation theorem for the process under similar regularity conditions as Durbin used. This latter result is stated in Theorem 3.1 in the next section. Note that under a wider set of assumptions Burke et al. (1979) provided almost sure convergence, and also, they provided a rate for the convergence.

Since the famous result of Durbin (1973) statistical procedures based on the estimated empirical process became widely used to test goodness-of-fit. These methods usually are not distribution free, that is, the limiting process, and thus the asymptotic critical values of the test statistics depend not only on $\mathcal{F}$, but also on the unknown parameter θ_0, as well. Consequently, we can not test directly the original hypotheses $\mathcal{H}_0$, just the fit of the sample to a fixed element of the distribution family. To make the situation worse, there is no known theoretical way to calculate the distribution of non-trivial functionals of the limiting process G, so the critical values can be determined only by computer simulation separately for each θ_0. To overcome these difficulties, we can apply the bootstrapped version of the estimated empirical process, which requires computer work, but provides distribution free procedures.

Let $\hat{F}_n(x)$ be an estimator of $F(x,\theta_0)$, $x\in\mathbb{R}$, based on $X_1,\dots,X_n$, and generate independent sample elements $X^*_{1,n},\dots,X^*_{m_n,n}$ having distribution function $\hat{F}_n$. Consider the corresponding empirical distribution function

$$F^*_{m_n,n}(x)=\frac{1}{m_n}\sum_{i=1}^{m_n}\mathbb{1}_{\{X^*_{i,n}\le x\}}\,,\qquad x\in\mathbb{R}\,,$$

and let θ^*_n be an estimator of θ_0 based on the bootstrapped variables. The bootstrapped estimated empirical process is the estimated empirical process pertaining to the generated sample, and it is defined by the form

$$\bar{\alpha}^*_{m_n,n}(x)=m_n^{1/2}\big[F^*_{m_n,n}(x)-F(x,\theta^*_n)\big]\,,\qquad x\in\mathbb{R}\,. \tag{3.2}$$

The resampling method is called parametric bootstrap if $\hat{F}_n(x)=F(x,\hat{\theta}_n)$, and the method is called non-parametric bootstrap or Efron bootstrap if $\hat{F}_n(x)=F_n(x)$, $x\in\mathbb{R}$. In the latter case the bootstrapped values $X^*_{1,n},\dots,X^*_{m_n,n}$ are obtained by sampling n times with replacement from the original observations such that for each of the n selections, each X_k has probability $1/n$ of being chosen.

To understand the essence of the bootstrap technique, consider any continuous functional $\psi(\hat{\alpha}_n)$ of the process $\hat{\alpha}_n$ as a test statistics for $\mathcal{H}_0$. Since $\hat{\alpha}_n$ converges weakly to G, the asymptotic critical values of the statistics are the proper theoretical quantiles of $\psi(G)$. Whenever the bootstrapped process $\bar{\alpha}^*_{m_n,n}$ converges in distribution to the same limit G, the critical values can be replaced by the empirical quantiles of $\psi(\bar{\alpha}^*_{m_n,n})$ based on a large number of independent sample generations, which process can be made by using computers. The most important benefit of this method is that we do not need to specify the parameter θ_0 in the algorithm.

The parametric bootstrap was introduced by Stute et al. (1993), and it was designed directly to test goodness-of-fit for parametric distribution families. The idea behind their method might be the Monte Carlo technique, with which one can investigate the distribution of a random variable by generating large number of observations. Since in the applications we do not know the parameter θ_0, heuristically, it seems to be a good idea to replace it with the estimator $\hat{\theta}_n$, and generate realizations of the process $\hat{\alpha}_n$ based on samples having distribution function $F(x,\hat{\theta}_n)$, $x\in\mathbb{R}$. The parametric bootstrap method works exactly this way. For a fixed sample size $m_n\in\mathbb{N}$ consider independent sample variables $X^*_{1,n},\dots,X^*_{m_n,n}$ coming from the distribution $F(x,\hat{\theta}_n)$, $x\in\mathbb{R}$, and let

$$\hat{\theta}^*_n=\hat{\theta}^*_n(X^*_{1,n},\dots,X^*_{m_n,n})$$

be an estimator of the parameter $\hat{\theta}_n$. The parametric bootstrap estimated empirical process is the parameter estimated empirical process based on the bootstrapped sample, and it is defined by the form

$$\hat{\alpha}^*_{m_n,n}(x)=m_n^{1/2}\big[F^*_{m_n,n}(x)-F(x,\hat{\theta}^*_n)\big]\,,\qquad x\in\mathbb{R}\,. \tag{3.3}$$

The process was investigated by several researchers in the recent years. Stute et al. (1993) and Babu and Rao (2004) proved that $\hat{\alpha}^*_{m_n,n}$ converges to G in the case $m_n=n$

for continuous distribution families under slightly different groups of assumptions. Henze (1996) showed the convergence of the non-negative integer valued version defined by $Z_n(k) = \hat{\alpha}^*_{m_n,n}(k)$, $k = 0, 1, 2, \ldots$, as a random element of the Banach space of all real sequences converging to zero, endowed with the supremum norm. An interesting and useful variant of the procedure was investigated by Genz and Haeusler (2006) for processes under auxiliary information.

The non-parametric bootstrap of Efron (1979) originally was designed for estimating variance and for constructing confidence intervals, not for testing goodness-of-fit. Burke and Gombay (1988) were the first who defined a non-parametric bootstrap based empirical process similar to the one in formula (3.4) as a tool for testing composite hypothesis, and they proved the convergence of their process to G. Later, Babu and Rao (2004), in the same paper where they investigated $\hat{\alpha}^*_{m_n,n}$, introduced the non-parametric version of the process. To obtain a non-parametric bootstrap sample having size m_n select sample variables $X^*_{1,n}, \ldots, X^*_{m_n,n}$ among the original observations $X_1, \ldots, X_n$ with replacement such that each X_k has probability $1/n$ of being chosen each time. Let

$$\tilde{\theta}^*_n = \tilde{\theta}^*_n(X^*_{1,n}, \ldots, X^*_{m_n,n})$$

be an estimator of θ_0, and let $F^*_{m_n,n}(x)$, $x \in \mathbb{R}$, stand for the bootstrapped empirical distribution function. The non-parametric bootstrap estimated empirical process is

$$\tilde{\alpha}^*_{m_n,n} = m_n^{1/2}\left[F^*_{m_n,n}(x) - F(x, \tilde{\theta}^*_n)\right], \qquad x \in \mathbb{R}. \tag{3.4}$$

Babu and Rao (2004) found that the process does not converge to G in distribution, because it requires bias correction, and they proved that in $\tilde{\alpha}^*_{m_n,n} - \hat{\alpha}_n$ converges to G for $m_n = n$ in continuous distribution families. As a result, the critical values of a given test statistics $\psi(\hat{\alpha}_n)$ can not be determined as the empirical quantiles of $\psi(\tilde{\alpha}^*_{n,n})$ as in the parametric case, but the method works if we apply the functional $\psi(\tilde{\alpha}^*_{n,n} - \hat{\alpha}_n)$. For more details see Babu and Rao (2004), where the parametric and the non-parametric bootstrap procedure are illustrated in the uniform, the normal and the Cauchy family.

The aim of the chapter is to state and prove weak approximation theorems for the parametric and the non-parametric bootstrapped estimated empirical process under some general conditions, which contain the most important continuous and discrete distribution families and estimation methods. That is, we will construct a representation of $\hat{\alpha}^*_{m_n,n}$ and $\tilde{\alpha}^*_{m_n,n}$, and copies G_n of the limiting process G, $n = 1, 2, \ldots$, such that the supremum distance between the bootstrapped processes and G_{m_n} converges to zero in probability. This implies the weak convergence of $\hat{\alpha}^*_{m_n,n}$ and $\tilde{\alpha}^*_{m_n,n}$ in the space $D[-\infty, \infty]$. We present our main results in Section 3.2 with a short discussion of the conditions. The proofs can be found in Section 3.3. The approximations will be highly valued later in Section 4.9, where we will introduce the related probability generator processes, and we will apply the results of this chapter to prove their convergence. The bootstrap test algorithm and its theoretical background for the Kolmogorov–Smirnov type functional of the processes is provided in Section 3.4. Finally, in Section 3.5 we check the validity of the assumptions of the method for the Poisson and for the normal distribution, and in Section 3.6 we report on a simulation study in these families.

3.2 Assumptions and results

Consider the distribution family $\mathcal{F} = \{F(x,\theta) : x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^d\}$, and let θ_0 be a fixed and let $\theta = (\theta^{(1)}, \dots, \theta^{(d)})$ be an arbitrary vector in the set Θ. Let $X_1, X_2, \dots$ and $Y_1, Y_2, \dots$ be sequences of independent and identically distributed random variables having distribution function $F(x,\theta_0)$ and $F(x,\theta)$, $x \in \mathbb{R}$, respectively. Also, consider a statistical function

$$\theta_n = \theta_n(Y_1, \dots, Y_n)$$

as an estimator of the general parameter θ, and let $\hat{\theta}_n = \theta_n(X_1, \dots, X_n)$ be the estimation of θ_0 based on $X_1, \dots, X_n$. Let m_n be the bootstrap sample size, and based on the bootstrapped sample let $\hat{\theta}_n^*$ and $\tilde{\theta}_n^*$ denote the estimator of $\hat{\theta}_n$ in the parametric and of θ_0 in the non-parametric case, respectively. Unless otherwise specified, all asymptotic statements are meant as $n \to \infty$.

Assumption 1. We will use the following assumptions in our main results.

(a1) The vector

$$\nabla_\theta F(x,\theta) = \left(\frac{\partial}{\partial \theta^{(1)}} F(x,\theta), \dots, \frac{\partial}{\partial \theta^{(d)}} F(x,\theta) \right)$$

of partial derivatives exists for all $(x,\theta) \in \mathbb{R} \times \Lambda$, where the set $\Lambda \subseteq \Theta$ is a proper neighborhood of θ_0.

(a2) $\nabla_\theta F(x,\theta)$, $x \in \mathbb{R}$, converges uniformly to $\nabla_\theta F(x,\theta_0)$, $x \in \mathbb{R}$, as $\theta \to \theta_0$.

(a3) $\nabla_\theta F(x,\theta_0)$, $x \in \mathbb{R}$, is bounded.

(a4) There exist Borel measurable functions $l(\theta_0) : \mathbb{R} \to \mathbb{R}^d$ and $\varepsilon_n(\theta_0) : \mathbb{R}^n \to \mathbb{R}^d$ such that

$$\hat{\theta}_n - \theta_0 = \frac{1}{n} \sum_{i=1}^{n} l(X_i, \theta_0) + n^{-1/2} \varepsilon_n(\theta_0) \qquad \text{a.s.}$$

holds with $\varepsilon_n(\theta_0) = \varepsilon_n(X_1, \dots, X_n, \theta_0)$.

(a5) There exist Borel measurable functions $l : \mathbb{R} \times \Lambda \to \mathbb{R}^d$ and $\varepsilon_n : \mathbb{R}^n \times \Lambda \to \mathbb{R}^d$ such that

$$\theta_n - \theta = \frac{1}{n} \sum_{i=1}^{n} l(Y_i, \theta) + n^{-1/2} \varepsilon_n(\theta) \qquad \text{a.s.}$$

holds with $\varepsilon_n(\theta) = \varepsilon_n(Y_1, \dots, Y_n, \theta)$ for every $\theta \in \Lambda$.

(a6) There exist a Borel measurable function $\varepsilon_{m,n}(\theta_0) : \mathbb{R}^{n+m} \to \mathbb{R}^d$ such that

$$\tilde{\theta}_n^* - \hat{\theta}_n = \frac{1}{m} \sum_{i=1}^{m} l(X_{i,n}^*, \theta_0) - \frac{1}{n} \sum_{i=1}^{n} l(X_i, \theta_0) + m^{-1/2} \varepsilon_{m,n}(\theta_0) \qquad \text{a.s.}$$

holds with the function l defined in (a4), and with the non-parametric bootstrap sample $X_{1,n}^*, \dots, X_{m,n}^*$, and with $\varepsilon_{m,n}(\theta_0) = \varepsilon_{m,n}(X_1, \dots, X_n, X_{1,n}^*, \dots, X_{m,n}^*, \theta_0)$.

(a7) $E\, l(X_i, \theta_0) = 0$.

(a8) $E\, l(Y_i, \theta) = 0$ for every $\theta \in \Lambda$.

(a9) $M(\theta_0) = E\, l(X_i, \theta_0)^T l(X_i, \theta_0)$ is a finite non-negative definite matrix.

(a10) $M(\theta) = E\, l(Y_i, \theta)^T l(Y_i, \theta)$ is a finite non-negative definite matrix for every $\theta \in \Lambda$.

(a11) The function $M(\theta)$, $\theta \in \Lambda$, is continuous at θ_0.

(a12) Each component of $l(x, \theta_0)$, $x \in \mathbb{R}$, is of bounded variation on every finite interval.

(a13) $l(x, \theta)$ converges uniformly to $l(x, \theta_0)$, $x \in \mathbb{R}$, on every finite interval as $\theta \to \theta_0$.

(a14) $\varepsilon_n(\theta_0) \xrightarrow{P} 0$.

(a15) $\varepsilon_{m_n}(\hat{\theta}_n) \xrightarrow{P} 0$.

(a16) $\varepsilon_{m_n,n}(\theta_0) \xrightarrow{P} 0$.

Throughout the chapter the primary notation of vectors refers to row vectors, and V^T and $V^{(k)}$ will stand for the transpose and the k-th component of a row vector V. As further notation, $M^{(j,k)}(\theta)$ will denote the element of the matrix $M(\theta)$ in the j-th row and k-th column. Also, for a d-dimensional vector valued function $g(x, \theta)$ we use the notation

$$g^2(x, \theta) = \Big((g^{(k)}(x, \theta))^2, \dots, (g^{(k)}(x, \theta))^2 \Big)$$

The norm $\| \cdot \|$ on $\mathbb{R}^d$ is the usual Euclidean norm. The following theorem is due to Burke et al. (1979).

Theorem 3.1. *(Burke, Csörgő, Csörgő and Révész, 1978) Assume that the distribution family $\mathcal{F}$, the fixed parameter θ_0 and the applied estimation method satisfy conditions (a1)–(a4), (a7), (a9), (a12) and (a14). Then, on a suitable probability space, one can construct independent random variables $X_1, X_2, \dots$, having common distribution function $F(x, \theta_0)$, and a sequence of Brownian bridges $B_1, B_2, \dots$, such that the parameter estimated empirical process $\hat{\alpha}_n(x)$, $x \in \mathbb{R}$, based on the variables $X_1, \dots, X_n$ satisfies*

$$\sup_{x \in \mathbb{R}} \big| \hat{\alpha}_n(x) - G_n(x) \big| \xrightarrow{P} 0\,, \qquad n \to \infty\,,$$

with the process

$$G_n(x) = B_n\big(F(x, \theta_0)\big) - \left[\int_{\mathbb{R}} l(x, \theta_0)\, dB_n\big(F(x, \theta_0)\big) \right] \nabla_\theta F(x, \theta_0)^T\,, \qquad x \in \mathbb{R}\,.$$

As a consequence of Theorem 3.1 we obtain the well-known result of Durbin (1973), the weak convergence of $\hat{\alpha}_n$ to the process

$$G(x) = B\big(F(x, \theta_0)\big) - \left[\int_{\mathbb{R}} l(x, \theta_0)\, dB\big(F(x, \theta_0)\big) \right] \nabla_\theta F(x, \theta_0)^T\,, \qquad x \in \mathbb{R}\,, \tag{3.5}$$

where $B(u)$, $0 \le u \le 1$, is an arbitrary Brownian bridge. Note that G is a mean-zero Gaussian process having covariance function

$$\begin{aligned}\operatorname{Cov}\big(G(x), G(y)\big) = F(\min(x,y),\theta_0) - F(x,\theta_0)F(y,\theta_0) - J(x)\nabla_\theta F(y,\theta_0)^T \\ - J(y)\nabla_\theta F(x,\theta_0)^T + \nabla_\theta F(x,\theta_0) M(\theta_0) \nabla_\theta F(y,\theta_0)^T ,\end{aligned}$$

for $x, y \in \mathbb{R}$, with $J(x) = \int_{-\infty}^{x} l(z,\theta_0)\, dF(z,\theta_0)$. Now, we state our main results in the subject of bootstrapped empirical processes. Theorem 3.2 deals with the parametric and Theorem 3.3 is about the non-parametric case.

Theorem 3.2. *Assume that the bootstrap sample size $m_n \to \infty$, and assume that the distribution family $\mathcal{F}$, the fixed parameter θ_0 and the applied estimation method satisfy conditions (a1)–(a3), (a5), (a8), (a10)–(a15). Then, on a suitable probability space, one can construct random variables X_i and $X_{i,\theta}$, $\theta \in \Theta$, $i = 1, 2, \dots$, having distribution function $F(x,\theta_0)$ and $F(x,\theta)$, respectively, and a sequence of Brownian bridges $B_1, B_2, \dots$, such that the random variables $X_1, X_{1,\theta}, X_2, X_{2,\theta}, \dots$ are independent for every θ, and the parametric bootstrapped estimated empirical process of (3.3) based on the variables $X_1, \dots, X_n$ and on the parametric bootstrap sample*

$$\big(X^*_{1,n}, \dots, X^*_{m_n,n}\big) = \big(X_{1,\hat\theta_n}, \dots, X_{m_n,\hat\theta_n}\big)\,, \qquad n = 1, 2, \dots\,,$$

satisfies

$$\sup_{x\in\mathbb{R}} \big|\hat\alpha^*_{m_n,n}(x) - G_{m_n}(x)\big| \xrightarrow{P} 0\,, \qquad n \to \infty\,,$$

where $G_1, G_2, \dots$ are defined by (3.5) based on $B_1, B_2, \dots$

Theorem 3.3. *Assume that the bootstrap sample size $m_n \to \infty$, and assume that the distribution family $\mathcal{F}$, the parameter θ_0 and the estimation method satisfy conditions (a1)–(a4), (a6), (a7), (a9), (a12), (a14) and (a16). Then, on a suitable probability space, one can construct independent random variables $X_1, X_2, \dots$ with distribution function $F(x,\theta_0)$, and non-parametric bootstrap sample variables $X^*_{1,n}, \dots, X^*_{m_n,n}$, for $n = 1, 2, \dots$, and a sequence of Brownian bridges $B_1, B_2, \dots$, such that the non-parametric bootstrapped estimated empirical process defined in formula (3.4) based on the variables $X_1, \dots, X_n$ and bootstrap sample $X^*_{1,n}, \dots, X^*_{m_n,n}$ satisfies*

$$\sup_{x\in\mathbb{R}} \left|\tilde\alpha^*_{m_n,n}(x) - \left(\frac{m_n}{n}\right)^{1/2} \hat\alpha_n(x) - G_{m_n}(x)\right| \xrightarrow{P} 0\,, \qquad n \to \infty\,,$$

where $\hat\alpha_n$ is the estimated empirical process of (3.1) and $G_1, G_2, \dots$ are defined by (3.5) based on $B_1, B_2, \dots$

Since the process G_{m_n} have the same distribution as G for every n, we obtain the following consequence of Theorems 3.2 and 3.3.

Corollary 3.4. *Under the assumptions of Theorems 3.2 and 3.3 the processes*

$$\hat\alpha^*_{m_n,n}(x) \qquad \text{and} \qquad \tilde\alpha^*_{m_n,n}(x) - \left(\frac{m_n}{n}\right)^{1/2} \hat\alpha_n(x)\,, \qquad x \in \mathbb{R}\,,$$

converges weakly to $G(x)$, $x \in \mathbb{R}$, in the space $D[-\infty, \infty]$.

For the proofs of the theorems we drew inspiration from two papers. The construction of the random elements is based on the method of Csörgő and Mason (1989) who introduced the idea of representing the original and the bootstrapped sample variables on the product of two KMT spaces. Fortunately, this technique does not require any regularity condition on the model. Also, we adopt the approximation method which was applied for the non-bootstrapped process $\hat{\alpha}_n$ by Burke et al. (1979), from where we inherit the conditions of Theorem 3.1 along with the technique. That is, we require the existence and the smoothness of the function $\nabla_\theta F(x,\theta)$ in a neighborhood of θ_0, and we must also take some assumptions on the regularity of the applied estimation method at the single point θ_0. In the non-parametric bootstrap case we do not need much more additional assumptions, we only need a similar sum representation for the bootstrapped parameter estimator $\tilde{\theta}_n^*$ as we already have in the non-bootstrapped model. Contrary, in the parametric bootstrap case the bootstrapped sample comes from the distribution $F(x,\hat{\theta}_n)$, $x \in \mathbb{R}$, and hence, we must extend the assumptions on the estimation method for a neighborhood of θ_0, and we need uniformity for the functions $M(\theta)$ and $l(x,\theta)$. Surveying the earlier proofs for the convergence of the bootstrapped processes by Stute et al. (1993), Babu and Rao (2004) and the version entertained by Henze (1996) we meet similar conditions.

3.3 Proofs of Theorems 3.2 and 3.3

The construction of the random variables and function of Theorems 3.2 and 3.3 is based on the Hungarian construction presented in Theorem 2.2. This provides independent variables $U_{1,1}, U_{1,2}, \ldots$ distributed uniformly on the interval $[0,1]$ and a sequence of Brownian bridges $B_{1,1}(u), B_{1,2}(u), \ldots$ on an appropriate probability space $(\Omega_1, \mathcal{A}_1, P_1)$, such that the related uniform empirical process $\beta_{1,n}(u)$, $0 \le u \le 1$, defined in (2.1) satisfies

$$\sup_{0\le u\le 1} \left|\beta_{1,n}(u) - B_{1,n}(u)\right| = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right), \qquad \text{a.s.} \tag{3.6}$$

Now, let $(\Omega_2, \mathcal{A}_2, P_2)$ be an other KMT probability space which carries the random variables $U_{2,i}$, empirical processes $\beta_{2,i}$ and Brownian bridges $B_{2,i}$, $i = 1, 2, \ldots$. Define $(\Omega, \mathcal{A}, P)$ as the product probability space of $(\Omega_1, \mathcal{A}_1, P_1)$ and $(\Omega_2, \mathcal{A}_2, P_2)$, and consider the random elements

$$U_i(\omega) = U_{1,i}(\omega_1), \quad U_i^*(\omega) = U_{2,i}(\omega_2), \quad B_i(u,\omega) = B_{2,i}(u,\omega_2), \tag{3.7}$$

$\omega = (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 = \Omega$, $i = 1, 2, \ldots$. The empirical process based on the U_i^*'s is $\beta_n(\omega) = \beta_{2,n}(\omega_2)$, and (3.6) implies that

$$\sup_{0\le u\le 1} |H_n(u)| = \sup_{0\le u\le 1} \left|\beta_n(u) - B_n(u)\right| = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right) \tag{3.8}$$

holds with $H_n(u) = \beta_n(u) - B_n(u)$, $0 \le u \le 1$.

Let F_n be the empirical distribution of the sample variables $X_1, \dots, X_n$, and let

$$F^{-1}(u,\theta) = \inf\{x : F(x,\theta) \le u\}\,, \quad F_n^{-1}(u) = \inf\{x : F_n(x) \le u\}\,, \quad 0 \le u \le 1\,,$$

denote the quantile function pertaining to $F(x,\theta)$ and $F_n(x)$, respectively. Using the elementary properties of the quantile function, the variables

$$X_i = F^{-1}(U_i, \theta_0)\,, \qquad i = 1, 2, \dots \tag{3.9}$$

are independent and have the same distribution function $F(x,\theta_0)$, $x \in \mathbb{R}$. Also,

$$X_{i,\theta} = F^{-1}(U_i^*, \theta)\,, \qquad i = 1, 2, \dots\,, \quad \theta \in \Theta\,, \tag{3.10}$$

are independent from each other and from the X_i's, and they have the same distribution function $F(x,\theta)$ for every $\theta \in \Theta$, that is, $X_{1,\hat\theta_n}, \dots, X_{m_n,\hat\theta_n}$ is a parametric bootstrap sample for every n. The conditional empirical process of the bootstrap sample with respect to the variables $X_1, \dots, X_n$ is

$$\begin{aligned} m_n^{1/2}\big[F^*_{m_n,n}(x) - F(x,\hat\theta_n)\big] &= m_n^{1/2}\left[\frac{1}{m_n}\sum_{i=1}^{m_n} \mathbb{1}_{\{X_{i,\hat\theta_n} \le x\}} - F(x,\hat\theta_n)\right] \\ &= m_n^{1/2}\left[\frac{1}{m_n}\sum_{i=1}^{m_n} \mathbb{1}_{\{U_i^* \le F(x,\hat\theta_n)\}} - F(x,\hat\theta_n)\right] = \beta_{m_n}\big(F(x,\hat\theta_n)\big) \end{aligned} \tag{3.11}$$

By similar reasons

$$X^*_{i,n} = F_n^{-1}(U_i^*)\,, \qquad i = 1, \dots, m_n\,, \tag{3.12}$$

are conditionally independent from each other and have conditional theoretical distribution function $F_n(x)$, $x \in \mathbb{R}$, given $X_1, \dots, X_n$. Hence, $X^*_{1,n}, \dots, X^*_{m_n,n}$ is a nonparametric bootstrap sample based on $X_1, \dots, X_n$ for $n = 1, 2, \dots$, and its conditional empirical process is

$$m_n^{1/2}\big[F^*_{m_n,n}(x) - F_n(x)\big] = \beta_{m_n}\big(F_n(x)\big)\,. \tag{3.13}$$

The rest is to show that the constructed variables and Brownian bridges satisfy the statements in Theorems 3.2 and 3.3.

Throughout the proofs we may and do assume that Λ is a sphere with center θ_0 and radius $r > 0$. Using (a2) and (a3) we can and do assume that the components of the function $\nabla_\theta F(x,\theta)$ are bounded on $\mathbb{R} \times \Lambda$ by a positive value M_0. Also, in the parametric bootstrap case by (a11) we can assume that the components of $M(\theta)$ are bounded on Λ by the same M_0. In the proofs we often use the fact that if a real valued random sequence A_n, $n = 1, 2, \dots$ converges to 0 in probability, then there exists

$$p_n \to 0 \qquad \text{such that} \qquad P\big(|A_n| > p_n\big) < p_n\,, \qquad n = 1, 2, \dots \tag{3.14}$$

Also, if E_n, $n = 1, 2, \dots$ is a sequence of events such that $P(E_n) \to 1$, and we have $P(|A_n| > \varepsilon | E_n) \to 0$ for any $\varepsilon > 0$, then applying the complementer event $\bar E_n = \Omega \setminus E_n$ we find that

$$\begin{aligned} P\big(|A_n| > \varepsilon\big) &= P\big(|A_n| > \varepsilon \mid E_n\big)P(E_n) + P\big(|A_n| > \varepsilon \mid \bar E_n\big)P(\bar E_n) \\ &\le P\big(|A_n| > \varepsilon \mid E_n\big) + P(\bar E_n) \to 0\,. \end{aligned}$$

That is,

$$P\big(|A_n| > \varepsilon \mid E_n\big) \to 0\,,\ \forall \varepsilon > 0\,, \quad \text{and} \quad P(E_n) \to 1\,, \quad \text{imply} \quad A_n \xrightarrow{P} 0\,. \tag{3.15}$$

Introduce the modulus of continuity of a deterministic or random function $f(u)$, $0 \le u \le 1$, by the formula

$$w(f,p) = \sup_{\substack{0\le u\le v\le 1\\ v-u\le p}} |f(v) - f(u)|\,,$$

and observe that $w(B_m,p)$, $m = 1,2,\dots$ have the same distribution as $w(B,p)$ with an arbitrary Brownian bridge B. It is known that based on a standard Wiener process $W(u)$ the random function $B(u) = W(u) - uW(1)$, $0 \le u \le 1$, is a Brownian bridge, and Lévy's famous result for the modulus of continuity of the Wiener implies that

$$w(W,p) \sim \big(2p\log(1/p)\big)^{1/2}\,, \qquad p \to 0\,, \qquad \text{a.s.}$$

We obtain that the modulus of continuity of the Brownian bridge is

$$\begin{aligned} w(B,p) &= \sup_{\substack{0\le u\le v\le 1\\ v-u\le p}} \Big|\big(W(v) - vW(1)\big) - \big(W(u) - uW(1)\big)\Big| \\ &\le \sup_{\substack{0\le u\le v\le 1\\ v-u\le p}} \big|W(v) - W(u)\big| + \sup_{\substack{0\le u\le v\le 1\\ v-u\le p}} \big|vW(1) - uW(1)\big| \le w(W,p) + p|W(1)| \to 0\,, \end{aligned}$$

almost surely as $p \to 0$, and it follows that

$$P\big(w(B_{m_n},p_n) \le \varepsilon\big) \to 1\,, \qquad \forall \varepsilon > 0\,, \qquad n \to \infty\,. \tag{3.16}$$

Proof of Theorem 3.2. First, we examine the asymptotic behavior of the estimated parameters $\hat\theta_n$ and $\hat\theta_n^*$. From (a5) we obtain

$$\big\|\hat\theta_n - \theta_0\big\| \le \sum_{k=1}^{d} \left|\frac{1}{n}\sum_{i=1}^{n} l^{(k)}(X_i,\theta_0)\right| + n^{-1/2}\|\varepsilon_n(\theta_0)\| = \sum_{k=1}^{d} |S_{k,n}| + n^{-1/2}\|\varepsilon_n(\theta_0)\|\,.$$

where the terms $l^{(k)}(X_i,\theta_0)$, $i = 1,2,\dots$, are independent and identically distributed with finite mean 0 and variance $M^{(k,k)}(\theta_0)$ by (a7) and (a9) for every k. The law of large numbers implies that $S_{k,n} \to 0$ almost surely, and by (a14) the second term converges to 0 in probability, too. Hence,

$$\|\hat\theta_n - \theta_0\| \xrightarrow{P} 0\,. \tag{3.17}$$

Similarly, we have

$$\begin{aligned} \big\|\hat\theta_n^* - \hat\theta_n\big\| &\le \sum_{k=1}^{d} \left|\frac{1}{m_n}\sum_{i=1}^{m_n} l^{(k)}(X_{i,\hat\theta_n},\hat\theta_n)\right| + m_n^{-1/2}\|\varepsilon_{m_n}(\hat\theta_n)\| \\ &= \sum_{k=1}^{d} |S_{k,n}^*| + m_n^{-1/2}\|\varepsilon_{m_n}(\hat\theta_n)\|\,. \end{aligned}$$

Using (a8) and (a10) we find that $E[S^*_{k,n} \mid \hat\theta_n] = 0$ and

$$\operatorname{Var}[S^*_{k,n} \mid \hat\theta_n] = \frac{1}{m_n^2}\sum_{i=1}^{m_n} \operatorname{Var} l^{(k)}(X_{i,\hat\theta_n}, \hat\theta_n) = \frac{1}{m_n^2}\sum_{i=1}^{m_n} M^{(k,k)}(\hat\theta_n) \le \frac{dM_0}{m_n}\,,$$

whenever $\hat\theta_n \in \Lambda$, from which

$$E\big[S^*_{k,n} \mid \hat\theta_n \in \Lambda\big] = 0 \qquad \text{and} \qquad \operatorname{Var}\big[S^*_{k,n} \mid \hat\theta_n \in \Lambda\big] \le \frac{dM_0}{m_n}\,.$$

Applying the Chebyshev inequality with the conditional mean and variance we have

$$P\big(|S^*_{k,n}| > \varepsilon \mid \hat\theta_n \in \Lambda\big) \le \frac{\operatorname{Var}\big[S^*_{k,n} \mid \hat\theta_n \in \Lambda\big]}{\varepsilon^2} \le \frac{dM_0}{m_n\varepsilon^2} \to 0\,,$$

for every $\varepsilon > 0$. Since $P(\hat\theta_n \in \Theta) \to 1$ by (3.17), the remark in (3.15) implies that $S^*_{k,n} \to 0$ in probability. Then, using assumption (a15) we have

$$\|\hat\theta^*_n - \hat\theta_n\| \xrightarrow{P} 0\,, \tag{3.18}$$

and formula (3.17) implies that $P(\hat\theta^*_n \in \Lambda) \to 1$ because Λ is a sphere with center θ_0. In the remaining part of the proof we work on the event $E_n = \{\hat\theta_n \in \Lambda\} \cap \{\hat\theta^*_n \in \Lambda\}$, and prove that

$$P\Big(\sup_{x\in\mathbb{R}} \big|\hat\alpha^*_{m_n,n}(x) - G_{m_n}(x)\big| > \varepsilon \,\Big|\, E_n\Big) \to 0 \tag{3.19}$$

holds in case of any $\varepsilon > 0$. Since $P(E_n) \to 1$, the convergence in (3.19) with (3.15) implies Theorem 3.2.

Given E_n condition (a1) implies the one-term Taylor expansion

$$F(x,\hat\theta_n) - F(x,\theta_0) = (\hat\theta_n - \theta_0)\nabla_\theta F(x,\theta_n(x))^T\,,$$

where $\theta_n(x)$ lies in Λ for every integer n and real x. Then, we have

$$\begin{aligned} S_n &= \sup_{x\in\mathbb{R}} \big|F(x,\hat\theta_n) - F(x,\theta_0)\big| \\ &\le \sum_{k=1}^{d} \big|(\hat\theta_n - \theta_0)^{(k)}\big| \sup_{x\in\mathbb{R}} \big|\nabla_\theta F(x,\theta_n(x))^{(k)}\big| \le d\|\hat\theta_n - \theta_0\| M_0 \xrightarrow{P} 0\,. \end{aligned} \tag{3.20}$$

Let

$$T_{2,n} = \sup_{x\in\mathbb{R}} \big|B_{m_n}(F(x,\hat\theta_n)) - B_{m_n}(F(x,\theta_0))\big|\,,$$

and consider a sequence $p_n \to 0$ satisfying (3.14) with $A_n = S_n$. By (3.7) and (3.9) the random variables $X_1, X_2, \dots$ are independent from the processes $B_1, B_2, \dots$, and hence, we have

$$P(T_{2,n} \le \varepsilon) \ge P\big(w(B_{m_n}, p_n) \le \varepsilon, S_n \le p_n\big) = P\big(w(B_{m_n}, p_n) \le \varepsilon\big)P(S_n \le p_n)\,,$$

for any $\varepsilon > 0$. By (3.16) and (3.20) both probability on the right side converge to 1, that is, $T_{2,n} \to 0$ in probability. Next, define

$$K_n(x) = \beta_{m_n}(F(x, \hat{\theta}_n)) - B_{m_n}(F(x, \theta_0))\,.$$

Then, by (3.8) we have

$$\begin{aligned} T_n &= \sup_{x\in\mathbb{R}} |K_n(x)| = \sup_{x\in\mathbb{R}} \left|\beta_{m_n}(F(x, \hat{\theta}_n)) - B_{m_n}(F(x, \theta_0))\right| \\ &\le \sup_{x\in\mathbb{R}} \left|\beta_{m_n}(F(x, \hat{\theta}_n)) - B_{m_n}(F(x, \hat{\theta}_n))\right| + \sup_{x\in\mathbb{R}} \left|B_{m_n}(F(x, \hat{\theta}_n)) - B_{m_n}(F(x, \theta_0))\right| \\ &= \sup_{x\in\mathbb{R}} \left|H_{m_n}(F(x, \hat{\theta}_n))\right| + T_{2,n} = \sup_{0\le u\le 1} |H_{m_n}(u)| + T_{2,n} \xrightarrow{P} 0\,. \end{aligned} \tag{3.21}$$

Now we separate a crucial element of the proof in the following lemma.

Lemma 3.5. *Let* $I_n = \int_{\mathbb{R}} l(x, \theta_0)\, dB_{m_n}(F(x, \theta_0))$. *Then*

$$\left\| m_n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n) - I_n \right\| \xrightarrow{P} 0.$$

Proof. Let $V^{(k)}(y)$ denote the total variation of the component $l^{(k)}(x, \theta_0)$ on the interval $[-y, y]$, which is finite for any $y > 0$ by (a12), and let $V(y) = (V^{(1)}(y), \dots, V^{(d)}(y)$, $y > 0$. We show that for the functions

$$L_{1,n}(x) = l(x, \hat{\theta}_n) - l(x, \theta_0) \quad \text{and} \quad L_{2,n}(x) = l^2(x, \hat{\theta}_n) - l^2(x, \theta_0)\,,$$

we have

$$\sup_{-y\le x\le y} \|L_{1,n}(x)\| \xrightarrow{P} 0\,, \qquad \text{and} \qquad \sup_{-y\le x\le y} \|L_{2,n}(x)\| \xrightarrow{P} 0\,, \tag{3.22}$$

with any fixed $y > 0$. The first convergence immediately comes from assumption (a13) and (3.17). Also, if x lies in $[-y, y]$, then

$$\left|l^{(k)}(x, \theta_0)\right| \le \left|l^{(k)}(x, \theta_0) - l^{(k)}(0, \theta_0)\right| + \left|l^{(k)}(0, \theta_0)\right| \le V^{(k)}(y) + \left|l^{(k)}(0, \theta_0)\right|,$$

which implies that

$$\begin{aligned} \sup_{-y\le x\le y} \left|l^{(k)}(x, \hat{\theta}_n) + l^{(k)}(x, \theta_0)\right| &\le \sup_{-y\le x\le y} \left[\left|l^{(k)}(x, \hat{\theta}_n) - l^{(k)}(x, \theta_0)\right| + 2\left|l^{(k)}(x, \theta_0)\right|\right] \\ &\le \sup_{-y\le x\le y} |L_{1,n}^{(k)}(x)| + 2V^{(k)}(y) + 2\left|l^{(k)}(0, \theta_0)\right| \end{aligned}$$

is a bounded sequence for any fixed y and k. Hence,

$$\begin{aligned} \sup_{-y\le x\le y} \|L_{2,n}(x)\| &\le \sum_{k=1}^{d} \sup_{-y\le x\le y} \left[\left|l^{(k)}(x, \hat{\theta}_n) - l^{(k)}(x, \theta_0)\right| \left|l^{(k)}(x, \hat{\theta}_n) + l^{(k)}(x, \theta_0)\right|\right] \\ &\le \sum_{k=1}^{d} \sup_{-y\le x\le y} \left|L_{1,n}^{(k)}(x)\right| \sup_{-y\le x\le y} \left|l^{(k)}(x, \hat{\theta}_n) + l^{(k)}(x, \theta_0)\right| \to 0\,. \end{aligned}$$

In the next step we show that there exists real valued sequence $y_n \to \infty$ such that

$$T_n\|V(y_n)\| \xrightarrow{P} 0\,, \quad \sup_{-y_n\le x\le y_n} \|L_{1,n}(x)\| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{-y_n\le x\le y_n} \|L_{2,n}(x)\| \xrightarrow{P} 0\,. \tag{3.23}$$

Such a sequence can be constructed in the following way. If $\lim_{y\to\infty}\|V(y)\|$ is finite then the first convergence holds with arbitrary $y_{1,n} \to \infty$. Otherwise, consider p_n of (3.14) with $A_n = T_n$ and let

$$y_{1,n} = \sup\left\{y \ge 0 : \|V(y)\| < p_n^{-1/2}\right\}.$$

By (a12) the values $y_{1,n}$ are finite, and $y_{1,n} \to \infty$ because $\lim_{y\to\infty}\|V(y)\| = \infty$. Since we have $P(T_n\|V(y_{1,n})\| > p_n^{1/2}) < p_n$ for every n, the first convergence in (3.23) holds with $y_{1,n}$. From (3.22) we have

$$\sup_{-k\le x\le k} \|L_{1,n}(x)\| \xrightarrow{P} 0$$

for any fixed $k = 1, 2, \ldots$, and hence, there is a strictly increasing integer valued sequence $n_k \to \infty$ for which

$$P\left(\sup_{-k\le x\le k} \|L_{1,n}(x)\| > \frac{1}{k}\right) < \frac{1}{k}\,,$$

whenever $n > n_k$. Letting $y_{2,n} = k$, $n_k < n \le n_{k+1}$, $k = 1, 2, \ldots$, the second convergence in (3.23) follows with $y_{2,n} \to \infty$, and the third convergence can be guaranteed with a sequence $y_{3,n}$ based on similar construction. Thus $y_n = \min\{y_{1,n}, y_{2,n}, y_{3,n}\} \to \infty$ is a suitable sequence for all three statements in (3.23).

Using (a5), (a8) and (3.11), we have

$$\begin{aligned}
m_n^{1/2}(\hat\theta_n^* - \hat\theta_n) &= m_n^{1/2}\left[\frac{1}{m_n}\sum_{i=1}^{m_n} l\big(X_{i,\hat\theta_n}, \hat\theta_n\big) - E\Big[l\big(X_{1,\hat\theta_n}, \hat\theta_n\big)\big|\hat\theta_n\Big]\right] + \varepsilon_{m_n}(\hat\theta_n) \\
&= m_n^{1/2}\left[\int_{\mathbb{R}} l(x,\hat\theta_n)\,dF^*_{m_n,n}(x) - \int_{\mathbb{R}} l(x,\hat\theta_n)\,dF(x,\hat\theta_n)\right] + \varepsilon_{m_n}(\hat\theta_n) \\
&= \int_{\mathbb{R}} l(x,\hat\theta_n)\,d\beta_{m_n}(F(x,\hat\theta_n)) + \varepsilon_{m_n}(\hat\theta_n)\,.
\end{aligned}$$

To prove the lemma, consider the decomposition

$$\begin{aligned}
m_n^{1/2}(\hat\theta_n^* - \hat\theta_n) - I_n &= \int_{(-y_n,y_n]} L_{1,n}(x)\,d\beta_{m_n}(F(x,\hat\theta_n)) + \int_{(-y_n,y_n]} l(x,\theta_0)\,dK_n(x) \\
&\quad + \int_{R_n} l(x,\hat\theta_n)\,d\beta_{m_n}(F(x,\hat\theta_n)) - \int_{R_n} l(x,\theta_0)\,dB_n(F(x,\theta_0)) + \varepsilon_{m_n}(\hat\theta_n)
\end{aligned}$$

with $R_n = \mathbb{R}\setminus(-y_n, y_n]$, and let $I_{1,n}, I_{2,n}, I_{3,n}$ and $I_{4,n}$ denote the four integrals, respectively. We show that all of the terms converge to 0 in probability.

Define the function $f_n(u)$ as $L_{1,n}(F^{-1}(u,\hat\theta_n))$ on the interval $(F(-y_n,\hat\theta_n), F(y_n,\hat\theta_n)]$ and as 0 otherwise. From (3.23) we have

$$A_n = \sup_{0<u<1} \|f_n(u)\| = \sup_{-y_n<x\le y_n} \|L_{1,n}(x)\| \xrightarrow{P} 0\,.$$

Consider p_n of (3.14) with the sequence A_n and let $E_n' = \{\sup_{0<u<1} \|f_n(u)\| \le p_n\}$. Then, for any $\varepsilon > 0$ we have

$$\begin{aligned} P\big(\|I_{1,n}\| > \varepsilon \mid E_n'\big) &= P\bigg(\Big\| \int_{(0,1)} f_n(u)\, d\beta_{m_n}(u)\Big\| > \varepsilon \,\Big|\, E_n'\bigg) \\ &\le \sum_{k=1}^{d} P\bigg(\Big\| \int_{(0,1)} f_n^{(k)}(u)\, d\beta_{m_n}(u)\Big\| > \frac{\varepsilon}{d} \,\Big|\, E_n'\bigg) \le dC_1 \exp\Big(-\frac{C_2\varepsilon^2}{d^2 p_n^2}\Big) \xrightarrow{P} 0\,, \end{aligned}$$

where C_1 and C_2 are positive universal constants from Theorem 1 of Major (1988). Since $P(E_n') \to 1$, it follows that $I_{1,n} \xrightarrow{P} 0$ by (3.15). Next, integrating by parts and using the definition of the sequence y_n in (3.23), we get

$$\begin{aligned} \|I_{2,n}\| &\le \Big\| l(y_n,\theta_0)K_n(y_n) - l(-y_n,\theta_0)K_n(-y_n)\Big\| + \Big\| \int_{[-y_n,y_n)} K_n(x)\, dl(x,\theta_0)\Big\| \\ &\le \Big[\|l(y_n,\theta_0) - l(0,\theta_0)\| + \|l(-y_n,\theta_0) - l(0,\theta_0)\| + 2\|l(0,\theta_0)\|\Big] T_n + \|V(y_n)\| T_n \\ &\le 3\|V(y_n)\| T_n + 2\|l(0,\theta_0)\| T_n \xrightarrow{P} 0\,. \end{aligned} \tag{3.24}$$

By introducing a standard Wiener process $W(u)$ on an arbitrary probability space the process $B(u) = W(u) - uW(1)$, $0 \le u \le 1$, is a Brownian bridge, and using (a8), (a10) and the elementary properties of the stochastic integral the k-th component of $I_{4,n}$ has mean 0 and variance

$$\begin{aligned} \operatorname{Var} I_{4,n}^{(k)} &= \operatorname{Var}\bigg[\int_{R_n} l^{(k)}(x,\theta_0)\, dW(F(x,\theta_0)) - W(1)\int_{R_n} l^{(k)}(x,\theta_0)\, dF(x,\theta_0)\bigg] \\ &= \operatorname{Var}\bigg[\int_{R_n} l^{(k)}(x,\theta_0)\, dW(F(x,\theta_0))\bigg] + \bigg[\int_{R_n} l^{(k)}(x,\theta_0)\, dF(x,\theta_0)\bigg]^2 \operatorname{Var} W(1) \\ &\quad - 2\int_{R_n} l^{(k)}(x,\theta_0)\, dF(x,\theta_0)\; E\bigg[\int_{R_n} l^{(k)}(x,\theta_0)\, dW(F(x,\theta_0)) \int_{-\infty}^{\infty} 1\, dW(F(x,\theta_0))\bigg] \\ &= \int_{R_n} \big(l^{(k)}(x,\theta_0)\big)^2\, dF(x,\theta_0) + \bigg[\int_{R_n} l^{(k)}(x,\theta_0)\, dF(x,\theta_0)\bigg]^2 \\ &\quad - 2\int_{R_n} l^{(k)}(x,\theta_0)\, dF(x,\theta_0) \int_{R_n} l^{(k)}(x,\theta_0)\, dF(x,\theta_0) \\ &= \int_{R_n} \big(l^{(k)}(x,\theta_0)\big)^2\, dF(x,\theta_0) - \bigg[\int_{R_n} l^{(k)}(x,\theta_0)\, dF(x,\theta_0)\bigg]^2 \\ &\le \int_{R_n} \big(l^{(k)}(x,\theta_0)\big)^2\, dF(x,\theta_0) \to 0\,. \end{aligned} \tag{3.25}$$

Hence, by the Chebyshev inequality,

$$P(\|I_{4,n}\| > \varepsilon) \le \sum_{k=1}^{d} P\Big(\big|I_{4,n}^{(k)}\big| > \frac{\varepsilon}{d}\Big) \le \sum_{k=1}^{d} \Big(\frac{d}{\varepsilon}\Big)^2 \operatorname{Var} I_{4,n}^{(k)} \to 0 \tag{3.26}$$

with an arbitrary $\varepsilon > 0$, and thus, $I_{4,n} \to 0$ in probability. Also, using the conditional independence of the variables $X_{1,\hat\theta_n}, \dots, X_{m_n,\hat\theta_n}$ we obtain the mean

$$\begin{aligned} E\Big[\int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF^*_{m_n,n}(x) \,\Big|\, \hat\theta_n\Big] &= \frac{1}{m_n}\sum_{i=1}^{m_n} E\Big[l^{(k)}(X_{i,\hat\theta_n},\hat\theta_n)\mathbb{1}_{\{X_{i,\hat\theta_n}\in R_n\}} \,\Big|\, \hat\theta_n\Big] \\ &= \int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF(x,\hat\theta_n)\,, \end{aligned} \tag{3.27}$$

and the variance

$$\begin{aligned} E^2\Big[\int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF^*_{m_n,n}(x) \,\Big|\, \hat\theta_n\Big] &= \frac{1}{m_n^2}\sum_{i=1}^{m_n} E\Big[\big(l^{(k)}(X_{i,\hat\theta_n},\hat\theta_n)\big)^2\mathbb{1}_{\{X_{i,\hat\theta_n}\in R_n\}} \,\Big|\, \hat\theta_n\Big] \\ &+ \frac{1}{m_n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{m_n} E\Big[l^{(k)}(X_{i,\hat\theta_n},\hat\theta_n)\mathbb{1}_{\{X_{i,\hat\theta_n}\in R_n\}} \,\Big|\, \hat\theta_n\Big]\; E\Big[l^{(k)}(X_{j,\hat\theta_n},\hat\theta_n)\mathbb{1}_{\{X_{j,\hat\theta_n}\in R_n\}} \,\Big|\, \hat\theta_n\Big] \\ &= \frac{1}{m_n}\int_{R_n} \big(l^{(k)}(x,\hat\theta_n)\big)^2\, dF(x,\hat\theta_n) + \frac{m_n-1}{m_n}\Big[\int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF(x,\hat\theta_n)\Big]^2. \end{aligned} \tag{3.28}$$

Applying (3.27) and (3.28) the variable $I_{3,n}^{(k)}$ has conditional mean

$$\begin{aligned} &E\Big[I_{3,n}^{(k)} \,\Big|\, \hat\theta_n\Big] = E\Big[\int_{R_n} l^{(k)}(x,\hat\theta_n)\, d\beta_{m_n}(F(x,\hat\theta_n)) \,\Big|\, \hat\theta_n\Big] \\ &= m_n^{1/2} E\Big[\int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF^*_{m_n,n}(x) \,\Big|\, \hat\theta_n\Big] - m_n^{1/2}\int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF(x,\hat\theta_n) = 0\,, \end{aligned}$$

and conditional variance

$$\begin{aligned} &\operatorname{Var}\Big[I_{3,n}^{(k)} \,\Big|\, \hat\theta_n\Big] \\ &= m_n E^2\Big[\int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF^*_{m_n,n}(x) \,\Big|\, \hat\theta_n\Big] + m_n\Big[\int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF(x,\hat\theta_n)\Big]^2 \\ &\quad - 2m_n\int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF(x,\hat\theta_n)\; E\Big[\int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF^*_{m_n,n}(x) \,\Big|\, \hat\theta_n\Big] \\ &= \int_{R_n} \big(l^{(k)}(x,\hat\theta_n)\big)^2\, dF(x,\hat\theta_n) - \Big[\int_{R_n} l^{(k)}(x,\hat\theta_n)\, dF(x,\hat\theta_n)\Big]^2. \end{aligned} \tag{3.29}$$

This implies that

$$\begin{aligned}\operatorname{Var}\Big[I_{3,n}^{(k)} \,\Big|\, \hat\theta_n\Big] &\le \int_{R_n} \big(l^{(k)}(x,\hat\theta_n)\big)^2\, dF(x,\hat\theta_n) = \int_{R_n} \big(l^{(k)}(x,\theta_0)\big)^2\, dF(x,\theta_0)\\ &+ \bigg[\int_{\mathbb{R}} \big(l^{(k)}(x,\hat\theta_n)\big)^2\, dF(x,\hat\theta_n) - \int_{\mathbb{R}} \big(l^{(k)}(x,\theta_0)\big)^2\, dF(x,\theta_0)\bigg]\\ &- \int_{(-y_n,y_n]} L_{2,n}^{(k)}(x)\, dF(x,\hat\theta_n) - \int_{(-y_n,y_n]} \big(l^{(k)}(x,\theta_0)\big)^2\, d\big[F(x,\hat\theta_n) - F(x,\theta_0)\big] \xrightarrow{P} 0\,,\end{aligned}$$

because all of the four terms converge to 0 in probability. The first and the second terms do that by (a12) and (a13). The third term follows by (3.23) and by the fact that $F(x,\hat\theta_n)$ is of bounded variation on $\mathbb{R}$. Finally, using the technique in (3.24), one can show that the fourth term also goes to 0 in probability by (3.20) and (3.23). Now, let p_n be the sequence defined in (3.14) corresponding to

$$M_n = \max_{1\le k\le d} \operatorname{Var}\Big[I_{3,n}^{(k)} \,\Big|\, \hat\theta_n\Big] \xrightarrow{P} 0\,.$$

Using again the Chebyshev inequality we have

$$P\big(\|I_{3,n}\| > \varepsilon \mid M_n \le p_n\big) \le \sum_{k=1}^{d} P\Big(\big|I_{3,n}^{(k)}\big| > \frac{\varepsilon}{d} \,\Big|\, M_n \le p_n\Big) \le \sum_{k=1}^{d} \Big(\frac{d}{\varepsilon}\Big)^2 p_n \to 0$$

for every $\varepsilon > 0$, and this implies that $\|I_{3,n}\| \to 0$ in stochastic sense by (3.15) because $P(M_n \le p_n) \to 1$. The term $\varepsilon_{m_n}(\hat\theta_n)$ also converges to 0 by (a15), that is, we have (3.3), and the proof of Lemma 3.5 is completed. □

Proof of Theorem 3.2, continued. Returning to the proof of Theorem 3.2, if E_n occurs then by (a1) we can take the one-term Taylor expansion with respect to $\hat\theta_n$:

$$\begin{aligned}\hat\alpha_{m_n,n}^*(x) &= \beta_{m_n}(F(x,\hat\theta_n)) - m_n^{1/2}[F(x,\hat\theta_n^*) - F(x,\hat\theta_n)]\\ &= B_{m_n}(F(x,\hat\theta_n)) + H_{m_n}(F(x,\hat\theta_n)) - m_n^{1/2}(\hat\theta_n^* - \hat\theta_n)\nabla_\theta F(x,\theta_n^*(x))^T\\ &= G_{m_n}(x) + H_{m_n}(F(x,\hat\theta_n)) - H_{2,n}(x)\,,\end{aligned}$$

where $\theta_n^*(x)$ lies in the sphere Λ for every x and n, and

$$\sup_{x\in\mathbb{R}} \|\theta_n^*(x) - \theta_0\| \le \|\hat\theta_n^* - \hat\theta_n\| + \|\hat\theta_n - \theta_0\| \xrightarrow{P} 0\,.$$

Using (a2), Lemma 3.5, the stochastic boundedness of I_n and the boundedness of the function $\nabla_\theta F(x,\theta)$ on $\mathbb{R}\times\Lambda$ we obtain the convergence

$$H_{2,n}(x) = \big[n^{1/2}(\hat\theta_n^* - \hat\theta_n) - I_n\big]\nabla_\theta F(x,\theta_n^*(x))^T + I_n\big[\nabla_\theta F(x,\theta_n^*(x)) - \nabla_\theta F(x,\theta_0)\big]^T \xrightarrow{P} 0$$

uniformly in the variable x. Also, from (3.8) we get that

$$\sup_{x\in\mathbb{R}} |H_{m_n}(F(x,\hat\theta_n))| \le \sup_{0\le u\le 1} |H_{m_n}(u)| \to 0 \qquad \text{a.s.}$$

Hence, we have (3.19) and the proof of Theorem 3.2 is complete. □

Proof of Theorem 3.3. Consider the random variables defined in the preamble of the section. Since $F_n(x)$ is the empirical distribution function of the independent sample variables $X_1, \dots, X_n$ having common theoretical distribution function $F(x, \theta_0)$, $x \in \mathbb{R}$, the Glivenko–Cantelli theorem implies that

$$S_n = \sup_{x \in \mathbb{R}} \big| F_n(x) - F(x, \theta_0) \big| \to 0 \qquad \text{a.s.} \tag{3.30}$$

Let

$$T_{2,n} = \sup_{x \in \mathbb{R}} \big| B_{m_n}(F_n(x)) - B_{m_n}(F(x, \theta_0)) \big| \,,$$

and consider the sequence $p_n \to 0$ satisfying (3.14) with $A_n = S_n$. Using the fact that the random variables $X_1, X_2, \dots$ are independent from the processes $B_1, B_2, \dots$ we obtain from (3.16) and (3.30) that

$$P\big(|T_{2,n}| \le \varepsilon\big) \ge P\big(w(B_{m_n}, p_n) \le \varepsilon\big) P\big(|S_n| \le p_n\big) \to 1 \,,$$

and hence, $T_{2,n} \to 0$ in probability. Also, introducing the function

$$K_n(x) = \beta_{m_n}(F_n(x)) - B_{m_n}(F(x, \theta_0)) \,, \qquad x \in \mathbb{R} \,,$$

and using (3.8) it follows that

$$\begin{aligned} T_n = \sup_{x \in \mathbb{R}} |K_n(x)| &\le \sup_{x \in \mathbb{R}} \big| \beta_{m_n}(F_n(x)) - B_{m_n}(F_n(x)) \big| \\ &+ \sup_{x \in \mathbb{R}} \big| B_{m_n}(F_n(x)) - B_{m_n}(F(x, \theta_0)) \big| = \sup_{x \in \mathbb{R}} \big| H_{m_n}(F_n(x)) \big| + T_{2,n} \xrightarrow{P} 0 \,. \end{aligned} \tag{3.31}$$

The following statement is the non-parametric bootstrap version of Lemma 3.5, and it says that $m_n^{1/2}(\tilde\theta_n^* - \hat\theta_n)$ has the same asymptotic behavior as the related parametric bootstrapped variable.

Lemma 3.6. *Let $I_n = \int_{\mathbb{R}} l(x, \theta_0)\, dB_{m_n}(F(x, \theta_0))$ be the same as in Lemma 3.5. Then*

$$\big\| m_n^{1/2}(\tilde\theta_n^* - \hat\theta_n) - I_n \big\| \xrightarrow{P} 0 \,.$$

Proof. Let $V^{(k)}(y)$ be the same as in Lemma 3.5, that is, the total variation of the component $l^{(k)}(x, \theta_0)$ on the interval $[-y, y]$, and consider $V(y) = (V^{(1)}(y), \dots, V^{(d)}(y))$, $y > 0$. Similarly to (3.23) one can construct a sequence

$$y_n \to \infty \qquad \text{such that} \qquad T_n \|V(y_n)\| \xrightarrow{P} 0 \,.$$

By assumption (a6) and (3.13) we can write

$$\begin{aligned} m_n^{1/2}(\tilde\theta_n^* - \hat\theta_n) &= m_n^{1/2} \left[\int_{\mathbb{R}} l(x, \theta_0)\, dF_{m_n,n}^*(x) - \int_{\mathbb{R}} l(x, \theta_0)\, dF_n(x) \right] + \varepsilon_{m_n,n}(\theta_0) \\ &= \int_{\mathbb{R}} l(x, \theta_0) d\beta_{n_m}(F_n(x)) + \varepsilon_{m_n,n}(\theta_0) \,. \end{aligned}$$

Let $R_n = \mathbb{R} \setminus (-y_n, y_n]$ and consider the decomposition

$$m_n^{1/2}(\tilde{\theta}_n^* - \hat{\theta}_n) - I_n = \int_{(-y_n,y_n]} l(x,\theta_0)\, dK_n(x) + \int_{R_n} l(x,\theta_0)\, d\beta_{m_n}(F_n(x))$$
$$- \int_{R_n} l(x,\theta_0)\, dB_{m_n}(F(x,\theta_0)) + \varepsilon_{m_n,n}(\theta_0) = I_{1,n} + I_{2,n} + I_{3,n} + \varepsilon_{m_n,n}(\theta_0)\,.$$

Our goal is to prove that all terms converges to 0 in probability. Although the sequence y_n may differ from the sequence y_n of the parametric bootstrap case, one can prove the convergence of $I_{1,n}$ by using the same technique as in (3.24). Also, from the calculation in (3.25) the variance of the k-th component of $I_{3,n}$ is

$$\operatorname{Var} I_{3,n}^{(k)} = \int_{R_n} \big(l^{(k)}(x,\theta_0)\big)^2\, dF(x,\theta_0) - \bigg[\int_{R_n} l^{(k)}(x,\theta_0)\, dF(x,\theta_0)\bigg]^2 \to 0\,,$$

and the convergence of $I_{3,n}$ follows by applying the Chebyshev inequality just the same way as in (3.26). With the computation method of (3.27)-(3.29) one can obtain the conditional mean

$$E\big[I_{2,n}^{(k)} \mid X_1, \ldots, X_n\big] = E\big[I_{2,n}^{(k)} \mid F_n\big] = 0\,,$$

and the conditional variance

$$\operatorname{Var}\big[I_{2,n}^{(k)} \mid X_1, \ldots, X_n\big] = \int_{R_n} \big(l^{(k)}(x,\theta_0)\big)^2\, dF_n(x) - \bigg[\int_{R_n} l^{(k)}(x,\theta_0)\, dF_n(x)\bigg]^2$$
$$= \frac{1}{n}\sum_{i=1}^n \big(l^{(k)}(X_i,\theta_0)\big)^2 \mathbb{1}_{\{X_i \in R_n\}} - \bigg[\frac{1}{n}\sum_{i=1}^n l^{(k)}(X_i,\theta_0)\mathbb{1}_{\{X_i\in R_n\}}\bigg]^2$$
$$= \frac{n-1}{n^2}\sum_{i=1}^n \big(l^{(k)}(X_i,\theta_0)\big)^2 \mathbb{1}_{\{X_i \in R_n\}} - \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^n l^{(k)}(X_i,\theta_0)\mathbb{1}_{\{X_i\in R_n\}} l^{(k)}(X_j,\theta_0)\mathbb{1}_{\{X_j\in R_n\}}\,.$$

Then, $EI_{2,n}^{(k)} = 0$ and using the independence of the variables $X_1, \ldots, X_n$ we have

$$\operatorname{Var} I_{2,n}^{(k)} = E\Big[\operatorname{Var}\big[I_{2,n}^{(k)} \mid X_1, \ldots, X_n\big]\Big] + \operatorname{Var}\Big[E\big[I_{2,n}^{(k)} \mid X_1, \ldots, X_n\big]\Big]$$
$$= \frac{n-1}{n} E\Big[\big(l^{(k)}(X_1,\theta_0)\big)^2 \mathbb{1}_{\{X_1\in R_n\}}\Big] - \frac{n-1}{n} E^2\Big[l^{(k)}(X_1,\theta_0)\mathbb{1}_{\{X_1\in R_n\}}\Big]$$
$$= \frac{n-1}{n}\int_{R_n} \big(l^{(k)}(x,\theta_0)\big)^2\, dF(x,\theta_0) - \frac{n-1}{n}\bigg[\int_{R_n} l^{(k)}(x,\theta_0)\, dF(x,\theta_0)\bigg]^2$$
$$= \frac{n-1}{n}\operatorname{Var} I_{3,n}^{(k)} \le \operatorname{Var} I_{3,n}^{(k)} \to 0\,.$$

Applying the Chebyshev inequality similarly as in (3.26) one can obtain the stochastic convergence of $I_{2,n}$. Finally, $\varepsilon_{m_n,n}(\theta_0)$ also converges to 0 by (a16), which completes the proof of Lemma 3.6. □

Proof of Theorem 3.3, continued. Now we examine the limit behavior of the estimated parameters $\hat{\theta}_n$ and $\hat{\theta}_n^*$. The sample variables $X_1, \dots, X_n$ and the parameter estimation $\hat{\theta}_n$ are the same in the parametric and the non-parametric case, and using only (a4), (a7), (a9), (a14) and the strong law of large numbers we already showed in the proof of Theorem 3.2 that $\|\hat{\theta}_n - \theta_0\| \to 0$ in probability (see (3.17)). Since the stochastic integrals I_n, $n = 1, 2, \dots$, defined in Lemma 3.6 have identical distribution, the term $m_n^{1/2}(\tilde{\theta}_n^* - \hat{\theta}_n)$ is asymptotically normal, which implies that $\|\tilde{\theta}_n^* - \hat{\theta}_n\| \xrightarrow{P} 0$. Working on the event $E_n = \{\hat{\theta}_n \in \Lambda\} \cap \{\tilde{\theta}_n^* \in \Lambda\}$ we will show that

$$P\left(\sup_{x \in \mathbb{R}} \left| \tilde{\alpha}^*_{m_n,n}(x) - \left(\frac{m_n}{n}\right)^{1/2} \hat{\alpha}_n(x) - G_n(x) \right| > \varepsilon \,\middle|\, E_n \right) \to 0 \tag{3.32}$$

holds for arbitrary $\varepsilon > 0$. Since $P(E_n) \to 1$ the convergence in (3.32) implies Theorem 3.3 by (3.15).

If E_n occurs then by assumption (a1) we can take the one-term Taylor expansion with respect to $\hat{\theta}_n$:

$$\begin{aligned}
\tilde{\alpha}^*_{m_n,n}(x) - \left(\frac{m_n}{n}\right)^{1/2} \hat{\alpha}_n(x) &= \beta_{m_n}(F_n(x)) - m_n^{1/2}\big[F(x, \tilde{\theta}_n^*) - F(x, \hat{\theta}_n)\big] \\
&= B_{m_n}(F(x, \theta_0)) + K_n(x) - m_n^{1/2}\big(\tilde{\theta}_n^* - \hat{\theta}_n\big)\nabla_\theta F(x, \tilde{\theta}_n^*(x))^T \\
&= G_{m_n}(x) + L_n(x) - H_{2,n}(x)\,,
\end{aligned}$$

where $\tilde{\theta}_n^*(x)$ lies somewhere between $\hat{\theta}_n$ and $\tilde{\theta}_n^*$ for every $x \in \mathbb{R}$ and $n = 1, 2, \dots$ Then, $\tilde{\theta}_n^*(x) \in \Lambda$ and $\sup_{x \in \mathbb{R}} \|\tilde{\theta}_n^*(x) - \theta_0\| \to 0$ in probability. Using (a2), Lemma 3.6, the stochastic boundedness of I_n and the boundedness of $\nabla_\theta F(x, \theta)$ on $\mathbb{R} \times \Lambda$ we get that

$$H_{2,n}(x) = \big[n^{1/2}\big(\tilde{\theta}_n^* - \hat{\theta}_n\big) - I_n\big]\nabla_\theta F(x, \tilde{\theta}_n^*(x))^T + I_n\big[\nabla_\theta F(x, \tilde{\theta}_n^*(x)) - \nabla_\theta F(x, \theta_0)\big]^T \xrightarrow{P} 0$$

uniformly in x. Also, from (3.31) the sequence $K_n(x)$ converges to 0 uniformly in probability. Hence, we have (3.32), which leads to Theorem 3.3. □

3.4 The bootstrap algorithm

Consider independent and identically distributed observations $X_1, \dots, X_n$ having an unknown distribution function $F(x)$, $x \in \mathbb{R}$, and consider a distribution family

$$\mathcal{F} = \big\{F(x, \theta) : x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^d\big\}$$

endowed with a parameter estimation $\theta_n : \mathbb{R}^n \to \Theta$. In this setup one can test the fit of the sample to the family $\mathcal{F}$, that is, the null-hypotheses $\mathcal{H}_0 : F \in \mathcal{F}$ by applying the test statistics $\psi_n = \psi(\hat{\alpha}_n)$, where $\hat{\alpha}_n$ is the estimated empirical process defined in (3.1) and ψ is a real valued functional on the space $D[-\infty, \infty]$. If ψ is continuous then the weak convergence of $\hat{\alpha}_n$ to the limit process G implies that ψ_n converges to $\varphi = \psi(G)$ in distribution. If, additionally, the variable φ has continuous distribution function F_φ,

then the theoretical quantiles of φ serves as asymptotically correct critical values for ψ_n. That is, for a fixed significance level $0 < \alpha < 1$ the $(1-\alpha)$ quantile

$$c(\alpha) = \inf\big\{x \in \mathbb{R} : F_\varphi(x) \geq 1-\alpha\big\}$$

satisfies the convergence

$$P\big(\psi_n \leq c(\alpha)\big) \to 1-\alpha\,, \qquad n \to \infty\,,$$

and one can reject or accept the null-hypotheses $\mathcal{H}_0$ by using the test statistics ψ_n with the critical value $c(\alpha)$.

As we have specified in the introduction of this chapter, the main difficulty with this method is that the quantile $c(\alpha)$ can not be determined in theoretical way. However, if the distribution family $\mathcal{F}$ and the estimation statistics θ_n satisfies the conditions of Theorems 3.2 and/or 3.3 then one may obtain an asymptotically correct estimation for $c(\alpha)$ by using the parametric and/or the non-parametric bootstrap technique. For example, assume that the approximation

$$\sup_{x\in\mathbb{R}} \big|\hat{\alpha}^*_{m_n,n}(x) - G_{m_n}(x)\big| \xrightarrow{P} 0\,, \qquad n \to \infty\,,$$

holds with the parametric bootstrap estimated empirical process $\hat{\alpha}^*_{m_n,n}$, and consider the related statistics $\psi^{*p}_{m_n,n} = \psi(\hat{\alpha}^*_{m_n,n})$ and $\varphi_n = \psi(G_n)$, $n = 1, 2, \dots$. If the functional ψ is smooth enough then we have a good chance to prove the convergence

$$\big|\psi^{*p}_{m_n,n} - \varphi_n\big| \xrightarrow{P} 0\,, \qquad n \to \infty\,. \tag{3.33}$$

Furthermore, Theorem 3.2 implies that the copies $\varphi_1, \varphi_2, \dots$ are independent from the X_i's, and the common distribution function F_φ of the copies $\varphi_1, \varphi_2, \dots$ is continuous at the point $c(\alpha)$ by assumption. Then, it follows from Theorem 2.7 that the quantile

$$c^{*p}_n(\alpha) = \inf\Big\{x \in \mathbb{R} : P\big(\psi^{*p}_{m_n,n} \leq x \mid X_1, \dots, X_n\big) \geq 1-\alpha\Big\}$$

can be applied as an estimator of $c(\alpha)$, and it is an asymptotically correct critical value for the statistics ψ_n. Also, if the conditions of Theorem 3.3 are satisfied then we may obtain the approximation

$$\big|\psi^{*np}_{m_n,n} - \varphi_n\big| \xrightarrow{P} 0 \qquad \text{with} \qquad \psi^{*np}_{m_n,n} = \psi\bigg(\tilde{\alpha}^*_{m_n,n} - \Big(\frac{m_n}{n}\Big)^{1/2}\hat{\alpha}_n\bigg) \tag{3.34}$$

based on the non-parametric bootstrap estimated empirical process $\tilde{\alpha}^*_{m_n,n}$, and Theorem 2.7 ensures that the quantile

$$c^{*np}_n(\alpha) = \inf\Big\{x \in \mathbb{R} : P\big(\psi^{*np}_{m_n,n} \leq x \mid X_1, \dots, X_n\big) \geq 1-\alpha\Big\}$$

can replace $c(\alpha)$ in the applications.

The bootstrap method can be used to test the fit of the given sample $X_1, \dots, X_n$ to the family $\mathcal{F}$ by using the following algorithm. Depending on which approximation of (3.33) and/or (3.34) hold we can use the parametric and/or the non-parametric version of the method. Note that the technique requires a massive amount of random generations and analytical calculations, so the application of a computer is recommended.

1. Calculate the estimator $\hat{\theta}_n$ based on the observations $X_1, \dots, X_n$.
2. Calculate the test statistics ψ_n.
3. Generate independent parametric or non-parametric bootstrapped observations $X^*_{1,n}, \dots, X^*_{m_n,n}$ having distribution function $F(x, \hat{\theta}_n)$ or $F_n(x)$, respectively.
4. Calculate the estimator $\hat{\theta}^*_n$ or $\tilde{\theta}^*_n$ based on the bootstrapped sample.
5. Calculate the bootstrapped statistics $\psi^{*p}_{m_n,n}$ or $\psi^{*np}_{m_n,n}$.
6. Repeat the steps 3–5 R times, and let $\psi^*_{n,1} \le \dots \le \psi^*_{n,R}$ be the order statistics of the resulting R values of $\psi^{*p}_{m_n,n}$ or $\psi^{*np}_{m_n,n}$.
7. Let $c^*_{n,\alpha}$ be the $(1-\alpha)$ empirical quantile of $\psi^{*p}_{m_n,n}$ or $\psi^{*np}_{m_n,n}$, that is, the $\lceil R(1-\alpha) \rceil$-th largest order statistic, where $\lceil y \rceil = \min\{j \in \mathbb{Z} : y \le j\}$ for $y \in \mathbb{R}$.
8. Reject $\mathcal{H}_0$ if ψ_n is greater than $c^*_{n,\alpha}$.

In Section 3.6 we provide simulation studies for the bootstrap test algorithm using the Kolmogorov–Smirnov type statistics of the processes. To prove the theoretical base of the method we must show that the assumptions of Theorem 2.7 hold with

$$\psi : D[-\infty, \infty] \to \mathbb{R}\,, \qquad \psi(h) = \sup_{x \in \mathbb{R}} |h(x)|\,.$$

Observe that the supremum functional ψ is Lipschitz, because for any elements h_1 and h_2 of $D[-\infty, \infty]$ we have the inequality

$$\big|\psi(h_1) - \psi(h_2)\big| \le \sup_{x \in \mathbb{R}} \big|h_1(x) - h_2(x)\big|\,.$$

Under the assumption that the distribution family $\mathcal{F}$ and the parameter estimation statistics θ_n satisfy the conditions of Theorems 3.2 and/or 3.3, we obtain the desired approximation for the parametric bootstrap statistic

$$|\psi^{*p}_{m_n,n} - \varphi_{m_n}| = \big|\psi(\hat{\alpha}^*_{m_n,n}) - \psi(G_{m_n})\big| \le \sup_{x \in \mathbb{R}} \big|\hat{\alpha}^*_{m_n,n}(x) - G_{m_n}(x)\big| \xrightarrow{P} 0\,,$$

and/or for the non-parametric bootstrap statistic

$$|\psi^{*np}_{m_n,n} - \varphi_{m_n}| \le \sup_{x \in \mathbb{R}} \Big|\tilde{\alpha}^*_{m_n,n}(x) - \Big(\frac{m_n}{n}\Big)^{1/2} \hat{\alpha}_n(x) - G_{m_n}(x)\Big| \xrightarrow{P} 0\,,$$

as $n \to \infty$. It only remains to prove that the common distribution function F_φ of the copies $\varphi_1, \varphi_2, \dots$ is continuous.

Since the process $W(u) - uW(1)$ is a Brownian bridge for any representation $W(u)$, $0 \le u \le 1$, of the standard Wiener process, the process $B(F(x, \theta_0))$, $x \in \mathbb{R}$, has the same distribution as

$$W\big(F(x, \theta_0)\big) - F(x, \theta_0) W(1) = \int_{\mathbb{R}} \Big[\mathbb{1}_{(-\infty, x]}(y) - F(x, \theta_0)\Big]\, dW\big(F(y, \theta_0)\big)\,, \quad x \in \mathbb{R}\,,$$

where $\mathbb{1}$ stands for the indicator function of the interval appearing in the lower index. Also, we have

$$\begin{aligned}\xi &= \int_{\mathbb{R}} l(y,\theta_0)\, dB\big(F(y,\theta_0)\big) \overset{\mathcal{D}}{=} \int_{\mathbb{R}} l(y,\theta_0)\, dW\big(F(y,\theta_0)\big) - W(1)\int_{\mathbb{R}} l(y,\theta_0)\, dF(y,\theta_0) \\ &= \int_{\mathbb{R}} \left[l(y,\theta_0) - \int_{\mathbb{R}} l(z,\theta_0)\, dF(z,\theta_0) \right] dW\big(F(y,\theta_0)\big)\end{aligned}$$

from which the limit process $G(x)$, $x \in \mathbb{R}$, has the same distribution than the process defined by the stochastic integral

$$\int_{\mathbb{R}} \left[\mathbb{1}_{(-\infty,x]}(y) - F(x,\theta_0) + \left[l(y,\theta_0) - \int_{\mathbb{R}} l(z,\theta_0)\, dF(z,\theta_0) \right] \nabla_\theta F(x,\theta_0)^T \right] dW\big(F(y,\theta_0)\big)\,,$$

$x \in \mathbb{R}$. Lemma 2.9 in Section 2.3 implies that $G(x)$, $x \in \mathbb{R}$, is a centered Gaussian process, and one can obtain its covariance function presented in Section 3.2 by using the general covariance function provided by the referred lemma. Since the variable ξ has a d-dimensional normal distribution with mean 0, and $\nabla_\theta F(x,\theta_0)$, $x \in \mathbb{R}$, is a deterministic function, the product $\xi \nabla_\theta F(x,\theta_0)^T$ can be equal to the term $B(F(x,\theta_0))$ at every point $x \in \mathbb{R}$ only with probability 0. That is, with probability 1 the process

$$G(x) = B\big(F(x,\theta_0)\big) - \xi \nabla_\theta F(x,\theta_0)^T\,, \qquad x \in \mathbb{R}\,,$$

is not degenerate at every real x. Observe that we have the inequality

$$\varphi = \sup_{x\in\mathbb{R}} |G(x)| \le \sup_{0\le u\le 1} |B(u)| + |\xi| \sup_{x\in\mathbb{R}} \left|\nabla_\theta F(x,\theta_0)^T\right|, \tag{3.35}$$

where the absolute value of a vector is defined as the vector of the absolute values of the components. Since the Brownian bridge has finite supremum and ξ is normal, the variable φ is almost surely finite. If we additionally assume that $\nabla_\theta F(x,\theta_0)$, $x \in \mathbb{R}$, has càdlàg components, then using Theorem 2.10 we obtain that the distribution function F_φ of

$$\varphi = \psi(G) = \sup_{x\in\mathbb{R}} |G(x)| = \sup\big\{G(x), -G(x) : x \in \mathbb{Q}\big\}$$

is continuous on the interval (s_0, ∞), where

$$s_0 = \inf\big\{x \in \mathbb{R} : F_\varphi(x) > 0\big\} \in [0,\infty)$$

is the left endpoint of the support of F_φ. That is, by the referred theorem the distribution function may have a jump at s_0. However, we strongly believes that the value s_0 is equal to 0, but we could not find a way to prove our conjecture. Since the process G is almost surely not degenerate on the whole real line, if one could show that $s_0 = 0$ then it would follow that the distribution function F_φ is continuous at every $x \in \mathbb{R}$.

To strengthen the conjecture that $s_0 = 0$ take a look at the inequality in (3.35). It is known that for an arbitrary $s > 0$ the supremum of the Brownian bridge B is not

greater than s with a positive probability. Also, the variable ξ has normal distribution with mean 0, and the function $\nabla_\theta F(x, \theta_0)$, $x \in \mathbb{R}$, is bounded by assumption (a3), which imply that the second term on the right side of (3.35) can be arbitrary small, as well. If the terms would be independent than we would obtain that for any $s > 0$ the process G has a supremum smaller than s with a positive probability, and Theorem 2.10 would imply that the distribution function F_φ is continuous at every point on the real line. Unfortunately, we can not apply this idea to solve the present problem, because there is a strong dependence between the Brownian bridge B and the variable ξ. However, similar examples suggest that the value s_0 seems to be equal to 0 for the process G, as well.

3.5 Validity of the assumptions

In Section 3.6 we demonstrate the bootstrap method in simulation studies by testing fit to the Poisson and the normal distribution family. We showed it in the previous section that the Kolmogorov–Smirnov type statistics based on the parametric and the non-parametric bootstrap estimated empirical process can be applied if the underlying distribution family $\mathcal{F}$ and the applied estimation method satisfies the conditions of Theorems 3.2 and 3.3. In this section we check the validity of the assumptions for the Poisson and the normal family with the maximum likelihood estimation.

From the practical point of view, when we test the fit of a given sample to a family $\mathcal{F}$, we can not fix the parameter θ_0, and we must check the conditions for every possible θ_0 in Θ. The validity of (a1)–(a3) must usually be verified by direct calculations. The most frequently used estimation methods satisfy (a4)–(a10), if the family has a finite second moment, and the estimator θ_n has an asymptotically normal distribution, but sometimes we must choose a non-standard parametrization. Assumptions (a11)–(a13) do not seem to be stringent either. The hardest conditions are (a14)–(a16). Burke et al. (1979) showed the stochastic convergence of $\varepsilon_n(\theta_0)$ under some analytical conditions on the distribution family for the maximum likelihood estimation, and Chapter 5 of van der Vaart (1998) investigates the same problem in case of M- and Z-statistics in a much more general framework. For the convergence of the non-parametric bootstrap error term $\varepsilon_{m_n,n}(\theta_0)$ Burke and Gombay (1991) provides similar analytic conditions as Burke et al. (1979) in the non-bootstrap case. Also, Babu and Singh (1984) proved the validity of the assumptions of the non-parametric bootstrap case for empirical quantile type estimators and for L-statistics defined by the forms

$$\sum_{i=1}^{n} X_{i,n} \int_{(i-1)/n}^{i/n} w(t)\, dt \qquad \text{and} \qquad \frac{1}{n} \sum_{i=1}^{n} X_{i,n} w(i/n)\,,$$

where $X_{1,n} \le \cdots \le X_{n,n}$ is the ordered sample based on the observations $X_1, \ldots, X_n$ and $w(t)$, $0 < t \le 1$, is a sufficiently smooth weight function. Unfortunately, general results for the limiting behavior of the bootstrap error terms are not available. In most practical applications the most simple solution is to determine the error terms and to check their convergence by direct considerations.

If we apply the maximum likelihood estimation, then in many distribution families the function

$$l(x,\theta) = \nabla_\theta f(x,\theta) I^{-1}(\theta)\,, \qquad x \in \mathbb{R}\,, \quad \theta \in \Theta\,, \tag{3.36}$$

satisfies assumptions (a4)–(a6), where $f(x,\theta)$ is the density function of $F(x,\theta)$, $x \in \mathbb{R}$, with respect to a σ-finite measure μ on the real line, that is, $dF(x,\theta) = f(x,\theta)d\mu$ for every θ, and $I^{-1}(\theta)$ is the inverse of the Fisher information matrix

$$I(\theta) = \int_{\mathbb{R}} \nabla_\theta f(x,\theta) \nabla_\theta f(x,\theta)^T \, dF(x,\theta)\,, \qquad \theta \in \Theta\,.$$

Although, in many families representations (a4)–(a6) with $l(x,\theta)$ of (3.36) provide the asymptotic normality of the estimators $\hat{\theta}_n$, $\hat{\theta}_n^*$ and $\tilde{\theta}_n^*$, in general we have no rate of convergence, and the validity of assumptions (a14)–(a16) does not follow directly. However, in many cases the convergence of the error terms can be guarantied by using a simple trick.

Observe that the convergence of the error terms are obviously satisfied if the family $\mathcal{F}$ is parametrized by some of its moments and we apply the method of moments to estimate the parameter. For example, if the k-th component of θ is the moment of order α of the underlying distribution $F(x,\theta)$, $x \in \mathbb{R}$, and if we estimate this component with the empirical moment

$$\theta_n^{(k)} = \frac{X_1^\alpha + \cdots + X_n^\alpha}{n}\,,$$

then using a function $l(x,\theta)$ having k-th component $l^{(k)}(x,\theta) = x^\alpha - \theta^{(k)}$, $x \in \mathbb{R}$, $\theta \in \Lambda$, the corresponding components of the error terms in representations (a4)-(a6) vanish. Of course, in general we prefer the maximum likelihood estimation, and in many cases we can not use the method of moments either, because the family is not parametrized only by its moments. However, if the parameter θ is a one-to-one function of some of the moments, then the maximum likelihood estimation of the parameter is equivalent with the estimation of the corresponding moments, and the maximum likelihood provides the same estimation for the moments as the method of moments does. That is, in such cases it is a good idea to choose a new, equivalent parametrization. For example, if we consider the family of negative binomial distributions of order 1 parametrized by the success probability p, then the maximum likelihood method provides the estimator

$$\hat{p}_n = \frac{n}{X_1 + \cdots + X_n + n}\,.$$

At a first glance it is not clear not at all how we can represent $\hat{p}_n$ in formula (a4), and unfortunately, the function defined in (3.36) is provided in a very unfriendly form. However, if we choose an equivalent parametrization, and we parametrize the family with the expected value $\theta = 1/p$ of the distributions, then the maximum likelihood estimator of θ is the sample mean, and the error terms in assumptions (a4)–(a6) are constant 0. Similar tricks can be applied in many distribution families.

3.5.1 The Poisson distribution family

The Poisson family is a parametric collection of distributions $\mathcal{F} = \{\mathrm{Po}(\lambda), \lambda > 0\}$ with probability mass function and distribution function

$$p(k,\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}\,, \quad k = 0, 1, \dots\,, \qquad F(x,\lambda) = \sum_{k=0}^{\lfloor x\rfloor} p(k,\lambda)\,, \quad x \in \mathbb{R}\,.$$

The distribution function can be differentiated with respect to the variable λ, and the derivative is $\nabla_\lambda F(x,\lambda) = 0$ for $x < 0$, while

$$\nabla_\lambda F(x,\lambda) = -e^{-\lambda} + \sum_{k=1}^{\lfloor x\rfloor}\left[\frac{\lambda^{k-1}}{(k-1)!} - \frac{\lambda^k}{k!}\right]e^{-\lambda} = -\frac{\lambda^{\lfloor x\rfloor}}{\lfloor x\rfloor!}e^{-\lambda}\,, \qquad x \geq 0\,.$$

Since $\lim_{x\to\infty}\nabla_\lambda F(x,\lambda) = 0$, the derivative is bounded by a real bound $R(\lambda)$. Consider arbitrary fixed values $\lambda_0 > 0$ and $x \geq 0$, and let $k = \lfloor x\rfloor \geq 0$. If $\lambda \leq 2\lambda_0$, then

$$\begin{aligned}
\left|\nabla_\lambda F(x,\lambda) - \nabla_\lambda F(x,\lambda_0)\right| &\leq \frac{\lambda^k}{k!}\left|e^{-\lambda} - e^{-\lambda_0}\right| + \left|\frac{\lambda^k}{k!} - \frac{\lambda_0^k}{k!}\right|e^{-\lambda_0} \\
&\leq \frac{(2\lambda_0)^k}{k!}\left|e^{-\lambda} - e^{-\lambda_0}\right| + |\lambda - \lambda_0|\frac{k\lambda_*^{k-1}}{k!}e^{-\lambda_0} \\
&\leq \frac{(2\lambda_0)^k}{k!}e^{-\lambda_0}\left|e^{\lambda_0-\lambda} - 1\right| + |\lambda - \lambda_0|\frac{(2\lambda_0)^{k-1}}{(k-1)!}e^{-\lambda_0} \\
&\leq R(2\lambda_0)e^{\lambda_0}\left|e^{\lambda_0-\lambda} - 1\right| + |\lambda - \lambda_0|e^{\lambda_0}R(2\lambda_0)\,,
\end{aligned}$$

with a value λ_* lying between λ and λ_0 provided by the one-term Taylor expansion. Hence, $\nabla_\lambda F(x,\lambda)$ converges to $\nabla_\lambda F(x,\lambda_0)$ uniformly in x as $\lambda \to \lambda_0$, and conditions (a1)–(a3) are satisfied.

Let $Y_1, \dots, Y_n$ be independent sample variables having distribution $\mathrm{Po}(\lambda)$ with an arbitrary λ. The maximum likelihood estimation of the parameter is the sample mean

$$\lambda_n = \frac{Y_1 + \dots + Y_n}{n}\,.$$

Since λ is the expected value of the Poisson distribution, the method of moments provides the same estimation. Based on our remark at the beginning of this section on the method of moments with

$$l(x,\lambda) = x - \lambda\,, \qquad x \in \mathbb{R}\,, \quad \theta \in \Theta\,,$$

the error terms vanish in representations (a4)–(a6), that is, we have (a14)–(a16). Note that the formula in (3.36) provides the same function $l(x,\theta)$. It can be easily checked that we have (a7)–(a9) and (a12)–(a13). Finally, the identity

$$M(\lambda) = El(Y_1,\lambda)^2 = E(Y_1 - \lambda)^2 = \mathrm{Var}\,(Y_1) = \lambda$$

implies (a11)–(a12), and all assumptions are satisfied.

3.5.2 The normal distribution family

The normal distribution family $\mathcal{F} = \{\mathrm{N}(\theta), \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)\}$ is defined by the biparametric distribution function

$$F\big(x, \mu, \sigma^2\big) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-t^2/2}\, dt = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \phi(t)\, dt\,, \quad x \in \mathbb{R}\,,$$

where

$$\phi(t) = \frac{1}{(2\pi)^{1/2}} e^{-t^2/2}\,, \qquad t \in \mathbb{R}\,,$$

is the density function of the standard normal distribution $\mathrm{N}(0, 1)$. The vector of the derivatives with respect to the variables μ and σ^2 exists and it is given in the form

$$\nabla_{(\mu,\sigma^2)} F\big(x, \mu, \sigma^2\big) = \bigg(-\frac{1}{\sigma}\phi\Big(\frac{x-\mu}{\sigma}\Big), -\frac{x-\mu}{2\sigma^3}\phi\Big(\frac{x-\mu}{\sigma}\Big)\bigg)\,.$$

Since the functions $\phi(t)$ and $|t|\phi(t)$, $t \in \mathbb{R}$, are bounded by $1/(2\pi)^{1/2}$ and $1/(2\pi e)^{1/2}$, respectively, we have

$$\big\|\nabla_{(\mu,\sigma^2)} F(x, \mu, \sigma^2)\big\|^2 \le \bigg(\frac{1}{(2\pi)^{1/2}\sigma}\bigg)^2 + \bigg(\frac{1}{(8\pi e)^{1/2}\sigma^2}\bigg)^2\,,$$

and the validity of assumption (a3) follows. For condition (a2) consider an arbitrary fixed parameter $\theta_0 = (\mu_0, \sigma_0^2) \in \Theta$ and a value $\varepsilon > 0$. Observe that there exists a real $T > |\mu_0|$ such that $\|\nabla_{(\mu_0,\sigma_0^2)} F(x, \mu, \sigma^2)\|$ is increasing on $(-\infty, -T]$ and decreasing on $[T, \infty)$, and also, the norm is smaller than $\varepsilon/2$ on these intervals. Since the derivative is a continuous function of the parameters, we have

$$\big\|\nabla_{(\mu,\sigma^2)} F(-T, \mu, \sigma^2)\big\| \le \varepsilon/2 \qquad \text{and} \qquad \big\|\nabla_{(\mu,\sigma^2)} F(T, \mu, \sigma^2)\big\| \le \varepsilon/2$$

if (μ, σ^2) lies close enough to (μ_0, σ_0^2), and also, we can assume that $\|\nabla_{(\mu,\sigma^2)} F(x, \mu, \sigma^2)\|$ is increasing on the interval $(-\infty, -T]$ and decreasing on $[T, \infty)$. Hence, we have

$$\begin{aligned}
&\sup_{|x|\ge T} \big\|\nabla_{(\mu,\sigma^2)} F(x, \mu, \sigma^2) - \nabla_{(\mu,\sigma^2)} F(x, \mu_0, \sigma_0^2)\big\| \\
&\le \sup_{|x|\ge T} \big\|\nabla_{(\mu,\sigma^2)} F(x, \mu, \sigma^2)\big\| + \sup_{|x|\ge T} \big\|\nabla_{(\mu,\sigma^2)} F(x, \mu_0, \sigma_0^2)\big\| < \varepsilon\,,
\end{aligned}$$

for every (μ, σ^2) lying in a suitably small neighborhood of (μ_0, σ_0^2). That is, we obtain the convergence of the derivative $\nabla_{(\mu,\sigma^2)} F(x, \mu, \sigma^2)$ to $\nabla_{(\mu,\sigma^2)} F(x, \mu_0, \sigma_0^2)$ uniformly in x on the set $\mathbb{R} \setminus [-T, T]$ as (μ, σ^2) goes to (μ_0, σ_0^2). Since the derivative is continuous in variable x in case of any parameter (μ, σ^2), the function $\nabla_{(\mu,\sigma^2)} F(x, \mu, \sigma^2)$ converges uniformly to $\nabla_{(\mu,\sigma^2)} F(x, \mu_0, \sigma_0^2)$ on the interval $[-T, T]$. This implies and we have uniform convergence on the whole real line, and assumption (a2) is satisfied.

For the remaining conditions note that the maximum likelihood estimators of the mean μ and the variance σ^2 based on an arbitrary sample $Y_1, \dots, Y_n$ are

$$\mu_n = \frac{\sum_{i=1}^n Y_i}{n} \qquad \text{and} \qquad {\sigma^2}_n = \frac{\sum_{i=1}^n (Y_i - \mu_n)^2}{n}\,. \tag{3.37}$$

The bootstrap estimators $\hat{\mu}_n^*$ and $\tilde{\mu}_n^*$ are the empirical means of the parametric and the non-parametric bootstrapped variables, respectively, and ${\widehat{\sigma^2}}_n^*$ and ${\widetilde{\sigma^2}}_n^*$ are the empirical variances based on the corresponding bootstrapped samples. It can be easily seen that representations (a4)–(a6) hold by applying the vector valued function

$$l\big(x, \mu, \sigma^2\big) = \Big(x - \mu, (x-\mu)^2 - \sigma^2\Big)\,, \qquad x \in \mathbb{R}\,, \mu \in \mathbb{R}\,, \sigma^2 > 0\,.$$

We note that formula (3.36) provides the same function $l(x, \mu, \sigma^2)$. If the X_i's have a common distribution $\mathrm{N}(\mu_0, \sigma_0^2)$ then the non-bootstrapped error term

$$\varepsilon_n\big(\mu_0, \sigma_0^2\big) = \varepsilon_n\big(X_1, \dots, X_n, \mu_0, \sigma_0^2\big) = \Big(0, -n^{1/2}(\hat{\mu}_n - \mu_0)^2\Big) \xrightarrow{P} 0$$

by the central limit theorem. Using the central limit theorem on the triangular array of the bootstrapped variables one can prove stochastic convergence for the parametric bootstrapped error term

$$\varepsilon_{m_n}\big(\hat{\mu}_n, \widehat{\sigma^2}_n\big) = \varepsilon_{m_n}\big(X_{1,n}^*, \dots, X_{m_n,n}^*, \hat{\mu}_n, \widehat{\sigma^2}_n\big) = \Big(0, -n^{1/2}(\hat{\mu}_n^* - \hat{\mu}_n)^2\Big)\,.$$

The representation of $\hat{\theta}_n - \theta_0$ in (a4) and the identity $\hat{\theta}_n^* - \hat{\theta}_n = (\tilde{\theta}_n^* - \theta_0) - (\hat{\theta}_n - \theta_0)$ lead to the non-parametric bootstrapped error term

$$\begin{aligned}\varepsilon_{m_n,n}\big(\mu_0, \sigma_0^2\big) &= \Big(0, -n^{1/2}(\tilde{\mu}_n^* - \mu_0)^2\Big) - \varepsilon_n\big(\mu_0, \sigma_0^2\big) \\ &= \Big(0, -n^{1/2}(\tilde{\mu}_n^* - \hat{\mu}_n)(\tilde{\mu}_n^* + \hat{\mu}_n - 2\mu_0)\Big)\,.\end{aligned}$$

From the bootstrap central limit theorem of Bickel and Freedman (1981) we get the convergence of the conditional distribution

$$\mathcal{L}\Big(n^{1/2}(\tilde{\mu}_n^* - \hat{\mu}_n) \mid X_1, \dots, X_n\Big) \xrightarrow{\mathcal{D}} N(0,1)\,, \qquad n \to \infty\,,$$

and it follows that $|\tilde{\mu}_n^* - \hat{\mu}_n|$ converges conditionally to 0 in stochastic sense. Since the condition can be omitted by Lemma 1.2 of Csörgő and Rosalsky (2003), and $\hat{\mu}_n$ goes to μ_0 by the law of large numbers, we obtain that $\tilde{\mu}_n^* + \hat{\mu}_n - 2\mu_0$ converges to 0 in probability. Hence, assumption (a16) is satisfied.

It is obvious that the function $l(x, \mu, \sigma^2)$ satisfies assumptions (a12)–(a13). If the variable Y has distribution $\mathrm{N}(\mu, \sigma^2)$ then we obtain the mean $El(Y, \mu, \sigma^2) = 0$ and the covariance matrix

$$M(\mu, \sigma^2) = \begin{bmatrix} E(Y-\mu)^2 & E(Y-\mu)^3 \\ E(Y-\mu)^3 & E\big[(Y-\mu)^2 - \sigma^2\big]^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix},$$

which satisfy assumptions (a7)–(a11).

We note that the maximum likelihood estimation (3.37) and the above method for proving conditions (a4)–(a16) can be applied not just in the normal family but in many other cases, when the distribution family is parametrized by the mean and the variance of the elements and the members have finite fourth moment. In general, the natural parametrization of biparametric distribution families is not based on the moments, but in most cases the family can be reparameterized. Since the difference between the empirical variance $\widehat{\sigma^2}_n$ and its non-biased counterpart $n\widehat{\sigma^2}_n/(n-1)$ converges to 0 with asymptotic rate $\mathcal{O}(1/n)$, one can also use the latter to estimate σ^2.

3.6 Simulation studies

To demonstrate the non-parametric and the parametric bootstrap method in practical applications, we made simulation studies in the Poisson and the normal distribution family. As a parameter estimation method we used the maximum likelihood estimation, which satisfies the conditions of Theorems 3.2 and 3.3 by our investigation in Section 3.5. We applied the Kolmogorov–Smirnov type statistics based on the parameter estimated empirical processes, that is, we applied the algorithm presented in Section 3.4 with the non-bootstrapped, the non-parametric bootstrap and the parametric bootstrap statistics

$$\psi_n = \sup_{x\in\mathbb{R}} \left|\hat{\alpha}_n(x)\right|, \quad \psi^{*np}_{m_n,n} = \sup_{x\in\mathbb{R}} \left|\tilde{\alpha}^*_{m_n,n}(x) - \hat{\alpha}_n(x)\right| \quad \text{and} \quad \psi^{*p}_{m_n,n} = \sup_{x\in\mathbb{R}} \left|\hat{\alpha}^*_{m_n,n}(x)\right|.$$

We considered the sample size $n = 50$ and the bootstrapped sample size $m_n = n$, and we made $R = 100$ bootstrap replications.

3.6.1 Testing for the Poisson distribution

In our first simulation, we considered identically distributed random samples from the negative binomial family $\{\mathrm{NB}(r,p), r = 1,2,\ldots, 0 < p \le 1\}$ having probability mass function

$$p(k,r,p) = \binom{k+r-1}{r-1}(1-p)^k p^r\,, \qquad k = 0,1,\ldots\,,$$

and we tested the fit for the Poisson distribution using both the non-parametric and the parametric bootstrap method. We made the tests for every odd r between 1 and 41, and for each r we varied the parameter p between 0.2 and 0.99 by stepping 0.05. To investigate the empirical power of the procedures, we generated variables and executed the above algorithm $N = 1000$ times for each pair (r,p). To compare the efficiency of the bootstrap technique to other methods see Gürtler and Henze (2000), where large scale simulation studies are presented for the Poisson distribution based on various true distributions including members from the negative binomial family.

The bivariate empirical power function of the statistics $\psi^{*np}_{m_n,n}$ and $\psi^{*p}_{m_n,n}$ for significance level $\alpha = 0.05$ can be seen in Figure 3.1, and Figure 3.2 presents the related contour diagrams. In the contour charts the interval $[0,1]$ is divided into eleven equal

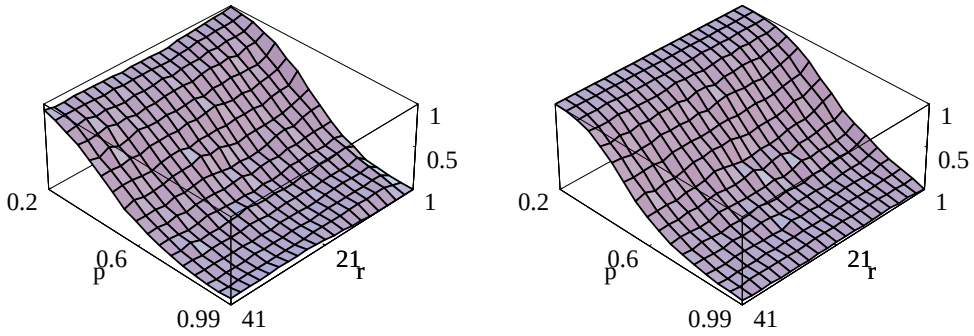

Figure 3.1: The empirical power of $\psi^{*np}_{m_n,n}$ and $\psi^{*p}_{m_n,n}$ for the Poisson distribution against negative binomial samples for significance level $\alpha = 0.05$.

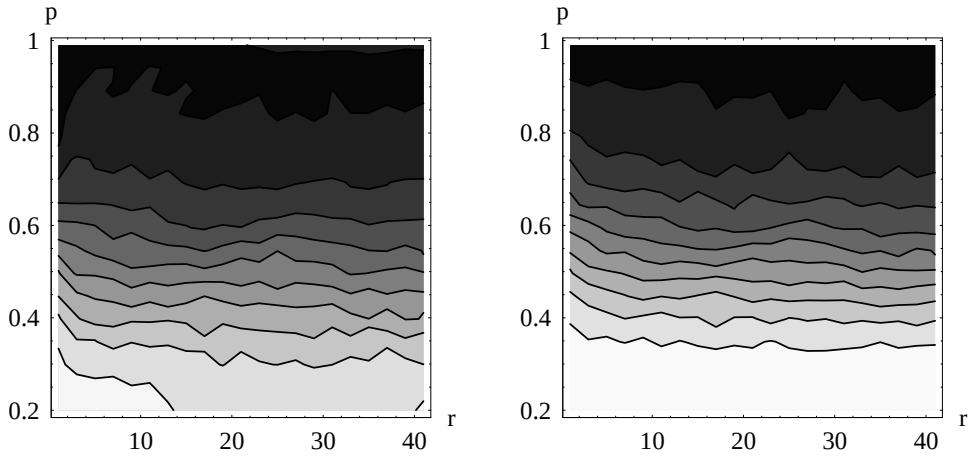

Figure 3.2: The contour diagrams of the functions plotted in Figure 3.1.

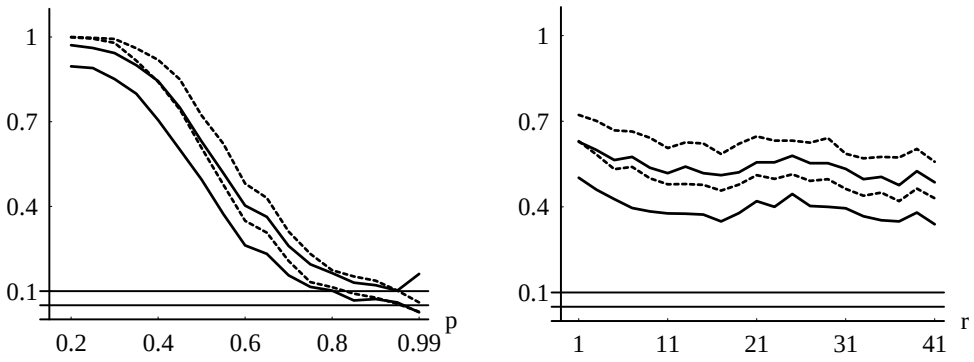

Figure 3.3: The empirical power of the tests for fixed $r = 20$ (on the left) and for fixed $p = 0.5$ (on the right).

subinterval, and the eleven tones of color gray indicates that the power is in a given section. The deeper tone the lower power, that is, in the black area the power is below 0.09, and in the white region the power is above 0.91. In Figure 3.3 we plotted the power for fixed order $r = 15$ (on the left) and for fixed $p = 0.5$ (on the right). The continuous and the dashed curves represent the non-parametric and the parametric power, respectively, and in both cases the upper curve belongs to significance level $\alpha = 0.1$ and the lower curve belongs to $\alpha = 0.05$. We found that the empirical power depends mostly on the parameter p. Since $\mathrm{NB}(r, 0)$ is the degenerate distribution at 0, and it is an element of the Poisson family, we may expect that the empirical power is close to the significance level α if the parameter p is close to 1 for every r. This idea is valid for large r's, but does not work if r is small. We can observe the latter fact in the left side diagram of Figure 3.3 at the point $p = 0.99$, and the author cannot explain it. The power grows if we lower p, and reaches 0.9 when $p = 0.2$ and $\alpha = 0.05$ for every investigated r. This is because for small a p the variance of the negative binomial distribution is much greater than the mean, and hence, the negative binomial samples are more distinguishable from the Poisson ones. While the empirical power of the non-parametric and the parametric bootstrap method are close in case of large p's, the parametric version has greater power for small values of p. In Figure 3.2 the black parts approximately overlap each other, but the white region is much wider in the parametric case. For a fixed p the power is slowly decreasing as r grows. Also, if $r \to \infty$ and $p = p_r \to 1$ such that $r(1-p) \to \lambda > 0$, then the distribution $\mathrm{NB}(r, p)$ converges to $\mathrm{Po}(\lambda)$ by Bartko (1966). We represented this fact in Figure 3.4, where we plotted the power of the tests based on negative binomial samples having parameters r and $p = 1 - \lambda/r$, (for $r = \lambda$ we chose $p = 0.2$.) On the left side $\lambda = 1$, and $\lambda = 5$ on the right, and the significance levels are 0.1 and 0.05, again. We can see that the empirical power goes to the theoretical significance level as p grows in each case, but for larger λ the convergence is slower.

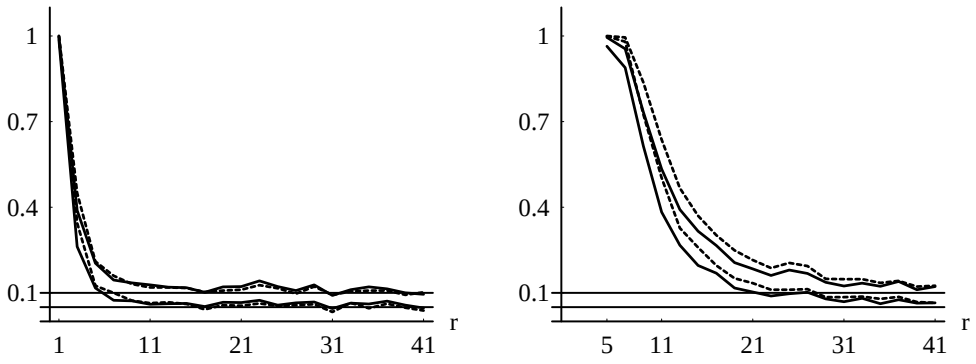

Figure 3.4: The empirical power of the tests against negative binomial samples of order r and parameter $p = 1 - 1/r$ (on the left) and $p = 1 - 5/r$ (on the right).

3.6.2 Testing for the normal distribution

In our second Monte Carlo simulation we tested normality by generating location and scale contaminated normal samples. We say that a variable has location contaminated normal distribution having parameters m and p, if the observation is randomly selected from the standard normal distribution with probability $1-p$ and from $\mathrm{N}(m,1)$ with probability p. The scale contaminated normal distribution with parameters σ^2 and p can be defined similarly as the mixture of $\mathrm{N}(0,1)$ and $\mathrm{N}(0,\sigma^2)$. For comparative simulation studies for the normal distribution see Gan and Koehler (1990).

First, we generated location contaminated normal samples. We raised the parameter m from 0 to 6 by stepping 0.5 and varied p between 0 and 1 by stepping 0.05. We repeated the tests $N = 1000$ times for each (m,p). The power function of $\psi^{*np}_{m_n,n}$ and $\psi^{*p}_{m_n,n}$ for significance level $\alpha = 0.05$ can be found in Figure 3.5, and Figure 3.6 shows the corresponding contour diagrams. Also, in Figure 3.7 we plotted the power for fixed $p = 0.5$ (on the left) and for fixed $m = 4$ (on the right). Since the true distribution of the generated samples is a mixture of two normals having equal variances, the power function is symmetric for $p = 0.5$. Also, if $p = 0$ or $p = 1$ or $m = 0$, then the generated variables have normal distribution, and the empirical power is close to the significance level. We can see that the power is very low, it is under 0.2 if $m \leq 2$, and then, it rises very steeply on the interval $[2,4]$ nearing to 1. Interestingly, in the non-parametric bootstrap case the power function is approximately constant in p if $0.2 \leq p \leq 0.8$, and what is more, in the parametric case the test has unexpectedly low power for $p = 0.5$. The parametric bootstrap test has higher power for small and large p's, but it is weaker than the non-parametric method if p is near to 0.5.

Finally, we generated scale contaminated normal samples. We scaled the parameter σ logarithmically, namely, we made our tests for $\sigma = 1.5^s$, where we raised s from -10 to 10. Again, we varied p between 0 and 1 by stepping 0.05, and we made $N = 1000$ repetitions. The bivariate empirical power function for significance level $\alpha = 0.05$ can be seen in Figures 3.8 and 3.9, and Figure 3.10 shows the power for fixed $p = 0.2$ (on the left) and for fixed $s = 5$ (on the right) for significance levels $\alpha = 0.05$ and $\alpha = 0.1$. Since for $p = 0$, $p = 1$ and $s = 0$ the true distribution of the generated sample is normal, the empirical power is approximately equal to the significance level. We found that excluding these cases the parametric bootstrap has higher power than the non-parametric method.

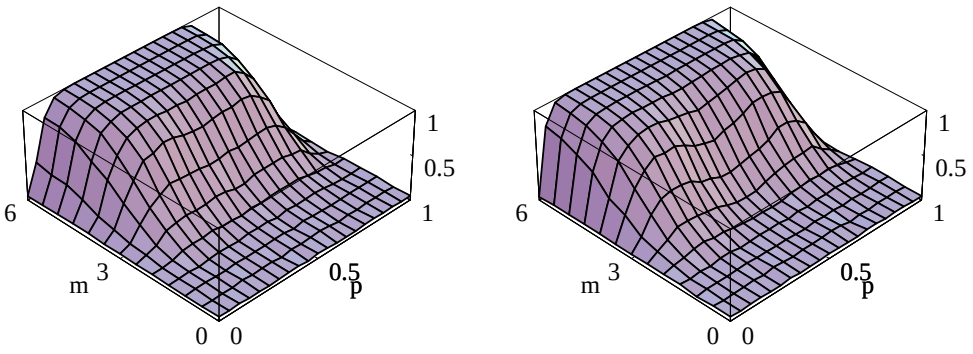

Figure 3.5: The empirical power of $\psi^{*np}_{m_n,n}$ and $\psi^{*p}_{m_n,n}$ for the normal distribution against location contaminated samples for significance level $\alpha = 0.05$.

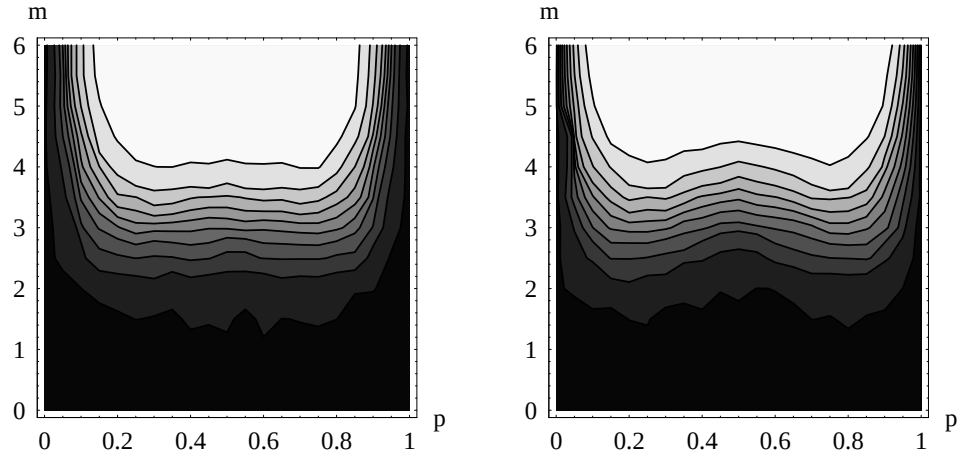

Figure 3.6: The contour diagrams of the functions plotted in Figure 3.5.

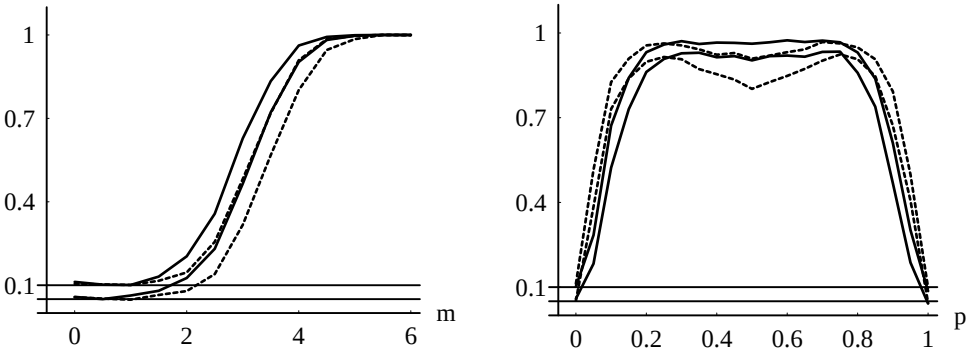

Figure 3.7: The empirical power of the tests for fixed $m = 4$ (on the left) and for fixed $p = 0.5$ (on the right).

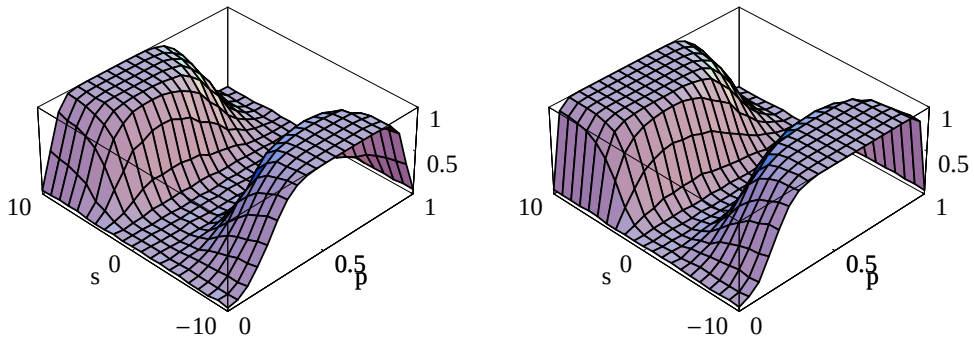

Figure 3.8: The empirical power of $\psi^{*np}_{m_n,n}$ and $\psi^{*p}_{m_n,n}$ for the normal distribution against scale contaminated samples for significance level $\alpha = 0.05$.

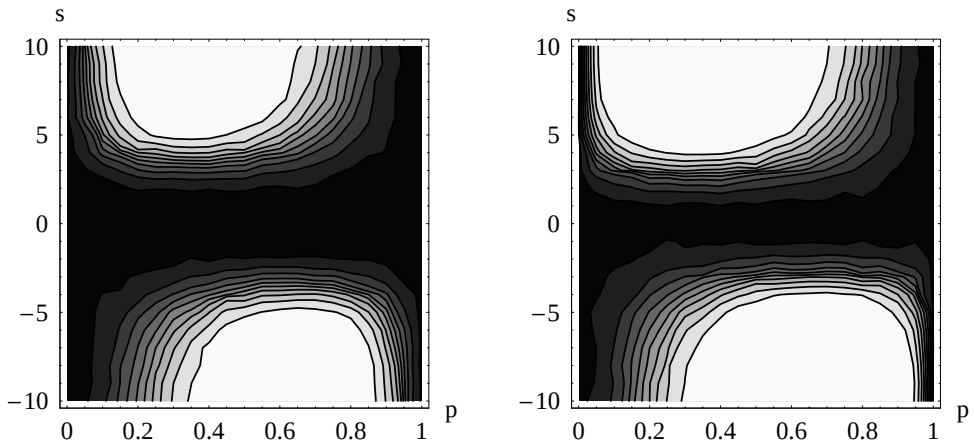

Figure 3.9: The contour diagrams of the functions plotted in Figure 3.8.

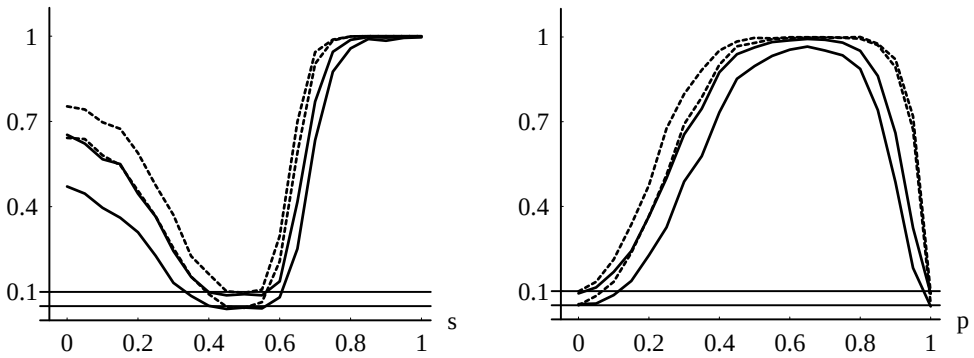

Figure 3.10: The empirical power of the tests for fixed $s = 5$ (on the left) and for fixed $p = 0.2$ (on the right).

Chapter 4

Empirical probability generating processes

4.1 Introduction and preliminary results

Let $X, X_1, X_2, \ldots$ be a sequence of independent and identically distributed nonnegative valued random variables having distribution function $F(x)$, $x \in \mathbb{R}$. Let

$$g(t) = Et^X = \int_{\mathbb{R}} t^x dF(x) \qquad \text{and} \qquad g_n(t) = \frac{1}{n}\sum_{j=1}^{n} t^{X_j}\,, \qquad 0 \le t \le 1\,,$$

be the common probability generating function and its empirical counterpart based on the first n observations. Throughout this chapter the symbol 0^0 is interpreted as 1, because we will need the continuity of the function t^x in variable x. Then the empirical probability generating process can be defined by

$$\gamma_n(t) = n^{1/2}\big[g_n(t) - g(t)\big]\,, \qquad 0 \le t \le 1\,.$$

In the last two decades several authors examined the process γ_n and its parametric versions in the case when the variable X has only non-negative integer values. The papers by Kemp and Kemp (1988), Kocherlakota and Kocherlakota (1986) and Nakamura and Pérez-Abreu (1993a,b) were overviewed by Dowling and Nakamura (1997), while Gürtler and Henze (2000) basically summarize the contents of Baringhaus, Gürtler and Henze (2000), Baringhaus and Henze (1992) and Rueda, Pérez-Abreu and O'Reilly (1991). The idea of the application of generating functions to solve various statistical problems is not unusual, similar transformed processes based on empirical characteristic and moment generating functions are well-known. (For example, see Csörgő (1981) and Csörgő, Csörgő, Horváth and Mason (1986).) In each case the theoretical basis of the method is the fact, that under appropriate conditions the transformed processes converge in distribution in some function space. In the case of the empirical probability generating process, Csörgő and Mason (1989) and Marques and Pérez-Abreu (1989) independently and by different methods proved that if X has finite variance, then γ_n

converges in distribution in $C[0,1]$, the complete and separable metric space of all continuous functions on the interval $[0,1]$ with respect to the supremum distance. The limiting process is

$$\int_{\mathbb{R}} t^x \, dB\big(F(x)\big)\,, \qquad 0 \le t \le 1\,,$$

where $B(u)$, $0 \le u \le 1$, is the Brownian bridge defined in Section 2.1. Remarkably, Proposition 3.1 of Rémillard and Theodorescu (2000) states that no tail condition on X is needed, and the empirical probability generating process converges in distribution in $C[0,1]$ for every non-negative integer valued random variable X. This statement is particularly striking in comparison to the asymptotic behavior of the empirical characteristic function process, as described by Csörgő (1981). Unfortunately, there is an oversight in the proof of Rémillard and Theodorescu (2000). In Section 4.5 we will show that their basic idea is nevertheless good, their proof can be corrected, and so their remarkable result is in fact true.

The aim of this chapter is to present a general approach to convergence problems for probability generating functions and processes and their derivatives, for the parameter estimated version of the empirical probability generating process, and for their bootstrapped analogues. Our results are general in the other sense, as well, that they hold not only for an integer valued variable, but for an arbitrary non-negative valued variable X. In Section 4.2 we investigate the existence of the generalized version of the empirical probability generating process, and provide some useful inequalities for it, which practically trivialize all the convergence problems by reducing them to convergence problems for the corresponding empirical processes. This way it also becomes possible to transfer rate of convergence and strong approximation results to empirical probability generating processes. The applications of the general results can be found in the later sections.

4.2 General results

In this section we investigate processes defined by the form $I(t) = \int_{\mathbb{R}} t^x dK(x)$, where $K(x)$, $x \in \mathbb{R}$, is some empirical type process (empirical or theoretical distribution function, empirical bootstrapped and/or parameter estimated process) based on the sample variables $X_1, \ldots, X_n$. Because in our case the sample comes from a distribution having only non-negative values, the related empirical type processes vanish on the negative half-line $(-\infty, 0)$. Also, these processes are càdlàg, that is, with probability 1 the trajectories are right-continuous and have left-side limit everywhere on the real line. Hence, we can assume that K satisfies all of these criterions.

Throughout in the section we assume that K can be represented by the form

$$K(x) = M(x) + A(x)\,, \qquad x \in \mathbb{R}\,, \tag{4.1}$$

where $M(x)$, $x \in \mathbb{R}$, is a zero mean locally square integrable martingale adapted to a filtration $\mathcal{F} = \{\mathcal{F}_x : x \ge 0\}$ on the underlying probability space, and $A(x)$, $x \in \mathbb{R}$,

is a deterministic or random function which is of bounded variation on every finite interval. Furthermore, we will assume that both M and A are càdlàg on the real line and vanish on $(-\infty, 0)$ with probability 1. By letting $A^+(x)$ and $A^-(x)$ be the positive and the negative variation of A on the interval $(-\infty, x]$, $x \in \mathbb{R}$, respectively, we also have $A = A^+ - A^-$. From the assumptions on A the processes A^+ and A^- are constant 0 on the negative half-line, and they are non-decreasing and càdlàg on the real line.

In Section 2.3 we have already investigated locally square integrable martingales, and we introduced some of their basic properties. Since the choice of the filtration $\mathcal{F}$ is of no significance during the applications of the later sections, we can choose a very comfortable one, we will work with the augmentation of the generated filtration

$$\mathcal{F}_x^0 = \big\{ M(y) : 0 \le y \le x \big\}\,, \qquad x \ge 0\,.$$

We recall that the process $\langle M \rangle_x$, $x \in \mathbb{R}$, is the quadratic characteristic of the local martingale M. Since M is càdlàg on the real line and vanishes on the negative half-line, the process $\langle M \rangle_x$ and its pointwise mean function $E\langle M \rangle_x$, $x \in \mathbb{R}$, also have these properties, and additionally, they are non-decreasing on the positive half-line.

If $f(x)$, $x \in \mathbb{R}$, is some locally bounded deterministic function, that is, it is bounded on every finite interval, then we define its integral on the real line with respect to the process K by the formula

$$\int_{\mathbb{R}} f(x)\, dK(x) = \int_{\mathbb{R}} f(x)\, dM(x) + \int_{\mathbb{R}} f(x)\, dA(x)\,,$$

and the integral is well-defined if both terms of the sum exist. The first term on the right side is a stochastic integral, which was defined in Section 2.3, and in Lemma 2.8 we showed that it is well-defined if and only if

$$\|f\|_{\mathbb{R}}^2 = E \int_{\mathbb{R}} f^2(x)\, d\langle M \rangle_x < \infty\,.$$

By technical reasons we consider the last integral in Lebesgue–Stieltjes sense, that is, the integral exists if and only if f can be integrated with respect to both A^+ and A^-. However, we will see that in our applications the integral is also well-defined in Riemann–Stieltjes sense.

Our base goal is to investigate of the generalized probability generating process

$$I(t) = \int_{\mathbb{R}} t^x\, dK(x)\,,$$

and its "differentiated" versions

$$I_r(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dK(x)\,, \qquad r = 0, 1, \ldots$$

Observe that we have $I(t) = I_0(t)$. As we have already noted earlier, we define the expression 0^0 by 1, just because the continuity of the function t^x in variable x is more

crucial than the continuity in t. It can be easily seen that the integral $I(0) = K(0)$ is finite, and $I_r(0)$ is not defined if $r = 1, 2, \ldots$, but it is not a trivial question at all that the integrals $I(t)$ and $I_r(t)$ exist or do not for any other t. Firstly, we investigate this question. For this we need the following definition and a technical lemma.

Definition. Consider a real valued random or deterministic function $L(x)$, $x \in \mathbb{R}$, which is càdlàg on the real line and vanishes on $(-\infty, 0)$. Let $E(L)$ be the set of those outcomes $\omega \in \Omega$ for which there exists a real constant $\delta = \delta(\omega, L) > 0$ such that the trajectory $L(x) = L(x, \omega)$, $x \in \mathbb{R}$, satisfies

$$\sup_{m\in\mathbb{Z}^+} \big|L\big(m\delta\big) - L\big((m-1)\delta\big)\big| < \infty \quad \text{or} \quad \limsup_{m\to\infty} \frac{\big|L\big((m+1)\delta\big) - L\big(m\delta\big)\big|}{\big|L\big(m\delta\big) - L\big((m-1)\delta\big)\big|} \le 1\,. \tag{4.2}$$

Also, let $E_1(L) \subseteq E(L)$ denote the set of those outcomes ω for which (4.2) holds with $\delta = 1$. We say that L has bounded or slowly growing increments if $E(L)$ has a sub-event having probability 1, and L has bounded or slowly growing increments on the unit intervals, if the same holds for $E_1(L)$. Note that if L is bounded on the positive half-line almost surely then it has bounded increments on the unit intervals.

Proposition 4.1. *Consider a random or deterministic function $L(x)$, $x \in \mathbb{R}$, which is càdlàg on the real line and vanishes on the negative half-line. Then, on the set $E(L)$ with $\delta = \delta(\omega, L)$ defined in the previous paragraph the series*

$$S_{r,\delta,L}(t) = \sum_{m=1}^{\infty} m^r t^{m-1} \big[L\big(m\delta\big) - L\big((m-1)\delta\big)\big]$$

converges absolutely for every $-1 < t < 1$. As a consequence, on the subset $E_1(L)$ the series $S_{r,1,L}(t)$ is also absolutely convergent with any $-1 < t < 1$.

Proof. Fix an arbitrary outcome $\omega \in E(L)$ and a real value $-1 < t < 1$, and consider the deterministic trajectory $L(x) = L(x, \omega)$, $x \in \mathbb{R}$. The increments of the function L are either bounded and we find that

$$\sum_{m=1}^{\infty} \Big|m^r t^{m-1} \big[L\big(m\delta\big) - L\big((m-1)\delta\big)\big]\Big| \le \sup_{m\in\mathbb{Z}^+} \big|L\big(m\delta\big) - L\big((m-1)\delta\big)\big| \sum_{m=1}^{\infty} m^r |t|^{m-1} < \infty\,;$$

or the increments grow slowly, from which we have

$$\limsup_{m\to\infty} \left[\Big(\frac{m+1}{m}\Big)^r \frac{|t|^m}{|t|^{m-1}} \frac{\big|L\big((m+1)\delta\big) - L\big(m\delta\big)\big|}{\big|L\big(m\delta\big) - L\big((m-1)\delta\big)\big|}\right] \le |t| < 1\,,$$

and using the ratio test we can obtain the absolute convergence of $S_{r,\delta,L}(t)$. □

Our next goal is to provide a sufficient condition under which the integral $I_r(t)$ exists. For any fixed $0 \le t \le 1$ and $r = 0, 1, \dots$ consider the decomposition

$$I_r(t) = I_r^M(t) + I_r^A(t) = I_r^M(t) + I_r^+(t) - I_r^-(t)\,,$$

where $I_r^M(t)$, $I_r^A(t)$, $I_r^+(t)$ and $I_r^-(t)$ are the integrals of $x(x-1)\cdots(x-r+1)t^{x-r}$ with respect to $M(x)$, $A(x)$, $A^+(x)$ and $A^-(x)$, $x \in \mathbb{R}$, respectively. Also, the r-th moment of the process K can be considered by the form

$$\int_{\mathbb{R}} x^r\, dK(x) = \int_{\mathbb{R}} x^r\, dM(x) + \int_{\mathbb{R}} x^r\, dA^+(x) - \int_{\mathbb{R}} x^r\, dA^-(x)\,.$$

Then, the variable $I_r(t)$ is finite if and only if $I_r^M(t)$, $I_r^+(t)$ and $I_r^-(t)$ are finite, and the integral $\int_{\mathbb{R}} x^r\, dK(x)$ is well-defined if and only if all terms in its sum representation exist. In the next two propositions we investigate the processes I_r^A and I_r^M.

Proposition 4.2. *Assume that the processes $A^+(x)$ and $A^-(x)$, $x \in \mathbb{R}$, have bounded or slowly growing increments. Then, the integral $I_r^A(t)$ is a well-defined and can be differentiated at any point t of the interval $(0,1)$, and it has derivative $I_{r+1}^A(t)$. Also, the integral $I_0^A(t)$ is finite and right-continuous at $t = 0$. Furthermore, $I_r^A(1)$ exists if and only if $A(x)$, $x \in \mathbb{R}$, has finite r-th moment $\int_{\mathbb{R}} x^r dA(x)$. In this case $I_{r-1}(t)$ have left-side derivative $I_r^A(1)$ at $t = 1$, and $I_r^A(t)$ is left-continuous at this point. (All statements are understood with probability 1.)*

Proof. Since A^+ and A^- have bounded or slowly growing increments it is enough to prove the statement for a fixed outcome ω coming from the set $E(A^+) \cap E(A^-)$.

First, we prove the statement in the case of a non-decreasing A, that is, when A^- vanishes with probability 1. In this case the fixed trajectory $A(x,\omega) = A^+(x,\omega)$, $x \in \mathbb{R}$, has bounded or slowly growing increments. Fix any $0 < t \le 1$ and $r = 0, 1, \dots$. The factors of the product $x(x-1)\cdots(x-r+1)$ has absolute value not greater than r if $0 \le x \le r$, and not greater than x if $x \ge r$, which implies that

$$\begin{aligned} \big|I_r^A(t)\big| &\le \int_{\mathbb{R}} \big|x(x-1)\cdots(x-r+1)\big| t^{x-r}\, dA(x) \\ &\le \int_{-\infty}^{r} r^r t^{x-r}\, dA(x) + \int_r^{\infty} x^r t^{x-r}\, dA(x) \\ &\le \Big(\frac{r}{t}\Big)^r \int_{-\infty}^{r} t^x\, dA(x) + \frac{1}{t^r}\int_0^{\infty} x^r t^x\, dA(x)\,. \end{aligned} \tag{4.3}$$

Since the process A has finite values

$$\int_{-\infty}^{r} t^x\, dA(x) \le \int_{-\infty}^{r} 1\, dA(x) = A(r) < \infty\,, \tag{4.4}$$

and (4.3) implies that $I_r^A(1)$ is finite if $\int_{\mathbb{R}} x^r dA(x)$ exists. For the opposed direction assume that the integral $I_r^A(1)$ is finite, and consider a real value $y \ge r$ such that

$x^r \le 2x(x-1)\cdots(x-r+1)$ holds for every $x \ge y$. Then, the integral of the product $x(x-1)\cdots(x-r+1)$ with respect to A is finite on $[y,\infty)$, and we obtain that

$$\begin{aligned}\int_{\mathbb{R}} x^r\,dA(x) &= \int_{-\infty}^{y} x^r\,dA(x) + \int_{y}^{\infty} x^r\,dA(x)\\ &\le \int_{-\infty}^{y} y^r\,dA(x) + \int_{y}^{\infty} 2x(x-1)\cdots(x-r+1)\,dA(x)\\ &\le y^r A(y) + 2\int_{y}^{\infty} x(x-1)\cdots(x-r+1)\,dA(x) < \infty\,.\end{aligned}$$

Now, let $0 < t < 1$ be arbitrary and consider the value $\delta = \delta(\omega, A)$ corresponding to the outcome $\omega \in E(A)$ appearing in the definition of functions having bounded or slowly growing increments. We have the inequality

$$\begin{aligned}\int_{\mathbb{R}} x^r t^x\,dA(x) &\le \int_{\{0\}} x^r t^x\,dA(x) + \sum_{m=1}^{\infty}\int_{(m-1)\delta}^{m\delta} (m\delta)^r t^{(m-1)\delta}\,dA(x)\\ &= 0^r t^0 A(0) + \delta^r \sum_{m=1}^{\infty} m^r (t^\delta)^{m-1}\big[A\big(m\delta\big) - A\big((m-1)\delta\big)\big] = 0^r A(0) + \delta^r S_{r,\delta,A}(t^\delta)\,,\end{aligned}$$

and the series $S_{r,\delta,A}(t^\delta)$ is convergent by Proposition 4.1. Using this with formulas (4.3) and (4.4) we obtain the existence of the integral $I_r^A(t)$.

In the next step, we show that the function I_{r-1}^A can be differentiated on the interval $(0,1]$. Consider any values $0 < s, t < 1$. Then, there exists an $s_* = s_*(s,t,x-r)$ between s and t such that

$$\frac{t^{x-r+1} - s^{x-r+1}}{t-s} = (x-r)s_*^{x-r}\,,$$

from which we obtain the equation

$$\begin{aligned}\frac{I_{r-1}^A(t) - I_{r-1}^A(s)}{t-s} &= \int_{\mathbb{R}} x(x-1)\cdots(x-r+2)\frac{t^{x-r+1} - s^{x-r+1}}{t-s}\,dA(x)\\ &= \int_{\mathbb{R}} x(x-1)\cdots(x-r+2)(x-r+1)s_*^{x-r}\,dA(x)\,.\end{aligned}$$

Let $s \to t$ and assume that $t/2 < s < (t+1)/2$. Then, $s_* \to t$ and $t/2 < s_* < (t+1)/2$, as well. Since s^{x-r} is a monotone function of the variable s on the positive half-line for every fixed $x-r$, we find that

$$s_*^{x-r} \le \max\left\{\Big(\frac{t}{2}\Big)^{x-r}, \Big(\frac{t+1}{2}\Big)^{x-r}\right\} \le \Big(\frac{t}{2}\Big)^{x-r} + \Big(\frac{t+1}{2}\Big)^{x-r}\,,$$

and the inequality remains valid if we multiply both sides by $|x(x-1)\cdots(x-r+1)|$. Observe that after the multiplication the right side can be integrated with respect to A because the expressions $I_r^A(t/2)$ and $I_r^A((t+1)/2)$ are both finite. Since $s_* \to t$ the dominated convergence theorem implies that

$$\frac{I_{r-1}^A(t) - I_{r-1}^A(s)}{t-s} \to \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\,dA(x) = I_r^A(t)\,.$$

Furthermore, if the integral $\int_{\mathbb{R}} x^r dA(x)$ is finite, then $I_r^A(1)$ exists and we can apply the presented method for $t = 1$ with only some minor change. If $s \to t = 1$ such that $t/2 \leq s \leq 1$, then the formulas remain valid if we replace $(t+1)/2$ with the constant 1. We find that the inequality

$$s_*^{x-r} \leq \left(\frac{t}{2}\right)^{x-r} + 1^{x-r} \tag{4.5}$$

holds, and we can prove the differentiability of I_{r-1}^A at $t = 1$ by using the dominated convergence theorem similarly as above. Note that inequality (4.5) is still valid if s_* is replaced by s, and by using the dominated convergence theorem again, we have

$$I_r^A(s) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)s^{x-r}\, dA(x) \to I_r^A(1)\,.$$

That is, $I_r^A(t)$ is continuous at $t = 1$.

In the special case $r = 0$ it is obvious that the variable $I_0^A(0) = A(0)$ exists, and we can apply our well-tried technique to prove continuity. By letting $s \to t = 0$ such that $0 \leq s \leq 1/2$ we have $s^x \leq (1/2)^x$ for every $x \geq 0$. Since the process I_0^A is finite at the point $1/2$, the function $(1/2)^x$ can be integrated with respect to A. Using the dominated convergence theorem again, we have

$$I_0^A(s) = \int_{\mathbb{R}} s^x\, dA(x) \to \int_{\mathbb{R}} 0^x\, dA(x) = I_0^A(0)\,, \qquad s \to 0\,,$$

and the right-continuity of $I_0^A(t)$ at the point $t = 0$ follows. This completes the proof of the proposition for a non-decreasing $A(x)$, $x \in \mathbb{R}$.

To handle the case of an arbitrary A observe that the processes A^+ and A^- are non-decreasing, and hence, the statement of the proposition is valid for the functions I_r^+ and I_r^-. Since we have

$$I_r^A(t) = I_r^+(t) - I_r^-(t) \tag{4.6}$$

for every possible r and t, the value $I_r^A(1)$ is well-defined if and only if the integrals $\int_{\mathbb{R}} x^r dA^+(x)$ and $\int_{\mathbb{R}} x^r dA^-(x)$ are finite, which condition is equivalent with the existence of the r-th moment $\int_{\mathbb{R}} x^r dA(x)$. The other statements immediately come from the representation (4.6) and the results of the lemma for a non-decreasing A. □

Note that if the integral $I_r^A(t)$ exists for some fixed t and r in Lebesgue-Stieltjes sense, then it is also well-defined as an improper Riemann–Stieltjes integral. To show this fact consider any $T_1 < 0$ and $T_2 \geq r$, and let $\mathbb{1}_B(x)$, $x \in \mathbb{R}$, stands for the indicator function of a given interval B. Also, let $d_L A(x)$ and $d_R A(x)$ denote integration with respect to the process $A(x)$, $x \in \mathbb{R}$, in Lebesgue–Stieltjes and in Riemann–Stieltjes sense, respectively. Assume that the integral $I_r^A(t)$ exists in Lebesgue–Stieltjes sense. Then, the function

$$f(x) = x(x-1)\cdots(x-r+1)t^{x-r}\,, \qquad x \in \mathbb{R}\,,$$

can be integrated with respect to A on $(T_1, T_2]$ in Lebesgue–Stieltjes sense. Since f is continuous, it follows that it can be integrated on $[T_1, T_2]$ also in Riemann–Stieltjes sense, and we have

$$\int_{T_1}^{T_2} f(x)\, d_R A(x) = \int_{T_1}^{T_2} f(x)\, d_L A(x) = \int_{-\infty}^{r} f(x)\, d_L A(x) + \int_{r}^{T_2} f(x)\, d_L A(x)\,.$$

Since f is positive on $(r, T_2]$, the monotone convergence theorem implies that

$$\int_{r}^{T_2} f(x)\, d_L A^+(x) = \int_{r}^{\infty} f(x) \mathbb{1}_{(r,T_2]}(x)\, d_L A^+(x) \to \int_{r}^{\infty} f(x)\, d_L A^+(x) < \infty\,,$$

as $T_2 \to \infty$, and the same holds if A^+ is replaced by A^-. Hence we have

$$\int_{T_1}^{T_2} f(x)\, d_R A(x) = \int_{-\infty}^{r} f(x)\, d_L A(x) + \int_{r}^{T_2} f(x)\, d_L A(x) \to \int_{\mathbb{R}} f(x)\, d_R A(x)\,,$$

as $T_1 \to -\infty$ and $T_2 \to \infty$. That is, the integral $I_r^A(t)$ exists in Riemann–Stieltjes sense, and it has the same value as in Lebesgue–Stieltjes sense. Unfortunately, the reverse fails, because it is possible that $\int_{-\infty}^{T_2} f(x) d_R A(x)$ has finite limit, but the integrals $\int_{-\infty}^{T_2} f(x) d_L A^+(x)$ and $\int_{-\infty}^{T_2} f(x) d_L A^-(x)$ explode as $T_2 \to \infty$. Of course, in such a case one can not apply Proposition 4.2.

Proposition 4.3. *Assume that the function $E\langle M\rangle_x$, $x \in \mathbb{R}$, has bounded or slowly growing increments. Then, the process $I_r^M(t)$, $r = 0, 1, \ldots$ is well-defined and continuous in L^2 sense on $(0, 1)$, and it has a sample-continuous modification $\bar{I}_r^M(t)$ on this interval. In the case $r = 0$, the variable $I_0^M(0)$ is well-defined, and the processes $I_0^M(t)$ and $\bar{I}_0^M(t)$ are right-continuous at the point $t = 0$ in L^2 sense. Also, the integral $I_r^M(1)$ exists if and only if $M(x)$, $x \in \mathbb{R}$, has finite r-th moment $\int_{\mathbb{R}} x^r dM(x)$. If this condition is satisfied, then $I_r^M(t)$ and $\bar{I}_r^M(t)$ are left-continuous at $t = 1$ in L^2 sense. Furthermore, if K has finite $(r+1)$-th moment, then we can choose a modification $\bar{I}_r^M$ which is continuous almost surely at $t = 1$.*

Proof. First, we prove the finiteness of $I_r^M(t)$. Using Lemma 2.8 the stochastic integral $I_r^M(t)$ is well-defined for fixed $0 \le t \le 1$ and $r = 0, 1, \ldots$ if and only if

$$I_r^E(t) = E \int_{\mathbb{R}} \Big[x(x-1)\cdots(x-r+1) t^{x-d} \Big]^2 d\langle M\rangle_x < \infty\,.$$

For $r = 0$ and $t = 0$ we have

$$I_0^E(0) = E \int_{\mathbb{R}} t^0\, d\langle M\rangle_x = E \int_{\{0\}} 1\, d\langle M\rangle_x = E\langle M\rangle_0 < \infty\,,$$

and hence, $I_0^M(0)$ exists. Now, consider any $0 < t \le 1$ and $r = 0, 1, \ldots$, and use the same arguments as in formula (4.3) in the previous proof. That is, observe that the

factors in the product $x(x-1)\cdots(x-r+1)$ has absolute value not greater than r if $0 \le x \le r$, and not greater than x if $x \ge r$. From this we obtain the inequality

$$\begin{aligned} I_r^E(t) &= E\int_{\mathbb{R}} \Big[x(x-1)\cdots(x-r+1)t^{x-r}\Big]^2 d\langle M\rangle_x < \infty \\ &\le E\int_{-\infty}^{r} r^{2r}t^{2(x-r)}\, d\langle M\rangle_x + E\int_{r}^{\infty} x^{2r}t^{2(x-r)}\, d\langle M\rangle_x \\ &\le \Big(\frac{r}{t}\Big)^{2r} E\int_{-\infty}^{r} t^{2x}\, d\langle M\rangle_x + \frac{1}{t^{2r}} E\int_{0}^{\infty} x^{2r}t^{2x}\, d\langle M\rangle_x\,. \end{aligned} \tag{4.7}$$

Since the mean $E\langle M\rangle_x$ is finite for every $x \in \mathbb{R}$, it follows that

$$E\int_{-\infty}^{r} t^{2x}\, d\langle M\rangle_x \le E\int_{-\infty}^{r} 1\, d\langle M\rangle_x = E\langle M\rangle_r < \infty\,. \tag{4.8}$$

If the stochastic integral $\int_{\mathbb{R}} x^r dM(x)$ is well-defined then it has variance

$$E\int_{\mathbb{R}} x^{2r}\, d\langle M\rangle_x < \infty\,,$$

which implies the finiteness of $I_r^E(1)$ and we obtain the existence of $I_r^M(1)$. Contrary, suppose that the integral $I_r^M(1)$ is well-defined, and choose a value $y \ge r$ such that $2x(x-1)\cdots(x-r+1) \ge x^r$ holds for every $x \ge y$. We obtain that

$$\begin{aligned} E\int_{\mathbb{R}} x^{2r}\, d\langle M\rangle_x &= E\int_{-\infty}^{y} x^{2r}\, d\langle M\rangle_x + E\int_{y}^{\infty} x^{2r}\, d\langle M\rangle_x \\ &\le E\int_{-\infty}^{y} y^{2r}\, d\langle M\rangle_x + E\int_{y}^{\infty} \Big[2x(x-1)\cdots(x-r+1)\Big]^2 d\langle M\rangle_x \\ &\le y^{2r} E\langle M\rangle_y + 4I_r^E(1) < \infty\,, \end{aligned}$$

from which the integral $\int_{\mathbb{R}} x^r dM(x)$ exists.

Fix an arbitrary point $0 < t < 1$ and consider the value $\delta = \delta(E\langle M\rangle)$ coming form the definition of functions having bounded or slowly growing increments. Note that $E\langle M\rangle$ is deterministic, and hence, δ does not depend on the outcome ω. Since the process $\langle M\rangle$ is non-decreasing, we have the inequality

$$\begin{aligned} E\int_{0}^{\infty} x^{2r}t^{2x}\, d\langle M\rangle_x &\le E\left[\sum_{m=1}^{\infty} (m\delta)^{2r} t^{(m-1)\delta}\Big[\langle M\rangle_{m\delta} - \langle M\rangle_{(m-1)\delta}\Big]\right] \\ &= \delta^{2r}\sum_{m=1}^{\infty} m^{2r}(t^\delta)^{m-1}\Big[E\langle M\rangle_{m\delta} - E\langle M\rangle_{(m-1)\delta}\Big] = \delta^{2r} S_{2r,\delta,\langle M\rangle}(t^\delta)\,, \end{aligned}$$

and by Proposition 4.1 the series on the right side is convergent. Then, (4.7) and (4.8) imply the finiteness of $I_r^E(t)$, form which the stochastic integral $I_r^M(t)$ is well-defined.

Let us examine the continuity of the process I_r^M on the interval $(0,1)$. By considering arbitrary values $0 < s, t < 1$ there exists a constant $s_* = s_*(s,t,x-r)$ between s and t such that

$$t^{x-r} - s^{x-r} = (t-s)(x-r)s_*^{x-r-1}\,.$$

From this we obtain the identity

$$\begin{aligned} E\Big(\big|I_r^M(t) - I_r^M(s)\big|^2\Big) &= E\bigg[\int_{\mathbb{R}} x(x-1)\cdots(x-r+1)(t^{x-r}-s^{x-r})\,dM(x)\bigg]^2 \\ &= E\int_{\mathbb{R}}\Big[x(x-1)\cdots(x-r+1)(t^{x-r}-s^{x-r})\Big]^2 d\langle M\rangle_x \\ &= (t-s)^2 E\int_{\mathbb{R}}\Big[x(x-1)\cdots(x-r+1)(x-r)s_*^{x-r-1}\Big]^2 d\langle M\rangle_x\,. \end{aligned} \tag{4.9}$$

For a fixed integer $n \geq 2$ let $\tau = (1/2)^n$, and assume that $s, t \in [\tau, 1-\tau]$. Then, the value s_*^2 lies between τ^2 and $(1-\tau)^2$, and we have

$$s_*^{2(x-r-1)} \leq \max\big\{\tau^{2(x-r-1)}, (1-\tau)^{2(x-r-1)}\big\} \leq \tau^{2(x-r-1)} + (1-\tau)^{2(x-r-1)}$$

for every $x \geq 0$. Multiplying this inequality by $[x(x-1)\cdots(x-r)]^2$, integrating both sides with respect to $\langle M\rangle$, taking the mean, and using (4.9) we get that

$$E\Big(\big|I_r^M(t) - I_r^M(s)\big|^2\Big) \leq (t-s)^2\Big[I_{r+1}^E(\tau) + I_{r+1}^E(1-\tau)\Big]\,. \tag{4.10}$$

Since the terms $I_{r+1}^E(\tau)$ and $I_{r+1}^E(1-\tau)$ are finite by the first part of the proof, the right side of (4.10) converges to 0 as $s \to t$, which implies the L^2 continuity of I_r^M at t. Since t is an arbitrary point in $[(1/2)^n, 1-(1/2)^n]$, the process I_r^M is continuous in L^2 sense on this interval, and by letting $n \to \infty$ we obtain the L^2 continuity on the whole $(0,1)$. Also, from (4.10) the Kolmogorov–Chentsov theorem (Theorem 2.8 of Karatzas and Shreve (1988)) implies that I_r^M has a sample-continuous modification $\bar{I}_{r,n}^M$ on $[(1/2)^n, 1-(1/2)^n]$ for every n. Note that by the dyadic construction method of the Kolmogorov–Chentsov theorem these modifications are expansions of each other, that is, we have

$$\bar{I}_{r,n+1}^M(t) = \bar{I}_{r,n}^M(t) \qquad \text{for every} \qquad t \in \bigg[\Big(\frac{1}{2}\Big)^n, 1-\Big(\frac{1}{2}\Big)^n\bigg], \quad n = 2,3,\ldots$$

From this we obtain that the process $\bar{I}_r^M(t)$, $0 \leq t \leq 1$, defined by the equations $\bar{I}_0^M(0) = I_0^M(0)$, $\bar{I}_r^M(1) = I_r^M(1)$,

$$\bar{I}_r^M(t) = \bar{I}_{r,n}^M(t)\,, \qquad t \in \bigg[\Big(\frac{1}{2}\Big)^n, 1-\Big(\frac{1}{2}\Big)^n\bigg], \quad n = 2,3,\ldots \tag{4.11}$$

is a sample-continuous modification of I_r^M on the interval $(0,1)$.

It only remains to prove the continuity of the processes at the endpoints of $[0,1]$. Consider the case $r=0$ and $t=0$, and recall that the function $t^x=0^x$ is defined as 0 if $x>0$ and as 1 if $x=0$. Then, for any $0\le s\le 1/2$ we obtain that

$$\begin{aligned} E\Big(\big|I_0^M(s)-I_0^M(0)\big|^2\Big) &= E\bigg[\int_{\mathbb{R}}\big(s^x-0^x\big)\,dM(x)\bigg]^2 \\ &= E\int_{\mathbb{R}}\big(s^x-0^x\big)^2 d\langle M\rangle_x = E\int_0^\infty s^{2x}\,d\langle M\rangle_x \\ &\le E\int_{\mathbb{R}}\Big(\frac12\Big)^{2x} d\langle M\rangle_x = I_0^E(1/2)<\infty\,. \end{aligned}$$

Since $s^{2x}\to 0$ for every $x>0$ as $s\to 0$, it follows from the dominated convergence theorem that

$$E\Big(\big|I_0^M(s)-I_0^M(0)\big|^2\Big) = E\int_0^\infty s^{2x}\,d\langle M\rangle_x\to 0\,,$$

and we get the L^2 continuity of $I_0^M(t)$ at $t=0$. Similarly, let $r=0,1,\dots$ be arbitrary and assume that the integral $I_r^M(1)$ exists. If $s\to t=1$ such that $1/2\le s\le 1$, then we have

$$\big|1-s^{x-r}\big|\le 1+s^{x-r}\le 1+\max\big\{1^{x-r}+(1/2)^{x-r}\big\}\le 1+\big(1+(1/2)^{x-r}\big)$$

for every $x\ge 0$, from which

$$\big[1-s^{x-r}\big]^2\le 4+4(1/2)^{x-r}+(1/4)^{x-r}\,.$$

After multiplying the inequality by $[x(x-1)\cdots(x-r+1)]^2$, integrating both sides with respect to $\langle M\rangle$, and taking the mean we get the finite sum $4I_r^E(1)+4I_r^E(1/2)+I_r^E(1/4)$ on the right. Then, the second equation of (4.9) and the dominant convergence theorem imply that

$$E\Big(\big|I_r^M(1)-I_r^M(s)\big|^2\Big) = E\int_{\mathbb{R}}\Big[x(x-1)\cdots(x-r+1)\big(1-s^{x-r}\big)\Big]^2 d\langle M\rangle_x\to 0\,,$$

because s^{x-r} converges to 1, and hence, the function in the integral converges to 0 for every $x\ge 0$ as $s\to 1$. That is, we have obtained the L^2 continuity of $I_0^M(t)$ and $I_r^M(t)$ at $t=0$ and $t=1$, respectively. Since $\bar I_r^M$ is a modification of the process I_r^M, it also satisfies these continuity properties.

Finally, assume that the process M has finite $(r+1)$-st moment. For a fixed integer $n\ge 2$ let $\tau=(1/2)^n$ and consider any values $s,t\in[\tau,1]$. We can repeat the computation presented in (4.9) and (4.10) with the minor change that $1-\tau$ is replaced by the constant 1. Then, the Kolmogorov–Chentsov theorem provides sample-continuous modifications on the intervals $[(1/2)^n,1]$, and we can obtain a sample-continuous modification of I_r^M on $(0,1]$ by using the method presented in (4.11). □

We can summarize the results of Propositions 4.2 and 4.3 in the following theorem.

Theorem 4.4. *Assume that the function $E\langle M\rangle_x$ and the processes $A^+(x)$ and $A^-(x)$, $x \in \mathbb{R}$, have bounded or slowly growing increments. Then, the integral $I_r(t)$ is well-defined and the process $I_r(t)$ has a sample continuous modification $\bar{I}_r(t)$ on the interval $(0,1)$ for any $r = 0, 1, \ldots$. Also, $\bar{I}_0(t)$ is right-continuous at the point $t = 0$ in stochastic sense. Furthermore, $I_r(1)$ is well-defined if and only if both M and A have finite r-th moment. If this condition holds, then $\bar{I}_r(t)$ is left-continuous at $t = 1$ in probability.*

Proof. Working with the process $\bar{I}_r^M$ provided by Proposition 4.3 consider the random function

$$\bar{I}_r(t) = \bar{I}_r^M(t) + I_r^A(t)\,,$$

in case of every t and r for which the processes on the right side are defined. Then, $\bar{I}_r$ is clearly a modification of I_r, and all of the statements of the theorem immediately follow from Propositions 4.2 and 4.3. The only harder question is the continuity of $\bar{I}_r$ at those endpoints of the interval $[0,1]$, where it is defined. Since the processes $\bar{I}_r^M$ and I_r^A are continuous at these points in L^2 and almost sure sense, respectively, we can state only stochastic continuity for the sum of them. □

After we proved the existence of the process I_r in the following we investigate its supremum on certain subintervals of $[0,1]$. We will see that this supremum can be bounded by a linear functional of the supremum $K(x)$, $x \in \mathbb{R}$, which will be crucial in the applications later. Proposition 4.6 provides a pointwise inequality, and the main result is stated in Theorem 4.7. But before that, we need a technical tool.

Proposition 4.5. *If $f(x)$, $x \in \mathbb{R}$, is a continuous deterministic function of bounded variation on every finite interval, then for any $a < 0 < b$ we have the integration by parts formula*

$$\int_a^b f(x)\,dK(x) = f(b)K(b) - \int_a^b K(x)\,df(x)\,.$$

Proof. Consider the representation $K(x) = M(x) + A(x)$, $x \in \mathbb{R}$. Since we have

$$\int_a^b f(x)\,dK(x) = \int_a^b f(x)\,dM(x) + \int_a^b f(x)\,dA(x)\,,$$

its enough to prove the statement for $K = A$ and for $K = M$, separately. The process A is of bounded variation on finite intervals, and the integral of f with respect to A is considered in Lebesgue–Stieltjes sense, so we can apply the standard integration by parts formula for the second term of the sum. Because A vanishes on the negative half-line we obtain

$$\begin{aligned}\int_a^b f(x)\,dA(x) &= f(b)A(b) - f(a)A(a) - \int_a^b A(x)\,df(x)\\ &= f(b)A(b) - \int_a^b A(x)\,df(x)\,.\end{aligned}$$

To obtain the formula for the stochastic integral let

$$M_-(x) = \lim_{y \uparrow x} M(y)\,, \qquad x \in \mathbb{R}\,,$$

be the left-continuous version of the process M. That is, M_- and M have the same value at the points, where M is continuous, and M_- is equal to the left-side limit of M at the points of discontinuity. The stochastic integration by parts formula provides the equation

$$\begin{aligned}\int_a^b f(x)\,dM(x) &= f(b)M(b) - f(a)M(a) - \int_a^b M_-(x)\,df(x) - \big[M, f\big]_b \\ &= f(b)A(b) - \int_a^b M_-(x)\,df(x) - \big[M, f\big]_b\,,\end{aligned}$$

where the correction term $[M, f]_b$, $b \geq 0$, is the so-called quadratic covariation of M and f. Since the function f is deterministic and continuous, Theorem 4.52 in Chapter I of Jacod and Shiryaev (2003) implies that the quadratic covariation vanishes for every b. Also, the local martingale M is càdlàg which ensures that it has only countably many points of discontinuity. Since f is continuous at every point of the interval $[a, b]$ we can replace M_- with M in the integral on the right side, and we have

$$\int_a^b f(x)\,dM(x) = f(b)A(b) - \int_a^b M(x)\,df(x)\,. \qquad \square$$

Proposition 4.6. *Assume that the integral $I(t)$ exists for some fixed $0 \leq t \leq 1$. Then,*

$$|I(t)| \leq \sup_{x \in \mathbb{R}} |K(x)| \qquad a.s.$$

Furthermore, fix $r = 1, 2, \ldots$ and assume that $I_r(t)$ is well-defined for some $0 < t < 1$. Then, we have

$$|I_r(t)| \leq C_0(t, r) \sup_{x \in \mathbb{R}} |K(x)| \qquad a.s.$$

with the constant

$$C_0(t, r) = \frac{(2r+4)x_0^r}{t^d}\,, \qquad \textit{where} \qquad x_0 = x_0(t, r) = \max\left\{\frac{r\big(r - \ln t\big)2^r}{-\ln t}, r\right\}.$$

Proof. Since the first inequality requires more sophisticated technique than the second one, we separate the proofs. Let $t \in [0, 1]$ be an arbitrary value for which the integral $I(t)$ exists. If $t = 0$, then $t^x = 1$ for $x = 0$, and $t^x = 0$ for $x > 0$, and thus,

$$|I(0)| = |K(0)| \leq \sup_{x \in \mathbb{R}} |K(x)|.$$

Also, for $t = 1$ we have

$$|I(1)| = \Big|\lim_{x \to \infty} K(x)\Big| \leq \sup_{x \in \mathbb{R}} |K(x)|.$$

Having the trivial boundary cases disposed of, let $0 < t < 1$, and choose any $a > 1$ and $b > 0$. Using the fact that the process K vanishes on the negative half-line, the integration by parts formula of Proposition 4.5 implies that

$$\int_{-\infty}^{b} t^x d\,K(x) = \int_{\log_t a}^{b} t^x\,dK(x) = t^b K(b) - \int_{\log_t a}^{b} K(x)\,dt^x\,.$$

Since t^x is a decreasing function of the variable x, its total variation on the interval $[\log_t a, b]$ is equal to $(a - t^b)$. Whence

$$\left|\int_{-\infty}^{b} t^x\,dK(x)\right| \le t^b|K(b)| + (a - t^b)\sup_{\log_t a\le x\le b}|K(x)| \le a\sup_{x\in\mathbb{R}}|K(x)|.$$

If now $a \to 1$ and $b \to \infty$, we obtain the first inequality.

For the second statement fix any values $0 < t < 1$ and $r = 1, 2, \ldots$ such that $I_r(t)$ is well-defined. Let

$$h_t(x) = \big[x(x-1)\cdots(x-r+1)\big]t^x = \big[a_r x^r + a_{r-1}x^{r-1} + \cdots + a_1 x + a_0\big]t^x\,.$$

Since the constants in the product on the left side of the equation are all smaller than r, the function $h_t(x)$, $x \in \mathbb{R}$, is positive on the interval (r, ∞). Also, the multinomial theorem implies that if $x \ge r$ then

$$|a_k| \le \binom{r}{k} r^{r-k} \le \binom{r}{k} r x^{r-k-1}\,, \qquad k = 0, \ldots, r-1\,,$$

and we have $a_r = 1 \le r$. The derivative of $h_t(x)$ with respect to the variable x is

$$\begin{aligned} h_t'(x) &= \left[\sum_{k=0}^{r-1} a_{k+1}(k+1)x^k\right]t^x + \left[\sum_{k=0}^{r} a_k x^k\right]t^x \ln t \\ &= \left[x^r \ln t + \sum_{k=0}^{r-1}\Big[a_{k+1}(k+1) + a_k \ln t\Big]x^k\right]t^x\,. \end{aligned}$$

The factor $\ln t$ is negative, so for every $x \ge r$ we have

$$\begin{aligned} \left|\sum_{k=0}^{r-1}\Big[a_{k+1}(k+1) + a_k \ln t\Big]x^k\right| &\le \sum_{k=0}^{r-1}|a_{k+1}|(k+1)x^k - \ln t\sum_{k=0}^{r-1}|a_k|x^k \\ &\le r^2 x^{r-1}\sum_{k=0}^{r-1}\binom{r}{k+1} - r x^{r-1}\ln t\sum_{k=0}^{r-1}\binom{r}{k} \le r\big(r - \ln t\big)x^{r-1}2^r\,. \end{aligned}$$

From this, we find that the inequality

$$h_t'(x) \le x^r \ln t + r\big(r - \ln t\big)2^r x^{r-1} < 0$$

holds if $x > x_0(t,r)$. That is, the function h_t is positive and decreasing on (x_0,∞). The derivative h_t' has at most r different zero points, which implies that the function h_t has at most r local extremum points and they are positioned in the set $(-\infty, x_0]$. The factors of the product $x(x-1)\cdots(x-r+1)$ have absolute value not greater than x_0 if $x \in [-1, x_0]$, so we have $|h_t(x)| \le x_0^r$ on this interval. Now, consider the local extremum points positioned on $[-1, x_0]$. Since h_t is bounded on this interval, its total variation between two adjoining extremum points is not greater than $2x_0^r$. Furthermore, $2x_0^r$ is an upper bound on the total variation between the smallest extreme on this interval and -1, and also, on the total variation between the largest extreme and x_0. Since the function h_t has at most r different extremum points, its total variation on $[-1, x_0]$ can not be greater than $2(r+1)x_0^r$.

Let $b \ge x_0$ be an arbitrary real value. Using Proposition 4.5 we obtain that

$$\int_{-\infty}^{b} h_t(x)\,dK(x) = \int_{-1}^{b} h_t(x)\,dK(x) = h_t(b)K(b) - \int_{-1}^{b} K(x)\,dh_t(x)\,.$$

The function h_t is decreasing and positive on $[x_0,\infty)$ from which its total variation on $[x_0, b]$ is not greater than x_0^r. Summing up, the total variation of h_t on $[-1, b]$ can not be greater than $(2r+3)x_0^r$. Since $0 \le h_t(b) \le x_0^r$ we have

$$\left| \int_{-\infty}^{b} f_t(x) t^{-d}\,dK(x) \right| \le \frac{x_0^r}{t^d}|K(b)| + \frac{(2r+3)x_0^r}{t^d} \sup_{x\in\mathbb{R}} |K(x)| \le C_0(t,r) \sup_{x\in\mathbb{R}} |K(x)|\,.$$

If $b \to \infty$ then we obtain the desired inequality, and the proof is complete. □

Theorem 4.7. *Assume that the function $E\langle M\rangle_x$ and the processes $A^+(x)$ and $A^-(x)$, $x \in \mathbb{R}$, has bounded or slowly growing increments, and consider any modification $\bar{I}_r(t)$ of the process $I_r(t)$ which is sample-continuous on the interval $(0,1)$. Then, with $\bar{I} = \bar{I}_0$ we have the inequality*

$$\sup_{0\le t<1} |\bar{I}(t)| \le \sup_{x\in\mathbb{R}} |K(x)| \qquad a.s.$$

Also, if $M(x)$, $A^-(x)$ and $A^+(x)$, $x \in \mathbb{R}$, have finite limit at infinity with probability 1, then the approximation can be extended to the point $t = 1$. Furthermore, in case of any $0 < \varepsilon < 1/2$ and $r = 1, 2, \ldots$ we have

$$\sup_{\varepsilon\le t\le 1-\varepsilon} |\bar{I}_r(t)| \le C_1(\varepsilon, r) \sup_{x\in\mathbb{R}} |K(x)| \qquad a.s.,$$

where the constant in the formula is independent from K, and it is

$$C_1(\varepsilon,r) = \frac{(2r+4)x_1^r}{t^d} \qquad with \qquad x_1 = x_1(\varepsilon,r) = \max\left\{ \frac{r\big(r - \ln\varepsilon\big)2^r}{-\ln(1-\varepsilon)}, r \right\}.$$

Note that Theorem 4.4 provides a sample-continuous modification of the process I_r on the interval $(0,1)$ for every $r = 0, 1, \ldots$, which can be applied in this theorem.

Proof. Using Proposition 4.6 and the fact that $\bar{I}(t)$ and $I(t)$ are equal with probability 1 at any single point $0 \le t < 1$ we have

$$\sup_{0\le t<1} |\bar{I}(t)| = \sup_{t\in[0,1)\cap\mathbb{Q}} |\bar{I}(t)| = \sup_{t\in[0,1)\cap\mathbb{Q}} |I(t)| \le \sup_{x\in\mathbb{R}} |K(x)| \qquad \text{a.s.} \tag{4.12}$$

Also, if M, A^- and A^+ have finite limit at infinity, then M and A have 0-th moments

$$\int_{\mathbb{R}} x^0\, dM(x) = \lim_{x\to\infty} M(x) < \infty$$

and

$$\int_{\mathbb{R}} x^0\, dA(x) = \int_{\mathbb{R}} x^0\, dA^+(x) - \int_{\mathbb{R}} x^0\, dA^-(x) = \lim_{x\to\infty} A^+(x) - \lim_{x\to\infty} A^-(x) < \infty\,,$$

respectively, and Theorem 4.4 implies that $I_0(1)$ is well-defined. Then, using Proposition 4.6 again the supremum on the left side of (4.12) can be extended to the point $t = 1$, as well.

For the second part of the statement fix any $0 < \varepsilon < 1/2$ and $r = 1, 2, \ldots$, and consider an arbitrary value t in the interval $[\varepsilon, 1-\varepsilon]$. Working with the constants provided by Proposition 4.6 we find that

$$x_0(t,r) \le x_1(\varepsilon,r) \qquad \text{and} \qquad C_0(t,r) \le C_1(\varepsilon,r)\,,$$

and we immediately get that

$$\sup_{\varepsilon\le t\le 1-\varepsilon} |\bar{I}_r(t)| = \sup_{t\in[\varepsilon,1-\varepsilon]\cap\mathbb{Q}} |\bar{I}_r(t)| = \sup_{t\in[\varepsilon,1-\varepsilon]\cap\mathbb{Q}} |I_r(t)| \le C_1(\varepsilon,t) \sup_{x\in\mathbb{R}} |K(x)| \qquad \text{a.s.}$$

With this inequality the theorem is completely proved. □

As we specified earlier the goal of this section is provide a flexible theoretical base what we can apply in the following sections. In the applications $K(x)$, $x \in \mathbb{R}$, will stand for some empirical type process related to some non-negative integer valued random variable X. We assumed that K satisfies some conditions which are common properties of these empirical type processes, but till now we did not make any assumptions on the background variable itself. Since in most practical applications of probability generating processes X is a non-negative integer valued variable, we must examine this special case in detail.

If the variable X is integer valued, then the related empirical type processes are constant on the intervals $[m, m+1)$, $m = 0, 1, \ldots$, and hence, we can assume that $K(x)$, $x \in \mathbb{R}$, also satisfies this additional condition. Then, K is of bounded variation on every finite interval, that is, we have $K = A$ on the real line and we do not need to bother about the martingale part M. Furthermore, if the integral $I_t(t)$ exists then it can be written in the sum formula

$$I_r(t) = \sum_{m=0}^{\infty} m(m-1)\cdots(m-r+1)t^{m-r}\big[K(m) - K(m-1)\big]\,.$$

Note that the convergence of the sum does not implies the existence of the integral $I_r(t)$ in Lebesgue–Stieltjes sense, but the finiteness of the related Riemann–Stieltjes integral follows. By this motivation in our last statement we consider the integral $I_r(t)$ in the more general framework, in Riemann–Stieltjes sense. To distinguish this special case in notation let

$$J_r(t) = \sum_{m=r}^{\infty} m(m-1)\cdots(m-r+1)t^{m-r}\big[K(m) - K(m-1)\big]\,, \tag{4.13}$$

and

$$J(t) = J_0(t) = \sum_{m=0}^{\infty} t^m\big[K(m) - K(m-1)\big]\,.$$

Furthermore, let $J_r^+(t)$ and $J_r^-(t)$ stand for the corresponding sums by replacing $K(x)$ with $A^+(x)$ and $A^-(x)$, $x \in \mathbb{R}$, in (4.13).

To examine the existence of the sum $J_r(t)$ consider the inequality

$$|J_r(t)| \le \sum_{m=0}^{\infty} m^r|t|^{m-r}\big|K(m) - K(m-1)\big| \le |K(0)| + |t|^{1-r}S_{r,1,K}(t)\,. \tag{4.14}$$

If the process K has bounded or slowly growing increments on the unit intervals then $J_r(t)$ converges absolutely for every $-1 < t < 1$ by Proposition 4.1. That is, in the case of an integer valued variable X the Riemann–Stieltjes integral $J_r(t)$ can be defined on a much wider set than $I_r(t)$ in the general setup. Additionally, if both A^+ and A^- has bounded or slowly growing increments on the unit intervals, then applying the same estimation as seen in (4.14) we obtain that the sums $J_r^+(t)$ and $J_r^-(t)$ are also absolute convergent for every $-1 < t < 1$. This immediately implies that the series $J_r(t) = J_r^+(t) - J_r^-(t)$ converges absolutely and it agrees with the Lebesgue–Stieltjes integral $I_r(t)$ on the interval $(-1, 1)$. Also, it can be easily seen that $J(1)$ exists if and only if K has finite limit at infinity with probability 1. If both A^+ and A^- has almost sure finite limit at infinity, then the sums $J^+(1)$ and $J^-(1)$ are convergent, as well, from which the Lebesgue–Stieltjes integral $I(1)$ is defined, and it is equal to $J(1)$.

We can summarize the results of our examination in the following statement.

Proposition 4.8. *Assume that the process $K(x)$, $x \in \mathbb{R}$, is constant on the intervals $[m, m+1)$, $m = 0, 1, \ldots$ If K has bounded or slowly growing increments on the unit intervals, then $J_r(t)$ is absolute convergent for any $-1 < t < 1$ and $r = 0, 1, \ldots$ Furthermore, if both $A^+(x)$ and $A^-(x)$, $x \in \mathbb{R}$, has bounded or slowly growing increments on the unit intervals, then $J_r(t)$ converges absolutely on $(-1, 1)$, again, and it is equal to the Lebesgue–Stieltjes integral $I_r(t)$ on this interval. Also, if K has finite limit at infinity with probability 1, then $J(1)$ is finite. Assuming that A^+ and A^- have almost sure finite limit at infinity, the sum $J(1)$ and the integral $I(1)$ exist, and they are equal.*

Our next statement is an inequality for the supremum of the process J_r, which is similar to the result of Theorem 4.7. We also show that the function $J_r(t)$ can be differentiated in variable t.

Theorem 4.9. *Assume that the process $K(x)$, $x \in \mathbb{R}$, is constant on each interval $[m, m+1)$, $m = 0, 1, \ldots$, and it has bounded or slowly growing increments on the unit intervals. Then, the series $J_r(t)$ can be differentiated by the variable t on $(-1, 1)$, and its derivative is $J_{r+1}(t)$ for any $r = 0, 1, \ldots$. Furthermore, we have the inequality*

$$\sup_{-\tau \le t \le \tau} |J_r(t)| \le C_2(\tau, r) \sup_{x \in \mathbb{R}} |K(x)|$$

for any $0 < \tau < 1$, where the constant $C_2(\tau, r)$, not depending on K, is

$$C_2(\tau, r) = 2 \sum_{m=r}^{\infty} \frac{m!}{(m-r)!} \tau^{m-r} \le 2 \sum_{m=r}^{\infty} m^r \tau^{m-r} < \infty \,.$$

Proof. With an arbitrary integer $k \ge r$ let us define the sum

$$J_{k,r}(t) = \sum_{m=r}^{k} m(m-1) \cdots (m-r+1) t^{m-r} \big[K(m) - K(m-1) \big] \,, \qquad -1 < t < 1 \,,$$

and note that $J'_{k,r}(t) = J_{k,r+1}(t)$ on the interval $(-1, 1)$. From Proposition 4.8 it follows that under the assumptions the power series $J_r(t)$ is absolute convergent on $(-1, 1)$. This implies that the partial sums $J_{k,r}(t)$ converges uniformly to $J_r(t)$ as $k \to \infty$ on the subinterval $[-\tau, \tau]$ for any fixed value $0 < s < 1$. Then, $J_r(t)$ is also differentiable and its derivative is $J_{r+1}(t)$ on the same interval. Since τ is an arbitrary value we have $J'_r(t) = J_{r+1}(t)$ on the whole $(-1, 1)$. For the second statement we have

$$\sup_{-s \le t \le s} |J_r(t)| \le \sum_{m=d}^{\infty} \frac{m!}{(m-r)!} \big| K(m) - K(m-1) \big| s^{m-r} \le C_2(s, d) \sup_{x \in \mathbb{R}} |K(x)| \,,$$

completing the proof. □

It deserves to point out that the constant $C_2(\tau, r)$ is tight. Indeed, the supremum of the function

$$K(x) = \begin{cases} (-1)^m \,, & m \le x < m+1 \,, \quad m = 0, 1, \ldots \\ 0 \,, & x < 0, \end{cases}$$

on the real line is 1, and we have the equation

$$J_r(-\tau) = \sum_{m=r}^{\infty} m(m-1) \cdots (m-r+1)(-\tau)^{m-r}(-2)^m = (-1)^r \, C_2(\tau, r)$$

for any $\tau > 0$ and $r = 0, 1, \ldots$

4.3 The empirical probability generating process

Using the general results of the previous section now we investigate the existence and the continuity of the empirical probability generating process, its derivatives and some corresponding Gaussian processes. Let $X, X_1, X_2, \dots$ be a sequence of independent non-negative valued variables having common distribution function $F(x)$, and let $F_n(x)$ stands for the empirical distribution function based on the sample $X_1, \dots, X_n$, $x \in \mathbb{R}$. Also, consider the empirical process

$$\alpha_n(x) = n^{1/2}\big[F_n(x) - F(x)\big]\,, \qquad x \in \mathbb{R}\,.$$

The theoretical and the empirical probability generating function corresponding to the sample variables can be written in the integral forms

$$g(t) = Et^X = \int_{\mathbb{R}} t^x\, dF(x)\,, \qquad g_n(t) = \frac{1}{n}\sum_{i=1}^{n} t^{X_i} = \int_{\mathbb{R}} t^x\, dF_n(x)\,. \tag{4.15}$$

Since $F(x)$ and $F_n(x)$, $x \in \mathbb{R}$, are bounded on the real line, they have bounded or slowly growing increments, and by Proposition 4.2 the functions g and g_n are well-defined and infinitely many times differentiable on the interval $(0,1)$ with r-th derivative

$$g^{(r)}(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dF(x) = E\big(X(X-1)\cdots(X-r+1)t^{X-r}\big)\,,$$

and

$$g_n^{(r)}(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i(X_i-1)\cdots(X_i-r+1)t^{X_i-r}\,.$$

Also, $g(t)$ and $g_n(t)$ are right-continuous at the point $t = 0$ without any regularity condition on the variable X. The derivative $g^{(r)}(t)$ is left-continuous at $t = 1$ for some $r = 0, 1, \dots$ if and only if the function F has finite r-th moment. Since the theoretical distribution function have finite limit as $x \to \infty$, this condition is satisfied for $r = 0$. For any higher r the existence of the r-th moment of F is equivalent with the finiteness of the expected value $E(X^r)$. Also, the empirical distribution function has r-th moment

$$\int_{\mathbb{R}} x^r\, dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i^r < \infty\,,$$

which implies that the function $g_n^{(r)}(t)$ is left-continuous at $t = 1$ for any $r = 0, 1, \dots$

In the special case, when the random variable X has only non-negative integer values, the theoretical and the empirical distribution function are constant on the intervals $[m, m+1)$, $m = 0, 1, \dots$ Since $F(x)$ and $F_n(x)$, $x \in \mathbb{R}$, are bounded, they have bounded increments on the unit intervals. Then, Proposition 4.8 and Theorem 4.9 imply that the theoretical and the empirical probability generating function can be extended to $(-1, 1]$, and they have r-th derivative $g^{(r)}$ and $g_n^{(r)}$ on the interval $(-1, 1)$.

By (4.15) the empirical probability generating process can be written in the form

$$\gamma_n(t) = n^{1/2}\big[g_n(x) - g(x)\big] = \int_{\mathbb{R}} t^x \, d\alpha_n(x)\,,$$

and the properties of the functions g and g_n imply that γ_n exists and sample-continuous on the interval $[0,1]$ for any non-negative valued random variable. Also, the process has continuous r-th derivative

$$\gamma_n^{(r)}(t) = n^{1/2}\big[g_n^{(r)}(x) - g^{(r)}(x)\big] = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r} \, d\alpha_n(x) \tag{4.16}$$

on $(0,1)$ for any $r = 1, 2, \ldots$, and $\gamma_n^{(r)}(t)$ exists and it is continuous at $t = 1$ if and only if X has finite r-th moment. Furthermore, if the variable X is non-negative integer valued, then the empirical probability generating process can be extended to $(-1,1]$, and has r-th derivative $\gamma_n^{(r)}$ on the interval $(-1,1)$ for any positive integer r. Also, we have

$$\begin{aligned}\gamma_n(t) &= \sum_{m=0}^{\infty} t^m\big[\alpha_n(m) - \alpha_n(m-1)\big] = \lim_{k\to\infty}\sum_{m=0}^{k} t^m\big[\alpha_n(m) - \alpha_n(m-1)\big] \\ &= \lim_{k\to\infty}\left[\sum_{m=0}^{k-1}\big(t^m - t^{m+1}\big)\alpha_n(m) + t^k\alpha_n(k)\right] = \sum_{m=0}^{\infty}\alpha_n(m)(1-t)t^m\,,\end{aligned} \tag{4.17}$$

for every $-1 < t \le 1$, since the process $\alpha_n(x)$ has limit 0 as $x \to \infty$.

In the following section we will show that the weak limit of the empirical probability generating process for an arbitrary non-negative valued variable X can be written in the form

$$Y(t) = \int_{\mathbb{R}} t^x \, dB\big(F(x)\big)\,,$$

and we will also investigate the corresponding "differentiated" processes

$$Y_r(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r} \, dB\big(F(x)\big)\,, \tag{4.18}$$

where $B(u)$, $0 \le u \le 1$ is the Brownian bridge defined in Section 2.1. We must note that the phrase "differentiated" is misleading in the sense that the process $Y(t)$ can not be differentiated at any t with probability 1. We use this word only to point out that $Y_r(t)$ is defined by a similar integral as the differentiated process $\gamma_n^{(r)}(t)$. Not motivated by the applications but by curiosity, first we examine the related integral with respect to the standard Wiener process $W(u)$, $u \ge 0$, that is, the process

$$Z_r(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r} \, dW\big(F(x)\big)\,. \tag{4.19}$$

First of all, observe that $W(F(x))$ is a martingale and has quadratic characteristic $\langle W(F)\rangle_x = F(x)$, $x \in \mathbb{R}$. Since F vanishes on the negative half-line and the standard

Wiener process is bounded on the interval $[0,1]$, the process $W(F)$ has bounded or slowly growing increments. The quadratic characteristic $\langle W(F)\rangle = F$ also has bounded or slowly growing increments, which implies that we can apply Proposition 4.3 for the process Z_r. We obtain that Z_r is well-defined and has a sample-continuous modification on $(-1,1)$. Also, by the same proposition $Z_0(0)$ exists. Furthermore, $Z_r(t)$ is defined at the point $t=1$ if and only if the quadratic characteristic F has finite $2r$-th moment, which condition is equivalent with the finiteness of the mean $E(X^{2r})$. Of course, this is satisfied for $r=0$.

By considering the processes

$$M(x) = W\big(F(x)\big)\,, \qquad A^+(x) = 0\,, \qquad A^-(x) = W(1)F(x)\,, \qquad x\in\mathbb{R}\,,$$

the time transformed Brownian bridge can be represented by the form

$$B\big(F(x)\big) = W\big(F(x)\big) - W(1)F(x) = M(x) + A^+(x) - A^-(x) = K(x)\,, \qquad x\in\mathbb{R}\,,$$

and using the notation of Section 4.2 we have

$$Y_r(t) = I_r^M(t) + I_r^+(t) - I_r^-(t) = Z_r(t) + 0 - W(1)g_r(t)$$

That is, $Y_r(t)$ is well-defined if $0<t<1$, and the sum on the right side has a sample-continuous modification on $(0,1)$. Note that applying Theorem 4.4 for the process Y_r we obtain exactly the same sample-continuous modification. Also, $Y(t) = Y_0(t)$ exists and continuous at $t=0$ and $t=1$ without any regularity condition on the distribution function F.

If we suppose that X is a non-negative integer valued variable then we can represent the processes Z_r in the form

$$Z_r(t) = \sum_{m=r}^{\infty} m(m-1)\cdots(m-r+1)t^{m-r}\big[W\big(F(m)\big) - W\big(F(m-1)\big)\big]\,, \qquad (4.20)$$

and the sums exist and sample-continuous on $(-1,1)$ by Proposition 4.8 and Theorem 4.9. Since the Wiener process $W(u)$ is continuous at $u=1$, it follows that $W(F(x))$ has finite limit as $x\to\infty$ almost surely. Using Proposition 4.8 again, we find that the sum $Z(t)$ is convergent at $t=1$. We must note that this implies the existence of the integral in (4.19) at $t=1$ only in Riemann–Stieltjes sense. Obviously, the integral is defined in Lebesgue–Stieltjes sense if and only if the series in (4.20) converges absolutely. Using Kolmogorov's three series theorem in the case $r=0$ one can show that this condition is satisfied if and only if the sum $\sum_{m=0}^{\infty}[F(m+1)-F(m)]^{1/2}$ is finite. By rearranging the terms we get the identity

$$\begin{aligned} Z(t) &= \lim_{k\to\infty}\sum_{m=0}^{k} t^m\big[W\big(F(m)\big) - W\big(F(m-1)\big)\big] \\ &= \lim_{k\to\infty}\left[\sum_{m=0}^{k-1}\big(t^m - t^{i+1}\big)W\big(F(m)\big) + t^k W\big(F(k)\big)\right] = \sum_{m=0}^{\infty} W\big(F(m)\big)(1-t)t^i\,, \end{aligned} \qquad (4.21)$$

which implies that the process $Z(t)$ is continuous also at the point $t = 1$. Since in the integer valued case the r-th derivative $g^{(r)}$ is continuous on $(-1, 1]$, we obtain that $Y_r(t) = Z_r(t) - W(1)g^{(r)}(t)$ is well-defined and continuous on this interval, as well. Of course, in certain cases the integral $Y_r(1)$ can be considered only in Riemann-Stieltjes sense. Using the method of formula (4.21) we have also that

$$Y(t) = \sum_{m=0}^{\infty} B\big(F(m)\big)(1-t)t^i\,. \tag{4.22}$$

In the next sections we will consider the processes on some closed intervals where they are continuous on, or at least have a continuous modification. We can summarize the results of our investigation in the next statement.

Proposition 4.10. *The processes $Y_r(t)$ and $Z_r(t)$ are well-defined and have a sample-continuous modifications on $[a, b]$, where the interval $[a, b]$ can be chosen as*

- *$[\varepsilon, 1-\varepsilon]$ with any $0 < \varepsilon < 1/2$, if X is an arbitrary non-negative valued variable;*
- *$[-\tau, \tau]$ with any $0 < \tau < 1$, if X is a non-negative integer valued variable;*
- *$[0, 1]$, if X is a non-negative integer valued variable and $r = 0$.*

Furthermore, $Y = Y_0$ and $Z = Z_0$ exist at $t = 0$ and $t = 1$ for any non-negative valued variable, but at $t = 1$ the integrals might be considered only in Riemann-Stieltjes sense. Also, the r-th derivative $\gamma_n^{(r)}(t)$ exists and continuous on $[a, b]$ in each case.

Since for a negative value t the function t^{x-r} is not defined for every $x \geq 0$, at a first glance it seems that we can not apply the friendly integral forms (4.16), (4.18) and (4.19) in the second case of Proposition (4.10). However, in this case the random or deterministic measures corresponding to the functions $F(x)$, $W(F(x))$ and $B(F(x))$, $x \in \mathbb{R}$, put mass only into the non-negative integers. Since at these points x the function t^{x-r} is defined for any negative t, we can still use formulas (4.16), (4.18) and (4.19). Furthermore, for simplicity, in the following $Y_r(t)$ and $Z_r(t)$, $a \leq t \leq b$, stand for the sample-continuous modifications provided by Proposition 4.10.

4.4 Properties of the process Y_r

In the following sections we investigate Kolmogorov–Smirnov and Cramér-von Mises type statistics based on the empirical probability generating process γ_n and its derivatives. In these applications we must ensure that the statistics have bounded density functions on certain subintervals of the real line. First, we show that Y_r and Z_r are Gaussian processes with continuous covariance function on the interval $[a, b]$ provided by Proposition 4.10.

Proposition 4.11. *The random functions Y_r and Z_r are Gaussian processes on the interval $[a,b]$ with pointwise mean 0 and covariance functions*

$$\operatorname{Cov}\big(Z_r(s),Z_r(t)\big)=\int_{\mathbb{R}}\big[x(x-1)\cdots(x-r+1)\big]^2(st)^{x-r}\,dF(x)$$

and

$$\operatorname{Cov}\big(Y_r(s),Y_r(t)\big)=\int_{\mathbb{R}}\big[x(x-1)\cdots(x-r+1)\big]^2(st)^{x-r}\,dF(x)-g^{(r)}(s)g^{(r)}(t)\,,$$

where $a\le s,t\le b$.

Proof. Observe that the process defined by the form

$$B^\star(t)=W(u)-uW(1)\,,\qquad 0\le u\le 1\,,$$

is a representation of the Brownian bridge, and the variable $W(1)$ can be written as

$$W(1)=\int_{\mathbb{R}}1\,dW\big(F(x)\big)\,.$$

By introducing the function

$$h_r(t,x)=x(x-1)\cdots(x-r+1)t^{x-r}\,,$$

the processes Y_r and Z_r can be written in the stochastic integral forms

$$Z_r(t)=\int_{\mathbb{R}}h_r(t,x)\,dW\big(F(x)\big)\tag{4.23}$$

and

$$\begin{aligned}Y_r(t)&=\int_{\mathbb{R}}h_r(t,x)\,d\big[W\big(F(x)\big)-F(x)W(1)\big]\\&=\int_{\mathbb{R}}h_r(t,x)\,dW\big(F(x)\big)-g^{(r)}(t)\int_{\mathbb{R}}1\,dW\big(F(x)\big)\\&=\int_{\mathbb{R}}\Big[h_r(t,x)-g^{(r)}(t)\Big]\,dW\big(F(x)\big)\,.\end{aligned}\tag{4.24}$$

At this point we must recall our remark at the end of the previous section. Although the function $h_r(t,x)$ is not defined in variable x on the whole positive half-line if $t<0$, the integrals (4.23) and (4.24) are valid for any $a\le t\le b$. This is because we consider the processes Y_r and Z_r at a negative t only in case of an integer valued variable X, and the function t^{x-r} is well-defined at integer points x for any t.

Since Y_r and Z_r exist on the interval $[a,b]$, Lemma 2.9 implies that they are Gaussian processes on this set with pointwise mean 0. Also, we obtain the desired covariance functions

$$\operatorname{Cov}\big(Z_r(s),Z_r(t)\big)=\int_{\mathbb{R}}h_r(s,x)h_r(t,x)\,dF(x)\,,$$

and

$$\begin{aligned}\operatorname{Cov}\big(Y_r(s),Y_r(t)\big) &= \int_{\mathbb{R}} \big[h_r(s,x)-g^{(r)}(s)\big]\big[h_r(t,x)-g^{(r)}(t)\big]dF(x)\\ &= \int_{\mathbb{R}} h_r(s,x)h_r(t,x)\,dF(x) + g^{(r)}(s)g^{(r)}(t)\\ &\qquad - g^{(r)}(s)\int_{\mathbb{R}} h_r(t,x)\,dF(x) - g^{(r)}(t)\int_{\mathbb{R}} h_r(s,x)\,dF(x)\\ &= \int_{\mathbb{R}} h_r(s,x)h_r(t,x)\,dF(x) - g^{(r)}(s)g^{(r)}(t)\,,\end{aligned}$$

for any $a \le s,t \le b$. □

Proposition 4.12. *The covariance functions of Y_r and Z_r are continuous on $[a,b]^2$.*

Proof. By Theorem 1.3.4 and Corollary 1.3.5 of Ash and Gardner (1975) the covariance functions are continuous on $[a,b]^2$ if and only if the processes Y_r and Z_r are continuous in L^2 sense on the interval $[a,b]$. Using formulas (4.18) and (4.19) with Proposition 4.3 we get that Y_0 and Z_0 are L^2 continuous on $[0,1]$ for any non-negative valued X, and Y_r and Z_r are continuous in the same sense on $[\varepsilon,1-\varepsilon]$ for any $\varepsilon>0$ and $r=0,1,\dots$

It remains to investigate the case when the variable X has only non-negative integer values and $[a,b]=[-\tau,\tau]$ with some $0<\tau<1$. Introduce the function

$$J_r(t) = \int_{\mathbb{R}} \big[x(x-1)\cdots(x-r+1)\big]^2 t^{x-r}\,dF(x)\,, \qquad -\tau \le t \le \tau\,.$$

Since the distribution function F is constant on the intervals $[m,m+1)$, $m=0,1,\dots$ and it is bounded on the real line, we get that

$$|J_r(t)| = \left|\sum_{m=r}^{\infty} \big[m(m-1)\cdots(m-r+1)\big]^2 t^{m-r}\big[F(m)-F(m-1)\big]\right| \le \sum_{m=r}^{\infty} m^{2r}\tau^{m-r}\,.$$

The right side of the formula is convergent by the ratio test, and hence, the series $J_r(t)$ converges absolutely and uniformly on $[-\tau,\tau]$. Since the terms of $J_r(t)$ are all continuous on this interval, we obtain the continuity of $J_r(t)$, as well. Then, the continuity of the derivative $g^{(r)}$ on $[-\tau,\tau]$ immediately implies that the covariance functions

$$\operatorname{Cov}\big(Z_r(s),Z_r(t)\big) = J_r(st) \qquad \text{and} \qquad \operatorname{Cov}\big(Y_r(s),Y_r(t)\big) = J_r(st) - g^{(r)}(s)g^{(r)}(t)$$

are continuous on the square $[-\tau,\tau]^2$. □

Proposition 4.13. *The process Y is degenerate at every point of the interval $(0,1)$ if and only if the variable X is degenerate. Otherwise, Y is not degenerate at any $0<t<1$. Also, for an arbitrary $r=1,2,\dots$ the process Y_r is degenerate on the whole $(0,1)$ if and only if X is constant with probability 1 or $P(X\in\{0,\dots,r-1\})=1$. Otherwise, the points where the Y_r is not degenerate are dense in $(0,1)$.*

Proof. Since Y_r is a well-defined Gaussian process on $[a,b] = [\varepsilon, 1-\varepsilon]$ with any values $0 < \varepsilon < 1/2$ and $r = 0, 1, \ldots$, it is Gaussian on the whole $(0,1)$ with the same mean and covariance function. Using Proposition 4.11 the variance of the integral $Y_r(t)$ is

$$\operatorname{Var} Y_r(t) = \int_{\mathbb{R}} \left[x(x-1)\cdots(x-r+1)t^{x-r}\right]^2 dF(x)$$
$$- \left[\int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dF(x)\right]^2 = \operatorname{Var}\left[X(X-1)\cdots(X-r+1)t^{X-r}\right].$$

In the case $r = 0$ we have $\operatorname{Var} Y(t) = \operatorname{Var} t^X$. If X is constant with probability 1, then the variable t^X is also degenerate for every $0 < t < 1$, which implies that $Y_r(t)$ has variance $\operatorname{Var} t^X = 0$. Contrary, assume that X is not constant with probability 1 and consider an arbitrary $0 < t < 1$. Since t^x is a strictly monotone function of the variable x, the power t^X is not constant, either. From this, it follows that $\operatorname{Var} Y_r(t) = \operatorname{Var} t^X$ is strictly positive.

Let $r = 1, 2, \ldots$ be an arbitrary fixed integer, and consider the function

$$h_r(t,x) = x(x-1)\cdots(x-r+1)t^{x-r}, \qquad 0 < t < 1, \quad x \in \mathbb{R}.$$

If X is constant with probability 1 or $P(X \in \{0, \ldots, r-1\}) = 1$, then the variable $h_r(t,X)$ is degenerate, and hence, $\operatorname{Var} Y_r(t) = 0$ for every value $0 < t < 1$.

For the opposite direction consider an arbitrary interval $[a', b'] \subseteq (0,1)$. Suppose that there exists a non-negative valued variable X such that $P(X \in \{0, \ldots, r-1\}) < 1$ and $Y_r(X)$ is degenerate for every $t \in [a', b']$. Then, $h_r(t,X)$ is constant with probability 1 for any fixed t in $[a', b']$. By derivating $h_r(t,x)$ with respect to x we showed it in the proof of Proposition 4.6 that $h_r(t,x)$, $x \in \mathbb{R}$, has finitely many extremum points, and between them the function is strictly monotone. This implies that $h_r(t,x)$ has the same value at most finitely many points $x \in \mathbb{R}$. Since $h_r(t,X)$ is constant the variable X lies in a finite set $E = \{x_1, \ldots, x_k\} \subseteq \mathbb{R}$ almost surely. We assume that E is minimal is the sense that we have $P(X = x_i) > 0$ for any $i = 1, \ldots, k$. Because $h_r(t,X)$ is degenerate for every t, the function $h_r(t,x)$ is constant in variable x on the set E for any $a' \le t \le b'$. If any of the values $0, \ldots, r-1$ would lie in the set E, then the function $h_r(t,x)$ would be constant 0 on E, which implies that $h_r(t,X)$ would be degenerate at 0. But, this alternative does not hold, because X has value in the set $\{0, \ldots, r-1\}$ with probability strictly smaller than 1, and hence, $h_r(t,X)$ is not constant 0. That is, none of the values $0, \ldots, r-1$ lies in E, which implies that $P(X \in \{0, \ldots, r-1\}) = 0$.

Consider any elements $x_i, x_j \in E$, and let $t \ge a'$ and $\varepsilon > 0$ such that $t+\varepsilon \le b'$. Since both $h_r(t,x)$ and $h_r(t+\varepsilon, x)$ are constant in x on the set E, we have the derivatives

$$\frac{dh_r(t,x_1)}{dt} = \lim_{\varepsilon \downarrow 0} \frac{h_r(t+\varepsilon, x_1) - h_r(t,x_1)}{h} = \lim_{\varepsilon \downarrow 0} \frac{h_r(t+\varepsilon, x_2) - h_r(t,x_2)}{h} = \frac{dh_r(t,x_2)}{dt}.$$

That is, the function

$$\frac{dh_r(t,x)}{dt} = x(x-1)\cdots(x-r+1)(x-r)t^{x-r-1}, \qquad x \in \mathbb{R},$$

is also constant on E. Since E does not contain the values $0, \dots, r-1$, the fraction

$$\frac{dh_r(t,x)}{dt} \Big/ h_r(t,x) = \frac{x(x-1)\cdots(x-r+1)(x-r)t^{x-r-1}}{x(x-1)\cdots(x-r+1)t^{x-r}} = \frac{x-r}{t}$$

is valid and constant on E for any t. This implies that E has only one element, and hence, X is constant with probability 1. That is, if X is not degenerate, then there exists a $t \in [a', b']$ such that $Y_r(t)$ is not degenerate. □

We note that by the proof of our Proposition 4.13 one can construct a variable X which is not constant with probability 1 and also $P(X \in \{0, \dots, r-1\}) < 1$, but $Y_r(t)$ is degenerate for a given $0 < t < 1$ and $r = 1, 2, \dots$ Just consider two positive values x_1 and x_2 not being equal to any of the integers $0, \dots, r-1$ such that $h_r(t, x_1) = h_r(t, x_2)$, and let us define the distribution of X by $P(X = x_1) = P(X = x_2) = 1/2$.

In the following statement we investigate the supremum functionals

$$S_r = \sup_{a \le x \le b} |Y_r(t)|\,, \qquad S_r^+ = \sup_{a \le x \le b} Y_r(t) \qquad \text{and} \qquad S_r^- = -\inf_{a \le x \le b} Y_r(t)\,.$$

Depending on which case of Proposition 4.10 provides the interval $[a, b]$, Theorem 4.7 or Theorem 4.9 provides us a finite positive constant $C = C(a, b, r)$ such that

$$S_r = \sup_{a \le x \le b} |Y_r(t)| \le C \sup_{x \in \mathbb{R}} \big|B\big(F(x)\big)\big| \le C \sup_{0 \le u \le 1} |B(u)| = CS' \qquad \text{a.s.}$$

Using this and the inequalities $|S_r^+| \le S_r$ and $|S_r^-| \le S_r$, we get that S_r, S_r^+ and S_r^- are finite with probability 1. Let $F_r(s)$, $F_r^+(s)$ and $F_r^-(s)$, $s \in \mathbb{R}$, denote the distribution functions of the variables, respectively. Consider the value

$$s_r = \inf\big\{s \in \mathbb{R} : F_r(s) > 0\big\}\,,$$

and similarly, let s_r^+ and s_r^- denote the left endpoints of the supports of the functions F_r^+ and F_r^-. Using the distribution function of S' provided by Kolmogorov we have

$$P\big(S_r \le s\big) \ge P\big(S' \le s/C\big) = 1 - \sum_{k=1}^{\infty} (-1)^{k+1} \exp\big(-2k^2(s/C)^2\big) > 0$$

with any positive s. From $S_r \ge 0$, it follows that $s_r = 0$, which clearly implies the inequalities $s_r^+ \le 0$ and $s_r^- \le 0$

Proposition 4.14. *The distribution functions of S_r, S_r^+ and S_r^- are absolute continuous on the intervals $(0, \infty)$, (s_r^+, ∞) and (s_r^-, ∞), respectively. Also, they have bounded density functions on $[s_1, \infty)$, $[s_2, \infty)$ and $[s_3, \infty)$ with arbitrary values $s_1 > 0$, $s_2 > s_r^+$ and $s_3 > s_r^-$.*

Proof. If the process $Y_r(t)$ is degenerate at every point t on the interval $[a, b]$, then it is constant 0 on this interval. Hence, in this case S_r, S_r^+ and S_r^- is equal to 0 with

probability 1, and the statement clearly follows. If the process is not degenerate at every point on $[a, b]$, then the continuity of Y_r implies the identities

$$S_r = \sup\left\{Y_r(t), -Y_r(t) : a \le t \le b, t \in \mathbb{Q}\right\},$$

$$S_r^+ = \sup\left\{Y_r(t) : a \le t \le b, t \in \mathbb{Q}\right\}, \qquad S_r^- = \sup\left\{-Y_r(t) : a \le t \le b, t \in \mathbb{Q}\right\},$$

and we obtain the statement by using Theorem 2.10. □

Finally, we prove a similar statement for the Cramér–von Mises type statistics of the process Y_r as we have in the previous proposition for the supremum functionals. First, we present a general theorem for sample-continuous Gaussian processes having continuous covariance function. Our result for Y_r is stated as corollary.

Proposition 4.15. *Consider a Gaussian process $G(t)$, $\alpha \le t \le \beta$, having pointwise mean 0 and continuous covariance function on $[\alpha, \beta]^2$. If G is not degenerate at every point of the interval $[\alpha, \beta]$, and its trajectories lie in the space $L^2[\alpha, \beta]$ with probability 1, then the integral*

$$I = \int_\alpha^\beta G^2(t)\, dt$$

is absolute continuous and has bounded density function on the real line.

Proof. Let $c(s, t)$, $\alpha \le s, t \le \beta$, stand for the covariance function of the process G. Using the results of Section 5.2 in Shorack and Wellner (1986) the Karhunen–Loève expansion provides the representation

$$G(t) = \sum_{k=1}^{\infty} \lambda_k^{1/2} \xi_k e_k(t)\,, \qquad \alpha \le t \le \beta\,, \tag{4.25}$$

where $\xi_1, \xi_2, \ldots$ are independent variables on the underlying probability space, the real constants $\lambda_1, \lambda_2, \ldots$ are the eigenvalues of the Fredholm operator

$$\Phi : L^2[\alpha, \beta] \to L^2[\alpha, \beta]\,, \qquad \Phi(h)(t) = \int_a^b c(s, t) h(s)\, ds\,,$$

and the deterministic functions $e_1, e_2, \ldots$ are the corresponding eigenvectors. Note that $e_1, e_2, \ldots$ are continuous on the interval $[\alpha, \beta]$ and form an orthonormal basis for the space $L^2[\alpha, \beta]$. Also, the series on the right side of (4.25) converges with probability 1 to the process G in the standard $L^2[\alpha, \beta]$ norm. Since G is Gaussian, the variables $\xi_1, \xi_2, \ldots$ have standard normal law, and not all of the coefficients $\lambda_1, \lambda_2, \ldots$ are equal to 0, because the process G is not degenerate at every point t on $[\alpha, \beta]$. By applying the properties of Hilbert spaces the representation in (4.25) immediately implies the well-known almost sure identity

$$\int_\alpha^\beta G^2(t)\, dt = \sum_{k=1}^{\infty} \lambda_k \xi_k^2\,.$$

Let the variable I denote the square integral of G on $[\alpha, \beta]$. The variables $\xi_1^2, \xi_2^2, \ldots$ have χ^2 distribution with degree of freedom 1, and we obtain the characteristic function of I in the form

$$\phi_I(t) = \prod_{k=1}^{\infty} \phi_{\xi_k^2}(\lambda_k t) = \prod_{k=1}^{\infty} \left(1 - 2i\lambda_k t\right)^{-1/2}, \qquad t \in \mathbb{R}\,.$$

The characteristic function has absolute value

$$|\phi_I(t)| = \left[\prod_{k=1}^{\infty} \left|1 - 2i\lambda_k t\right|\right]^{-1/2} = \left[\prod_{k=1}^{\infty} \left(1 + 4\lambda_k^2 t^2\right)\right]^{-1/4},$$

and by using the multinomial theorem and keeping only the main term 1 and the terms with factor t^6 we have

$$\prod_{k=1}^{\infty} \left(1 + 4\lambda_k^2 t^2\right) \geq 1 + 4^3 \sum_{i,j,k=1}^{\infty} \lambda_i^2 \lambda_j^2 \lambda_k^2 t^6 = 1 + Dt^6\,. \tag{4.26}$$

Since the characteristic function of a finite random variable is not constant 0 on the whole set $\mathbb{R} \setminus \{0\}$, the absolute value of $\phi_I(t)$ is strictly positive for some $t \neq 0$. At this point t the left side of (4.26) is finite, which implies that the constant D is also finite. Using (4.26) the absolute integral of the characteristic function can be bounded by

$$\int_{\mathbb{R}} |\phi_I(t)|\, dt \leq \int_{\mathbb{R}} \left(1 + Dt^6\right)^{-1/4} dt \leq 2 \int_0^{D^{1/6}} 2^{-1/4}\, dt + 2 \int_{D^{1/6}}^{\infty} \left(2Dt^6\right)^{-1/4} dt\,,$$

and the right side is clearly finite. Then, the variable I is absolute continuous and its density function $f_I(x)$, $x \in \mathbb{R}$, can be obtained from the characteristic function ϕ_I by the inverse Fourier transform. We get that

$$|f_I(x)| = \frac{1}{2\pi} \left| \int_{\mathbb{R}} e^{-itx} \phi_I(t)\, dt \right| \leq \frac{1}{2\pi} \int_{\mathbb{R}} |e^{-itx}||\phi_I(t)|\, dt = \frac{1}{2\pi} \int_{\mathbb{R}} |\phi_I(t)|\, dt\,,$$

and hence, the density function is bounded. □

Corollary 4.16. *Consider the interval $[a, b]$ of Proposition 4.10. If X is not constant with probability 1 and $P(X \in \{0, \ldots, r-1\}) < 1$, then the distribution function of the variable*

$$\int_a^b Y_r^2(t)\, dt$$

is absolute continuous and has bounded density function on the real line.

Proof. Since Y_r is sample-continuous on $[a, b]$, it lies in the space $L^2[a, b]$ almost surely. Also, by Proposition 4.13 the process is not degenerate at every point t on the interval, and Proposition 4.11 implies that Y_r has pointwise mean 0 and continuous covariance function. Then, the statement follows from Proposition 4.15. □

4.5 Weak convergence

Let us consider independent and identically distributed variables $X, X_1, X_2, \dots$ having only non-negative values with distribution function $F(x)$ and theoretical probability generating function $g(t)$. Also, let $F_n(x)$, $x \in \mathbb{R}$, and $g_n(t)$, $0 \le t \le 1$, denote the empirical distribution function and the empirical probability generating function based on the sample $X_1, \dots, X_n$. As we saw it in Section 4.3 the empirical probability generating process can be written in the form

$$\gamma_n(t) = n^{1/2}\big[g_n(t) - g(t)\big] = \int_{\mathbb{R}} t^x \, d\alpha_n(x)\,,$$

where $\alpha_n(x)$, $x \in \mathbb{R}$, is the empirical process corresponding to the given sample. Also, by using the Brownian bridge $B(u)$, $0 \le u \le 1$, consider the process

$$Y(t) = \int_{\mathbb{R}} t^x \, dB\big(F(x)\big)\,.$$

By the results of Proposition 4.10, the processes γ_n and Y are well-defined and sample-continuous on an arbitrary interval $[\varepsilon, 1-\varepsilon]$, $\varepsilon > 0$, and in case of an integer valued X they exist and continuous on the whole $[0,1]$.

In statistical applications the most important question is the weak convergence of the process γ_n in the space $C[0,1]$ for non-negative integer valued variables. The convergence was firstly proved independently by Csörgő and Mason (1989) and Marques and Pérez-Abreu (1989) for variables having finite variance. Later, Rémillard and Theodorescu (2000) generalized their results for arbitrary non-negative integer valued variables, but unfortunately, there is oversight in their proof. However, their basic idea is very interesting and the proof can be corrected, as we see in our next result.

Theorem 4.17 (Rémillard and Theodorescu, 2000)**.** *For an arbitrary non-negative integer valued random variable X the empirical probability generating process γ_n converges weakly to Y in the space $C[0,1]$ as $n \to \infty$.*

Proof. Since the distribution of the defined processes is the same for any representation of the sample variables $X_1, X_2, \dots$, we can assume without the loss of generality that

$$X_i = F^{-1}(U_i)\,, \qquad i = 1, 2, \dots\,,$$

where $U_1, U_2, \dots$ are independent random variables on a suitable probability space having uniform distribution on the interval $[0,1]$, and

$$F^{-1}(u) = \inf\big\{x \in \mathbb{R} : F(x) \ge 0\big\}\,, \qquad 0 < u < 1\,,$$

is the quantile function corresponding to the theoretical distribution function F. As we saw it in Section 2.1, in this representation the empirical process of the variables $X_1, \dots, X_n$ can be written in the form

$$\alpha_n(x) = \beta_n\big(F(x)\big)\,, \qquad x \in \mathbb{R}\,,$$

with the uniform empirical process $\beta_n(u)$, $0 \le u \le 1$, corresponding to $U_1, \ldots, U_n$. Then, using (4.17) and (4.22) the empirical probability generating process and the limiting process can be written as

$$\gamma_n(t) = \sum_{i=0}^{\infty} \beta_n\big(F(i)\big)(1-t)t^i \qquad \text{and} \qquad Y(t) = \sum_{i=0}^{\infty} B\big(F(i)\big)(1-t)t^i\,, \qquad 0 \le t \le 1\,.$$

Let C_0 be the subspace of all functions in $C[0,1]$ which are vanish at the point 1. In the original proof Rémillard and Theodorescu introduce the operator $\Psi : C_0 \to C[0,1]$,

$$\Psi(h)(t) = \sum_{k=0}^{\infty} h\big(F(k)\big)t^k(1-t)\,, \qquad 0 \le t \le 1\,, \tag{4.27}$$

and they show that Ψ is well-defined and continuous on C_0. Since we have $\gamma_n = \Psi(\beta_n)$ and $Y = \Psi(B)$, and β_n converges in distribution to the Brownian bridge B, the authors conclude that γ_n converges weakly to Y. The mistake in this argument is obvious: the process β_n is not continuous, so it does not lie in the space C_0. Hence, the weak convergence of β_n cannot be mapped to that of $\Psi(\beta_n)$ if Ψ is only defined on C_0.

The oversight can be corrected if we extend Ψ properly and prove a more general continuity property for it. Let $D[0,1]$ denote the space of all càdlàg functions defined on $[0,1]$ endowed with the Skorohod metric. Under the generated topology the space $D[0,1]$ is complete and separable, as described in Chapter 3 of Billingsley (1968). Let D_0 stand for the subspace of those elements of $D[0,1]$ which are continuous and vanish at the point 1, and define the operator Ψ on D_0 with the formula (4.27). First, we show that Ψ maps into $C[0,1]$, it is measurable on D_0 and it is continuous on $C_0 \subseteq D_0$ with respect to the Skorohod topology.

Consider any element h of the subspace D_0. Since h is continuous and vanishes at the point 1, the sequence $h(F(k))$, $k = 0, 1, \ldots$, is bounded and converges to 0. This implies that the series $\Psi(h)(t)$ is absolute convergent for any $0 \le t < 1$, and hence, the image $\Psi(h)$ is continuous on the interval $[0,1)$. We show that the continuity holds also at the point 1. For any fixed $\varepsilon > 0$ we can consider a positive integer m such that

$$\sup_{F(m) \le u \le 1} |h(u)| < \varepsilon/2\,.$$

Then, we obtain the inequality

$$\begin{aligned}
\big|\Psi(h)(t)\big| &\le \sum_{k=0}^{m-1} \big|h\big(F(k)\big)\big|(1-t)t^k + \sum_{k=m}^{\infty} \big|h\big(F(k)\big)\big|(1-t)t^k \\
&\le \sup_{0 \le u \le 1} |h(u)|(1-t)\sum_{k=0}^{m-1} t^k + \sup_{F(m) \le u \le 1} |h(u)|(1-t)\sum_{k=m}^{\infty} t^k \\
&\le \sup_{0 \le u \le 1} |h(u)|(1-t)\sum_{k=0}^{m-1} 1 + \sup_{F(m) \le u \le 1} |h(u)|(1-t)\frac{t^m}{1-t} \\
&= m(1-t)\sup_{0 \le u \le 1} |h(u)| + \sup_{F(m) \le u \le 1} |h(u)| \le \varepsilon/2 + \varepsilon/2\,,
\end{aligned} \tag{4.28}$$

for any $0 \le t \le 1$ satisfying the condition

$$t > 1 - \frac{\varepsilon}{2m \sup_{0\le u\le 1} |h(u)|}\,.$$

Hence, $\Psi(h)$ is an element of the space $C[0,1]$ for any $h \in D_0$.

Consider functions $h \in C_0$ and $h_n \in D_0$, $n = 1, 2, \ldots$, such that $h_n \to h$ with respect to the Skorohod topology. Since the convergence to an element of the space $C[0,1]$ in the Skorohod metric is equivalent with the convergence in the supremum distance, it follows that h_n converges uniformly to h on the interval $[0,1]$. Using the linearity of the operator Ψ inequality (4.28) implies that

$$\begin{aligned}\limsup_{0\le t\le 1} \big|\Psi(h_n)(t) - \Psi(h)(t)\big| &= \limsup_{0\le t\le 1} \big|\Psi(h_n - h)(t)\big| \\ &\le m \lim_{n\to\infty} \sup_{0\le u\le 1} \big|h_n(u) - h(u)\big| + \frac{\varepsilon}{2} = \frac{\varepsilon}{2}\,.\end{aligned}$$

Because ε is an arbitrary positive value, we obtain that $\Psi(h_n)$ converges uniformly to $\Psi(h)$, and hence, Ψ is continuous on the subspace $C_0 \subseteq D_0$. (Note that the argument is delicate in that Ψ is not continuous at some functions in D_0. See the counterexample after the proof.)

Next, we show that the terms in the representation of Ψ in (4.27) are all measurable operators $D_0 \to C[0,1]$, which implies that Ψ is also measurable. The inverse image of an open ball in $C[0,1]$ with center at the function φ and radius ε given by the k-th term of the series in (4.27) is

$$\Big\{h \in D_0 : \sup_{0\le t\le 1} \big|h\big(F(k)\big)t^k(1-t) - \varphi(t)\big|\Big\} = \bigcap_{t\in\mathbb{Q}\cap(0,1)} A_t(\varphi,\varepsilon) \cap D_0\,, \tag{4.29}$$

where for a fixed $t \in \mathbb{Q}\cap(0,1)$ the set $A_t(\varphi,\varepsilon)$ is defined by

$$A_t(\varphi,\varepsilon) = \left\{h \in D[0,1] : \frac{\varphi(t)-\varepsilon}{t^k(1-t)} < h\big(F(k)\big) < \frac{\varphi(t)+\varepsilon}{t^k(1-t)}\right\}.$$

While $A_t(\varphi,\varepsilon)$ itself is generally not a Skorohod open set, it is measurable. For this goal consider the projection

$$\pi_u : D[0,1] \to \mathbb{R}\,, \qquad \pi_u(g) = g(u)\,,$$

with any $0 \le u \le 1$, and note that π_u is a measurable function by the argument on page 121 of Billingsley (1968). This implies that the inverse image

$$A_t(\varphi,\varepsilon) = \pi^{-1}_{F(k)}\left(\frac{\varphi(t)-\varepsilon}{t^k(1-t)}\,,\ \frac{\varphi(t)+\varepsilon}{t^k(1-t)}\right)$$

is a measurable subset of $D[0,1]$. Using the standard definition of left-continuity the set D_0 can represented by the form

$$D_0 = \bigcap_{\varepsilon\in\mathbb{Q}^+} \bigcup_{\delta\in\mathbb{Q}^+} \bigcap_{\substack{t\in\mathbb{Q}\\ 1-\delta<t\le 1}} \pi_t^{-1}(-\varepsilon,\varepsilon)\,,$$

which is measurable, as well. Hence, we conclude that the inverse image in (4.29) is a measurable set, and the measurability of the operator $\Psi : D_0 \to C[0,1]$ follows.

Since the operator Ψ is continuous on the subspace $C_0 \subseteq D_0$ with respect to the Skorohod metric, and the uniform empirical process converges weakly to the Brownian bridge B in the Skorohod topology, and the process B lies in C_0 with probability 1, Corollary 1 to Theorem 1.5.1 in Billingsley (1968) implies the weak convergence of the image $\gamma_n = \Psi(\beta_n)$ to $Y = \Psi(B)$ in the space $C[0,1]$ as $n \to \infty$. □

It is a crucial point in the proof of Theorem 4.17 that the operator Ψ is continuous on C_0 in the Skorohod topology. It is a naturally arising question if one can extend this property for the whole subspace D_0 or Ψ has some points of discontinuity in the domain. Let

$$k_1 = \min\big\{k \in \mathbb{N} : P(X = k) > 0\big\}$$

be the smallest value of the non-negative integer valued random variables X. Also, this is the smallest jumping point of the related distribution function F. We show that Ψ is continuous on the whole domain D_0 if and only if the background variable is degenerate, that is, $P(X = k_1) = 1$.

If X is degenerate then the distribution function F vanishes on $(-\infty, k_1)$ and it is equal to 1 on the interval $[k_1, \infty)$. In this case Ψ can be written in the form

$$\Psi(h)(t) = \sum_{k=0}^{\infty} h\big(F(k)\big)(1-t)t^k = \sum_{k=k_1}^{\infty} h(1)(1-t)t^k = h(1)t^{k_1}\,, \qquad 0 \le t \le 1\,.$$

By considering functions $h, h_n \in D_0$, $n = 1, 2, \ldots$ such that $h_n \to h$ in the Skorohod topology, the basic properties of the Skorohod metric implies that $h_n(1) \to h(1)$, and the sequence $\Psi(h_n)$ converges to $\Psi(h)$ in $C[0,1]$. That is, Ψ is continuous on D_0.

On the contrary, assume that X is not degenerate, and let k_2 denote the second smallest value of X, that is, let

$$k_2 = \big\{k > k_1 : P(X = k_2) > 0\big\}\,.$$

Consider the functions

$$h(u) = \mathbb{1}_{[0,F(k_1))}(u)\,, \qquad h_n(u) = \mathbb{1}_{[0,F(k_1)+1/n)}(u)\,, \qquad 0 \le u \le 1\,.$$

where the notation $\mathbb{1}$ stands for the indicator function of the interval marked in the lower index. It is clear that h and h_n lie in D_0 for every n and $h \to h_n$ in the Skorohod metric. The image of the function h under the operator Ψ is

$$\Psi(h)(t) = \sum_{k=0}^{\infty} h\big(F(k)\big)t^k(1-t) = \sum_{k=0}^{k_1-1} t^k(1-t) = 1 - t^{k_1}\,, \qquad 0 \le t \le 1\,,$$

and if the integer n is large enough to satisfy the inequality $F(k_1) + 1/n < F(k_2)$ then

$$\Psi(h_n)(t) = \sum_{k=0}^{\infty} h_n\big(F(k)\big)t^k(1-t) = \sum_{k=0}^{k_2-1} t^k(1-t) = 1 - t^{k_2}\,, \qquad 0 \le t \le 1\,.$$

Since $k_2 > k_1$, the sequence $\Psi(h_n)$ does not converge to $\Psi(h)$ in $C[0,1]$, which implies that h is a point of discontinuity in the domain of Ψ.

We note that the weak convergence of γ_n to the process Y in $C[0,1]$ can also be proved in the usual way, by showing that the finite dimensional distributions converge and the sequence γ_n is tight. However, in comparison to the elegant idea of Rémillard and Theodorescu (2000), the classical method is much more complicated since the proof of the tightness requires a massive amount of calculations.

4.6 Strong approximations for the generating process

In this section we provide a uniform strong approximation for the empirical probability generating process γ_n and its derivatives, and based on this result we investigate the asymptotic behavior of some functionals of the processes.

Let $X_1, X_2, \dots$ be independent non-negative valued variables with the same distribution function $F(x)$, $x \in \mathbb{R}$, and for an arbitrary fixed $r = 0, 1, \dots$ consider the r-th derivative $\gamma_n^{(r)}$ of the probability generating process and the corresponding Gaussian process Y_r represented by formulas (4.16) and (4.18). Note that the processes are well-defined and sample-continuous on the interval $[a, b]$ provided by Proposition 4.10, and in the case $r = 0$ they exist also at $t = 0$ and $t = 1$ for any X.

Throughout the section we prove our statements by applying a suitable representation of the processes on the KMT probability space. Using the independent and uniform variables $U_1, U_2, \dots$ provided by Theorem 2.1 consider the random values

$$X_i = F^{-1}(U_i)\,, \qquad i = 1, 2, \dots\,,$$

with the quantile function F^{-1} of the distribution function F defined as

$$F^{-1}(u) = \inf\big\{x \in \mathbb{R} : F(x) \geq u\big\}\,, \qquad 0 < u < 1\,.$$

In Section 2.1 we found that using this representation the related empirical process can be written in the form

$$\alpha_n(x) = \beta_n\big(F(x)\big)\,, \qquad x \in \mathbb{R}\,,$$

with the uniform empirical process $\beta_n(u)$, $0 \leq u \leq 1$, appearing in Theorem 2.1. To highlight the fact that this construction takes place on a unique probability space we use a different notation for the corresponding empirical probability generating process and for its derivatives. The representation of $\gamma_n^{(r)}$ based on the variables $X_1, \dots, X_n$ is marked by $\gamma_{r,n}$. Then, using the general formula (4.16) we have the identity

$$\gamma_{r,n}(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, d\beta_n\big(F(x)\big)\,. \tag{4.30}$$

Also, let us define the independent and identically distributed variables $X_1', X_2', \dots$ as

$$X_i' = F^{-1}(U_i')\,, \qquad i = 1, 2, \dots \tag{4.31}$$

by applying the uniform variables $U_1', U_2', \ldots$ of the second part of Theorem 2.1. If β_n' stands for the uniform empirical process of the sample $U_1', \ldots, U_n'$ as in the referred theorem, then the empirical process and the derivative $\gamma_{r,n}' = \gamma_n^{(r)}$ based on $X_1', \ldots, X_n'$ can be written in the forms

$$\alpha_n'(x) = \beta_n'\big(F(x)\big)\,, \qquad \gamma_{r,n}'(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\,d\beta_n'\big(F(x)\big)\,. \quad (4.32)$$

Again, we use the notation $\gamma_{r,n}'$ instead of $\gamma_n^{(r)}$ to point out that the process is constructed on the KMT space.

Let $B_n(u)$, $n = 1, 2, \ldots$, be the sequence of Brownian bridges and let $K(u, y)$ be the Kiefer process provided by Theorem 2.1, $0 \le u \le 1$, $y \ge 0$, and define

$$Y_{r,n}(t) = \int_{\mathbb{R}} x(x-r)\cdots(x-r+1)t^{x-r}\,dB_n\big(F(x)\big)\,, \qquad 0 \le t \le 1\,,$$

and

$$Y_{r,n}'(t) = n^{-1/2}\int_{\mathbb{R}} x(x-r)\cdots(x-r+1)t^{x-r}\,dK\big(F(x), n\big)\,, \qquad 0 \le t \le 1\,. \quad (4.33)$$

Since for any fixed integer n the processes $B_n(u)$ and $n^{-1/2}K(u, n)$, $0 \le u \le 1$, are Brownian bridges, the integrals $Y_{r,n}(t)$ and $Y_{r,n}'(t)$ exist exactly at those points $t \in \mathbb{R}$ whereat $Y_r(t)$ is well-defined. That is, $Y_{r,n}$ and $Y_{r,n}'$ are finite and have sample-continuous modifications on the interval $[a, b]$ defined in Proposition 4.10. For simplicity, in the following $Y_{r,n}$ and $Y_{r,n}'$ stand for these modifications.

Our next results are uniform strong approximations for the processes $\gamma_{r,n}$ and $\gamma_{r,n}'$. Recall that for a sequence of random variables $V_1, V_2, \ldots$ and a sequence of positive constants $a_1, a_2, \ldots$ we write $V_n = \mathcal{O}(a_n)$, $n \to \infty$, if there exists a universal real constant C, not depending on the underlying distributions, such that

$$\limsup_{n\to\infty} \big|V_n/a_n\big| \le C\,, \qquad n \to \infty\,, \qquad \text{a.s.}$$

Theorem 4.18. *Consider the distribution function $F(x)$, $x \in \mathbb{R}$, of an arbitrary non-negative valued random variable X. Then, for the processes defined on the KMT space we have*

$$\sup_{0\le t\le 1} \big|\gamma_{0,n}(t) - Y_{0,n}(t)\big| = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right) \quad \textit{and} \quad \sup_{0\le t\le 1} \big|\gamma_{0,n}'(t) - Y_{0,n}'(t)\big| = \mathcal{O}\left(\frac{\log^2 n}{n^{1/2}}\right).$$

Also, for any $r = 1, 2, \ldots$ and $0 < \varepsilon < 1/2$ it holds that

$$\sup_{\varepsilon\le t\le 1-\varepsilon} \big|\gamma_{r,n}(t) - Y_{r,n}(t)\big| = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right) \quad \textit{and} \quad \sup_{\varepsilon\le t\le 1-\varepsilon} \big|\gamma_{r,n}'(t) - Y_{r,n}'(t)\big| = \mathcal{O}\left(\frac{\log^2 n}{n^{1/2}}\right).$$

Furthermore, if X has only non-negative integer values, then with any $r = 0, 1, \ldots$ and $0 < \tau < 1$ we obtain

$$\sup_{-\tau\le t\le \tau} \big|\gamma_{r,n}(t) - Y_{r,n}(t)\big| = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right) \quad \textit{and} \quad \sup_{-s\le t\le s} \big|\gamma_{r,n}(t)' - Y_{r,n}'(t)\big| = \mathcal{O}\left(\frac{\log^2 n}{n^{1/2}}\right).$$

All statements are understood almost surely as $n \to \infty$.

Proof. By Proposition 4.10 the processes on the left sides of the equations exist at every point of the intervals where we consider their supremums. Using the representations (4.30) and (4.32) we have the identities

$$\gamma_{r,n}(t) - Y_{r,n}(t) = \int_{\mathbb{R}} x(x-r)\cdots(x-r+1)t^{x-r}\, d\Big[\beta_n\big(F(x)\big) - B_n\big(F(x)\big)\Big]$$

and

$$\gamma'_{r,n}(t) - Y'_{r,n}(t) = \int_{\mathbb{R}} x(x-r)\cdots(x-r+1)t^{x-r}\, d\Big[\beta'_n\big(F(x)\big) - n^{-1/2}K\big(F(x),n\big)\Big]\,.$$

Applying our Theorem 4.7 with the process

$$K_n(x) = \alpha_n(x) - B_n\big(F(x)\big) = \beta_n\big(F(x)\big) - B_n\big(F(x)\big)\,, \qquad x\in\mathbb{R}\,,$$

and by the first formula of Theorem 2.2 we obtain the desired rate

$$\sup_{0\le t\le 1}\big|\gamma_{0,n}(t) - Y_{0,n}(t)\big| \le \sup_{x\in\mathbb{R}}\big|K_n(x)\big| \le \sup_{0\le u\le 1}\big|\beta_n(u) - B_n(u)\big| = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right).$$

The third and the fifth equations of the statement can be shown similarly by using Theorems 4.7 and 4.9, respectively. Also, with the choice

$$K'_n(x) = \alpha'_n(x) - n^{-1/2}K\big(F(x),n\big) = \beta'_n\big(F(x)\big) - n^{-1/2}K\big(F(x),n\big)\,, \qquad x\in\mathbb{R}\,,$$

Theorem 4.7 and the second formula of Theorem 2.2 imply

$$\sup_{0\le t\le 1}\big|\gamma'_{0,n}(t) - Y'_{0,n}(t)\big| \le \sup_{x\in\mathbb{R}}\big|K'_n(x)\big| \le \sup_{0\le u\le 1}\big|\beta'_n(u) - n^{-1/2}K(u,n)\big| = \mathcal{O}\left(\frac{\log^2 n}{n^{1/2}}\right).$$

One can prove the remaining equations just the same way by applying Theorems 4.7 and 4.9, again. □

As a direct consequence of Theorem 4.18 we obtain the weak convergence of the process γ_n and its derivatives. We note that, as a special case, the following statement also contains Theorem 4.17, the result of Rémillard and Theodorescu (2000).

Corollary 4.19. *Consider an arbitrary non-negative integer valued random variable X defined on any probability space. The process $\gamma_n^{(r)}$ converges in distribution to Y_r in the space $C[a,b]$ as $n\to\infty$, where the interval $[a,b]$ is provided by Proposition 4.10.*

Proof. By Proposition 4.10 the processes $\gamma_n^{(r)}$ and Y_r lie in the space $C[a,b]$. Theorem 4.18 implies that the supremum distance of the processes $\gamma_{r,n}$ and $Y_{r,n}$ constructed on the KMT space converges to 0 almost surely. Since the representations $\gamma_{r,n}$ and $Y_{r,n}$ have the same distribution in $C[a,b]$ as the general versions $\gamma_n^{(r)}$ and Y_r, respectively, we obtain the weak convergence. □

A typical application of the weak convergence of the probability generating process γ_n and its derivatives is testing the fit of a given sample $X_1, \dots, X_n$ to some hypothetic distribution. Let $g_{\mathcal{H}}^{(r)}$ stand for the r-th derivative of the probability generating function $g_{\mathcal{H}}$ corresponding to the hypothetic distribution and consider the process

$$\gamma_{n,\mathcal{H}}^{(r)}(t) = n^{1/2}\big[g_n^{(r)}(t) - g_{\mathcal{H}}^{(r)}(t)\big]\,, \qquad a \le t \le b\,.$$

Observe that $\gamma_{n,\mathcal{H}}^{(r)}$ is obtained from the formula (4.16) of $\gamma_n^{(r)}$ by replacing the unknown theoretical probability generating function g with $g_{\mathcal{H}}$. If we consider some continuous functional ψ on the space $C[a,b]$, then $\psi(\gamma_{n,\mathcal{H}}^{(r)})$ serves as a test statistic, and we can obtain critical values by determining the theoretical quantiles of the variable $\psi(Y_r)$. The theoretical basis of the method is the fact that if the hypothetic probability generating function is the true one, then the statistics $\psi(\gamma_{n,\mathcal{H}}^{(r)}) = \psi(\gamma_n^{(r)})$ converges in distribution to $\psi(Y_r)$ as $n \to \infty$. In the applications it is a crucial question that how fast the quantiles of the statistics $\psi(\gamma_n^{(r)})$ converge to those of $\psi(Y_r)$. The next statement provides an answer for this problem, a uniform rate of convergence.

Theorem 4.20. *Consider a non-negative valued variable X and an integer $r = 0, 1, \dots$ Let ψ denote a functional on the space $C[a,b]$ satisfying the Lipschitz condition*

$$\big|\psi(h_1) - \psi(h_2)\big| \le M \sup_{a \le u \le b} \big|h_1(u) - h_2(u)\big|\,, \qquad h_1, h_2 \in C[a,b]\,,$$

with some finite positive constant M, and assume that $\psi(Y_r)$ has bounded density function on $[s,\infty)$ with some $s \in \mathbb{R}$. Then we have

$$\sup_{x \ge s} \Big|P\big(\psi(\gamma_n^{(r)}) \le x\big) - P\big(\psi(Y_r) \le x\big)\Big| = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right).$$

Proof. We adapt the ideas of Komlós, Major and Tusnády (1975) which were applied in the proof of the related theorem for the uniform empirical process. (See Theorem 2.4 in our Section 2.1.) Since the statement is a property of the distributions of $\gamma_n^{(r)}$ and Y_r in the space $C[a,b]$, we can prove the theorem by using a suitable representation of the processes. Of course, we will work with the versions $\gamma_{r,n}$ and $Y_{r,n}$ constructed on the KMT space in Section 4.6. First, we need a stronger version of the approximations presented in our Theorem 4.18. By considering the process

$$K_n(x) = \beta_n\big(F(x)\big) - B_n\big(F(x)\big)\,, \qquad x \in \mathbb{R}\,,$$

Proposition 4.10 with Theorems 4.7 and 4.9 implies that

$$\sup_{a \le t \le b} \big|\gamma_{r,n}(t) - Y_{r,n}(t)\big| \le C \sup_{x \in \mathbb{R}} |K_n(x)| \le C \sup_{0 \le u \le 1} \big|\beta_n(u) - B_n(u)\big| \tag{4.34}$$

with a positive constant $C = C(a,b,r)$. From the Lipschitz property of the functional ψ we obtain the almost sure inequality

$$\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| \le M \sup_{a \le t \le b} \big|\gamma_{r,n}(t) - Y_{r,n}(t)\big| \le CM \sup_{0 \le u \le 1} \big|\beta_n(u) - B_n(u)\big|\,. \tag{4.35}$$

Consider the strictly positive constants c_1, c_2 and c_3 of Theorem 2.1, and let

$$y_n = \frac{2}{c_3}\log n \qquad \text{and} \qquad \varepsilon_n = CM\frac{y_n + c_1 \log n}{n^{1/2}} = \mathcal{O}\left(\frac{\log n}{n^{1/2}}\right).$$

The first statement of Theorem 2.1 with (4.35) implies that

$$\begin{aligned} P\Big(\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| > \varepsilon_n\Big) &\le P\left(\sup_{0\le u\le 1}\big|\beta_n(u) - B_n(u)\big| > \frac{y_n + c_1\log n}{n^{1/2}}\right) \\ &\le c_2 \exp(-c_3 y_n) = c_2/n^2\,. \end{aligned}$$

Let M_2 stand for a positive bound on the density function of $\psi(Y_r)$ on the interval $[s,\infty)$, and observe that for any fixed $x \ge s$ we have

$$\Big\{\psi(\gamma_{r,n}) \le x\Big\} \subseteq \Big\{\psi(Y_{r,n}) \le x + \varepsilon_n\Big\} \bigcup \Big\{\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| > \varepsilon_n\Big\}.$$

Since the functional $\psi(Y_{r,n})$ has the same distribution as $\psi(Y_r)$ for every $n = 1, 2, \ldots$ it follows that

$$\begin{aligned} P\Big(\psi(\gamma_{r,n}) \le x\Big) &\le P\Big(\psi(Y_{r,n}) \le x + \varepsilon_n\Big) + P\Big(\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| > \varepsilon_n\Big) \\ &\le P\Big(\psi(Y_r) \le x\Big) + P\Big(x < \psi(Y_r) \le x + \varepsilon_n\Big) + c_2/n^2 \\ &\le P\Big(\psi(Y_r) \le x\Big) + \varepsilon_n M_2 + c_2/n^2\,. \end{aligned}$$

Hence, we get the inequality

$$P\big(\psi(\gamma_{r,n}) \le x\big) - P\big(\psi(Y_r) \le x\big) \le \varepsilon_n M_2 + c_2/n^2\,.$$

We can easily obtain a similar lower bound for the difference, as well. From the formula

$$\Big\{\psi(Y_{r,n}) \le x - \varepsilon_n\Big\} \subseteq \Big\{\psi(\gamma_{r,n}) \le x\Big\} \bigcup \Big\{\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| > \varepsilon_n\Big\}.$$

it follows that

$$P\Big(\psi(Y_{r,n}) \le x - \varepsilon_n\Big) \le P\Big(\psi(\gamma_{r,n}) \le x\Big) + P\Big(\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| > \varepsilon_n\Big).$$

By rearranging the terms in the inequality we have

$$\begin{aligned} P\Big(\psi(\gamma_{r,n}) \le x\Big) &\ge P\Big(\psi(Y_{r,n}) \le x - \varepsilon_n\Big) - P\Big(\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| > \varepsilon_n\Big) \\ &\ge P\Big(\psi(Y_r) \le x\Big) - P\Big(x - \varepsilon_n < \psi(Y_r) \le x\Big) - c_2/n^2 \\ &\ge P\Big(\psi(Y_r) \le x\Big) - \varepsilon_n M_2 - c_2/n^2\,, \end{aligned}$$

which leads to

$$P\big(\psi(Y_r) \le x\big) - P\big(\psi(\gamma_{r,n}) \le x\big) \le \varepsilon_n M_2 + c_2/n^2\,.$$

Since the term $\varepsilon_n M_2$ has a slower rate of convergence than c_2/n^2 has, we get that

$$\Big|P\big(\psi(\gamma_{r,n}) \le x\big) - P\big(\psi(Y_r) \le x\big)\Big| \le \varepsilon_n M_2 + c_2/n^2 = \mathcal{O}\bigg(\frac{\log n}{n^{1/2}}\bigg)\,,$$

and the proof is complete. □

The best known and most important application of Theorem 4.20 is that when ψ is a supremum functional on $C[a,b]$, that is, one of

$$\psi_0(h) = \sup_{a\le t\le b} |h(t)|\,, \qquad \psi_+(h) = \sup_{a\le t\le b} h(t) \qquad \text{and} \qquad \psi_-(h) = -\inf_{a\le t\le b} h(t)\,.$$

Observe that with any functions $h_1, h_2 \in C[a,b]$ we have

$$\big|\psi_+(h_1) - \psi_+(h_2)\big| \le \big|\psi_0(h_1) - \psi_0(h_2)\big| \le \sup_{a\le u\le b} \big|h_1(u) - h_2(u)\big|\,,$$

and a similar inequality holds for ψ_-. Hence, the functionals ψ_0, ψ_+ and ψ_- satisfy the Lipschitz condition with $M = 1$. Also, in Proposition 4.14 we showed that the functionals $\psi_0(Y_r)$, $\psi_+(Y_r)$ and $\psi_-(Y_r)$ have bounded densities on certain subintervals of the real line. As a result we can apply Theorem 4.20 for these functionals and we obtain the uniform approximations presented in Corollary 4.21.

Corollary 4.21. *Consider the values $s_r^+ \le 0$ and $s_r^- \le 0$ of Proposition 4.14, and let $s_1 > 0$, $s_2 > s_r^+$ and $s_3 > s_r^-$ be arbitrary. Then, we have*

$$\sup_{x\ge s_1} \bigg|P\bigg(\sup_{a\le t\le b} \big|\gamma_n^{(r)}(t)\big| \le x\bigg) - P\bigg(\sup_{a\le t\le b} \big|Y_r(t)\big| \le x\bigg)\bigg| = \mathcal{O}\bigg(\frac{\log n}{n^{1/2}}\bigg)\,,$$

and the distribution function of

$$\sup_{a\le t\le b} \gamma_n^{(r)}(t) \qquad \textit{and} \qquad -\inf_{a\le t\le b} \gamma_n^{(r)}(t)$$

converge with the same rate to that of

$$\sup_{a\le t\le b} Y_r(t) \qquad \textit{and} \qquad -\inf_{a\le t\le b} Y_r(t)$$

on the intervals $[s_2,\infty)$ and $[s_3,\infty)$, respectively.

In the last result of this section we investigate the Cramér–von Mises type statistics of the process $\gamma_n^{(r)}$, and we show that its distribution function converges uniformly to the distribution function of the related statistics based on the limiting process Y_r. In the proof we apply a simplified version of the method which was invented to prove the corresponding statement for the uniform empirical process β_n by Csörgő (1976). Our theorem provides a rate of convergence $\mathcal{O}(n^{-1/2}(\log n)^{3/2})$, and it is a crucial point of the justification that the functional

$$\psi(Y_r) = \int_a^b Y_r^2(t)\,dt$$

has a bounded density function $f(x)$, $x \in \mathbb{R}$, by Corollary 4.16. Note that if one can show that $f(x)x^{1/2}$ is bounded, too, then using the original technique of Csörgő (1976), the rate $\mathcal{O}(n^{-1/2}\log n)$ can be achieved.

Theorem 4.22. *We have*

$$\sup_{x\in\mathbb{R}}\left|P\Big(\int_a^b \big(\gamma_n^{(r)}\big)^2(t)\,dt \le x\Big) - P\Big(\int_a^b Y_r^2(t)\,dt \le x\Big)\right| = \mathcal{O}\Big(\frac{(\log n)^{3/2}}{n^{1/2}}\Big)\,.$$

Proof. If the variable X is degenerate or $P(X \in \{0,\dots,r-1\}) = 1$ then $\gamma_n^{(r)}$ and Y_r vanish on the interval $[a,b]$ by formula (4.18) and Proposition 4.13. Hence, the square integrals of the processes are equal to 0 with probability one, and the statement follows. In the remaining we assume that X is not degenerate and $P(X \in \{0,\dots,r-1\}) < 1$.

Consider the processes β_n, B_n, $\gamma_{r,n}$ and $Y_{r,n}$, $n = 1,2,\dots$, defined on the KMT probability space, and let

$$\psi(\gamma_{r,n}) = \int_a^b \gamma_{r,n}^2(t)\,dt\,, \qquad \psi(Y_{r,n}) = \int_a^b Y_{r,n}^2(t)\,dt\,,$$

and

$$S_n = \sup_{0\le u\le 1}|B_n(u)|\,, \qquad T_n = \sup_{0\le u\le 1}\big|B_n(u) - \beta_n(u)\big|\,.$$

Depending on which case of Proposition 4.10 holds for the variable X and the integer r, by using Theorem 4.7 or Theorem 4.9 we obtain the inequality

$$\sup_{a\le t\le b}|Y_{r,n}(t)| \le C\sup_{x\in\mathbb{R}}\big|B_n\big(F(x)\big)\big| \le CS_n\,, \qquad \text{a.s.}$$

and (4.34) provides the approximation

$$\sup_{a\le t\le b}\big|\gamma_{r,n}(t) - Y_{r,n}(t)\big| \le C\sup_{0\le u\le 1}\big|\beta_n(u) - B_n(u)\big| = CT_n\,, \qquad \text{a.s.}$$

with a positive constant $C = C(a,b,r)$. By applying the elementary properties of the supremum functional we obtain the almost sure bound on the distance of $\psi(\gamma_{r,n})$ and $\psi(Y_{r,n})$ in the form

$$\begin{aligned}
\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| &\le \int_a^b \big|\gamma_{r,n}^2(t) - Y_{r,n}^2(t)\big|\,dt \le (b-a)\sup_{a\le t\le b}\big|\gamma_{r,n}^2(t) - Y_{r,n}^2(t)\big| \\
&\le (b-a)\sup_{a\le t\le b}\big|\gamma_{r,n}(t) - Y_{r,n}(t)\big|\sup_{a\le t\le b}\big|\gamma_{r,n}(t) + Y_{r,n}(t)\big| \\
&\le (b-a)\sup_{a\le t\le b}\big|\gamma_{r,n}(t) - Y_{r,n}(t)\big|\left[\sup_{a\le t\le b}\big|\gamma_{r,n}(t) - Y_{r,n}(t)\big| + 2\sup_{a\le t\le b}\big|Y_{r,n}(t)\big|\right] \\
&\le (b-a)C^2T_n\big(T_n + 2S_n\big)\,.
\end{aligned} \tag{4.36}$$

Consider the constants c_1, c_2 and c_3 provided by Theorem 2.1 and the sequences

$$y_n = \frac{2}{c_3}\log n\,, \qquad \varepsilon_n = \frac{y_n + c_1\log n}{n^{1/2}} \qquad \text{and} \qquad \delta_n = 3(b-a)C^2\varepsilon_n(\log n)^{1/2}\,.$$

Applying the first statement of Theorem 2.1 we obtain the inequality

$$P\big(T_n > \varepsilon_n\big) = P\bigg(\sup_{0\le u\le 1}\big|\beta_n(u) - B_n(u)\big| > \frac{y_n + c_1\log n}{n^{1/2}}\bigg) \le c_2\exp(-c_3 y_n) = c_2/n^2\,.$$

Also, by using the the well-known result of Smirnov for the supremum of the Brownian bridge (formula (1.5.3) in Csörgő and Révész (1981)) it follows that

$$P\big(S_n \ge (\log n)^{1/2}\big) = P\bigg(\sup_{0\le u\le 1}|B_n(u)| \ge (\log n)^{1/2}\bigg) \le 2\exp(-2\log n) = 2/n^2\,.$$

If n is large enough to satisfy the inequality $\varepsilon_n \le (\log n)^{1/2}$, then from (4.36) we get

$$\begin{aligned} P\Big(\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| \ge \delta_n, T_n \le \varepsilon_n\Big) &\le P\Big(T_n(T_n + 2S_n) \ge 3\varepsilon_n(\log n)^{1/2}, T_n \le \varepsilon_n\Big) \\ &\le P\Big(\varepsilon_n(\varepsilon_n + 2S_n) \ge 3\varepsilon_n(\log n)^{1/2}\Big) \le P\big(S_n \ge (\log n)^{1/2}\big) \le 2/n^2\,. \end{aligned}$$

Hence, we have

$$\begin{aligned} &P\Big(\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| \ge \delta_n\Big) \\ &\quad\le P\Big(\big|\psi(\gamma_{r,n}) - \psi(Y_{r,n})\big| \ge \delta_n, T_n \le \varepsilon_n\Big) + P\big(T_n > \varepsilon_n\big) \le (c_2+2)/n^2\,. \end{aligned}$$

Note that by Corollary 4.16 there exists a bound M on the common density function of the variables $\psi(Y_{r,n})$. One can complete the proof by applying the same method as we used in the second part in the proof of Theorem 4.20. Consider an arbitrary real x. Using the formula

$$\big\{\psi(\gamma_{r,n}) \le x\big\} \subseteq \big\{\psi(Y_{r,n}) \le x + \delta_n\big\} \cup \big\{|\psi(\gamma_{r,n}) - \varphi_n| \ge \delta_n\big\}\,,$$

one can obtain the inequality

$$P\big(\psi(\gamma_{r,n}) \le x\big) \le P\big(\psi(Y_{r,n}) \le x\big) + \delta_n M + (c_2+2)/n^2\,,$$

and

$$\big\{\psi(Y_{r,n}) \le x - \delta_n\big\} \subseteq \big\{\psi(\gamma_{r,n}) \le x\big\} \cup \big\{|\psi(\gamma_{r,n}) - \psi(Y_{r,n})| \ge \delta_n\big\}\,,$$

implies that

$$P\big(\psi(\gamma_{r,n}) \le x\big) \ge P\big(\psi(Y_{r,n}) \le x\big) - \delta_n M - (c_2+2)/n^2\,,$$

just as in the proof of Theorem 4.20. From these we have

$$\sup_{x>0}\Big|P\big(\psi(\gamma_{r,n}) \le x\big) - P\big(\psi(Y_{r,n}) \le x\big)\Big| \le \delta_n M + \frac{c_2+2}{n^2} = \mathcal{O}(\delta_n) = \mathcal{O}\bigg(\frac{(\log n)^{3/2}}{n^{1/2}}\bigg)\,,$$

since δ_n has a slower rate of convergence than the term $(c_2+2)/n^2$ has. □

4.7 Law of the iterated logarithm

We continue our investigation on the empirical probability generating process by working in the same framework as in the previous section. The variables $X, X_1, X_2, \ldots$ are independent and non-negative valued with the same distribution function $F(x)$. Also, $\alpha_n(x)$, $x \in \mathbb{R}$, and $\gamma_n^{(r)}(t)$, $a \le t \le b$, are the empirical process and the r-th derivative of the generating process γ_n being written up based on the sample $X_1, \ldots, X_n$. The interval $[a, b]$ is provided by Proposition 4.10. Also, let $\beta_n(u)$, $0 \le u \le 1$, stand for the uniform empirical process corresponding to the first n elements in the sequence of independent random variables $U_1, U_2, \ldots$ distributed uniformly on the interval $[0, 1]$, and let $K(u, y)$, $0 \le u \le 1$, $y \ge 0$, denote the Kiefer process represented on an arbitrary probability space.

The famous Smirnov–Chung law of the iterated logarithm (Theorem 5.1.1 in Csörgő and Révész (1981)) states that

$$\limsup_{n\to\infty} \frac{\sup_{0\le u\le 1} |\beta_n(u)|}{(\log\log n)^{1/2}} = \frac{1}{2^{1/2}} \qquad \text{a.s.,} \tag{4.37}$$

and hence, the supremum of the uniform empirical process has the rate

$$\sup_{0\le u\le 1} |\beta_n(u)| = \mathcal{O}\Big((\log\log n)^{1/2}\Big)\,, \qquad n \to \infty\,.$$

This result can be generalized for a variable X having an arbitrary distribution. Simply consider the quantile function $F^{-1}(u)$, $0 \le u \le 1$, corresponding to the distribution function F, and define the sample variables by the form

$$X_i = F^{-1}(U_i)\,, \qquad i = 1, 2, \ldots$$

Just as in the previous sections the distributed function of the constructed X_i's is F, and their empirical process can be written as

$$\alpha_n(x) = \beta_n\big(F(x)\big)\,, \qquad x \in \mathbb{R}\,.$$

This and the special form of the law of the iterated logarithm presented in (4.37) imply

$$\limsup_{n\to\infty} \frac{\sup_{x\in\mathbb{R}} |\alpha_n(x)|}{(\log\log n)^{1/2}} \le \limsup_{n\to\infty} \frac{\sup_{0\le u\le 1} |\beta_n(u)|}{(\log\log n)^{1/2}} = \frac{1}{2^{1/2}} \qquad \text{a.s.,} \tag{4.38}$$

and we have equation if the distribution function F is continuous everywhere on the real line. Since this inequality is a property of the distribution of the empirical process α_n in the space $D[-\infty, \infty]$, the rate (4.38) is valid in case of any representation of the variables $X_1, X_2, \ldots$ and the process α_n. Also, based on Section 1.15 of Csörgő and Révész (1981) one can show that the "normalized" Kiefer process $y^{-1/2}K(u, y)$, $0 \le u \le 1$, has the same rate of convergence as $y \to \infty$, that is,

$$\limsup_{y\to\infty} \frac{\sup_{0\le u\le 1} |K(u, y)|}{(y\log\log y)^{1/2}} = \frac{1}{2^{1/2}} \qquad \text{a.s.} \tag{4.39}$$

The first goal of this section is to present a law of the iterated logarithm for the empirical probability generating process γ_n and its derivatives. Also, we investigate a process which provides copies of the limiting process $Y_r(t)$, $a \le t \le b$, $n = 1, 2, \dots$. Let $K(u, y)$ stand for the Kiefer process and let

$$Y_r(t, y) = n^{-1/2} \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, K\big(dF(x), y\big)\,, \tag{4.40}$$

for $a \le t \le b$ and $y > 0$. Fix an arbitrary value $y > 0$. Since $y^{-1/2}K(u, y)$, $0 \le u \le 1$, is a Brownian bridge, the integral $Y_r(t, y)$ is defined exactly at those points where $Y_r(t)$ exists. Using Propositions 4.10 and 4.11 we get that $Y_r(t, y)$ is a well-defined Gaussian process on $[a, b]$ and it has a sample-continuous modification in variable t on this interval. In the following the notation $Y_r(t, y)$ stand for this modification. Of course, in general $Y_r(t, y)$ is not continuous in the parameter y, but we will not need this property at all. Also, in the case $r = 0$ the process $Y_r(t, y)$ is defined at $t = 0$ and $t = 1$, as well.

Note that if we consider $Y_r(t, y)$ based on the Kiefer process provided on the KMT space by Theorem 2.1, then for any positive integer n we have

$$Y_r(t, n) = Y'_{r,n}(t)\,, \qquad a \le t \le b\,,$$

since the right side is defined by the form (4.33). The examination of the process is motivated by the fact that using the covariance structure of $Y_r(t, n)$ one can investigate the dependence between the identically distributed but not independent processes $Y'_{r,n}$, $n = 1, 2, \dots$ We can state the following law of the iterated logarithm for the introduced processes.

Theorem 4.23. *For any non-negative valued random variable X we have*

$$\limsup_{n\to\infty} \frac{\sup_{0\le t\le 1} |\gamma_n(t)|}{(\log\log n)^{1/2}} = \limsup_{y\to\infty} \frac{\sup_{0\le t\le 1} |Y_0(t, y)|}{(\log\log y)^{1/2}} \le \frac{1}{2^{1/2}} \qquad a.s.$$

Also, with any $r = 1, 2 \dots$ and $0 < \varepsilon < 1/2$ it follows that

$$\limsup_{n\to\infty} \frac{\sup_{\varepsilon\le t\le 1-\varepsilon} |\gamma_n^{(r)}(t)|}{(\log\log n)^{1/2}} = \limsup_{y\to\infty} \frac{\sup_{\varepsilon\le t\le 1-\varepsilon} |Y_r(t, y)|}{(\log\log y)^{1/2}} \le \frac{C_1(\varepsilon, r)}{2^{1/2}} \qquad a.s.$$

Furthermore, if X is non-negative integer valued then with an arbitrary $r = 0, 1, \dots$ and $0 < \tau < 1$ we have

$$\limsup_{n\to\infty} \frac{\sup_{-\tau\le t\le \tau} |\gamma_n^{(r)}(t)|}{(\log\log n)^{1/2}} = \limsup_{y\to\infty} \frac{\sup_{-\tau\le t\le \tau} |Y_r(t, y)|}{(\log\log y)^{1/2}} \le \frac{C_2(\tau, r)}{2^{1/2}} \qquad a.s.$$

The constants $C_1(\varepsilon, r)$ and $C_2(\tau, r)$ are defined in Theorems 4.7 and 4.9, and they are independent from the distribution of the variable X.

Proof. Since the statements correspond to the distribution of the underlying processes we can work with any suitable representation of them. Let us consider the variables $X_1', X_2', \dots$, and the related empirical process $\alpha_n'(x)$, $x \in \mathbb{R}$, and the generating process $\gamma_{r,n}'(t)$, $a \le t \le b$, defined on the KMT space by formulas (4.31) and (4.32). Also, let us define $Y_r'(t,y)$ with equation (4.40) by using the Kiefer process of the KMT space.

Note that all of the processes are well-defined on the intervals where we consider their supremums. Using the first part of Theorem 4.7 we get that

$$\sup_{0\le t\le 1} |Y_0'(t,y)| \le \sup_{x\in\mathbb{R}} \left|y^{-1/2} K\big(F(x),y\big)\right| \le y^{-1/2} \sup_{0\le u\le 1} |K(u,y)|\,,$$

and the law of the iterated logarithm for the Kiefer process in (4.39) implies that

$$\limsup_{y\to\infty} \frac{\sup_{0\le t\le 1} |Y_0'(t,y)|}{(\log\log y)^{1/2}} \le \limsup_{y\to\infty} \frac{\sup_{0\le u\le 1} |K(u,y)|}{(y\log\log y)^{1/2}} = \frac{1}{2^{1/2}}\,. \tag{4.41}$$

From the identity $Y_0'(t,n) = Y_{0,n}'(t)$ and Theorem 4.18 it follows that

$$\sup_{0\le t\le 1} \left|\gamma_{0,n}'(t) - Y_0'(t,n)\right| \to 0\,, \qquad n\to\infty\,, \qquad \text{a.s.}$$

Also, for every $n = 1, 2, \dots$ we have

$$|\gamma_{0,n}'(t)| = \left|Y_0'(t,n) + \big(\gamma_{0,n}'(t) - Y_0'(t,n)\big)\right| \le |Y_0'(t,n)| + \left|\gamma_{0,n}'(t) - Y_0'(t,n)\right|, \tag{4.42}$$

and we get that

$$\begin{aligned} \limsup_{n\to\infty} \frac{\sup_{0\le t\le 1} |\gamma_{0,n}'(t)|}{(\log\log n)^{1/2}} &\le \limsup_{n\to\infty} \frac{\sup_{0\le t\le 1} |Y_0'(t,n)|}{(\log\log n)^{1/2}} \\ &+ \lim_{n\to\infty} \frac{\sup_{0\le t\le 1} |\gamma_{0,n}'(t) - Y_0'(t,n)|}{(\log\log n)^{1/2}} = \limsup_{n\to\infty} \frac{\sup_{0\le t\le 1} |Y_0'(t,n)|}{(\log\log n)^{1/2}} + 0\,. \end{aligned} \tag{4.43}$$

By changing the roles of $\gamma_{0,n}'(t)$ and $Y_0'(t,n)$ in formulas (4.42) and (4.43) we obtain that there is equation in (4.43), and the first statement of the theorem is showed.

One can prove the second and the third statement with the change that in formula (4.41) one must apply the second part of Theorem 4.7 and Theorem 4.9. □

Corollary 4.24. *Working on the interval $[a,b]$ provided by Proposition 4.10 we have*

$$\sup_{a\le t\le b} \left|\gamma_n^{(r)}(t)\right| = \sup_{a\le t\le b} \left|Y_r(t,n)\right| = \mathcal{O}\Big((\log\log n)^{1/2}\Big)\,.$$

Also, for the representations $\gamma_{n,r}$ and $Y_{r,n}$ defined on the KMT space we have the rates

$$\sup_{a\le t\le b} \left|\gamma_{n,r}^2(t) - Y_{r,n}^2(t)\right| = \mathcal{O}\left(\frac{\log n(\log\log n)^{1/2}}{n^{1/2}}\right)$$

and

$$\left|\int_a^b \gamma_{n,r}^2(t)\,dt - \int_a^b Y_{r,n}^2(t)\,dt\right| = \mathcal{O}\left(\frac{\log n(\log\log n)^{1/2}}{n^{1/2}}\right).$$

Proof. The first statement follows from Theorem 4.23. By using Theorem 4.18 and the method of the previous proof we have

$$\begin{aligned}\sup_{a\le t\le b}\bigl|\gamma_{n,r}^2(t)-Y_{r,n}^2(t)\bigr| &\le \sup_{a\le t\le b}\bigl|\gamma_{n,r}(t)-Y_{r,n}(t)\bigr| \sup_{a\le t\le b}\bigl|\gamma_{n,r}(t)+Y_{r,n}(t)\bigr| \\ &\le \sup_{a\le t\le b}\bigl|\gamma_{n,r}(t)-Y_{r,n}(t)\bigr|\Bigl[\sup_{a\le t\le b}\bigl|\gamma_{n,r}(t)-Y_{r,n}(t)\bigr|+2\sup_{a\le t\le b}\bigl|\gamma_{n,r}(t)\bigr|\Bigr] \\ &= \mathcal{O}\Bigl(\frac{\log n}{n^{1/2}}\Bigr)\Bigl[\mathcal{O}\Bigl(\frac{\log n}{n^{1/2}}\Bigr)+\mathcal{O}\Bigl((\log\log n)^{1/2}\Bigr)\Bigr] = \mathcal{O}\Bigl(\frac{\log n(\log\log n)^{1/2}}{n^{1/2}}\Bigr).\end{aligned}$$

From this we obtain the rate

$$\begin{aligned}\Bigl|\int_a^b \gamma_{n,r}^2(t)\,dt-\int_a^b Y_{r,n}^2(t)\,dt\Bigr| &\le \int_a^b\bigl|\gamma_{n,r}^2(t)-Y_{r,n}^2(t)\bigr|\,dt \\ &\le (b-a)\sup_{a\le t\le b}\bigl|\gamma_{n,r}^2(t)-Y_{r,n}^2(t)\bigr| = \mathcal{O}\Bigl(\frac{\log n(\log\log n)^{1/2}}{n^{1/2}}\Bigr).\end{aligned}$$

□

We saw it in Theorem 4.19 that $\gamma_n^{(r)}$ converges in distribution to the process Y_r in the space $C[a,b]$, but these results do not say anything about the limiting properties for a fixed outcome ω of the underlying probability space. It is known that the uniform empirical process β_n may be convergent only on an event having probability 0, because it is relative compact. We will see that $\gamma_n^{(r)}$ behaves similarly.

Consider random elements $\xi_1,\xi_2,\dots$ having values in some metric space $\mathbb{X}$. We say that the sequence $\xi_1,\xi_2,\dots$ is relative compact in the space $\mathbb{X}$ with limit set $\mathbb{Y}\subseteq\mathbb{X}$ if there exists an event Ω_0 having probability 1 on the underlying probability space such that for every $\omega\in\Omega_0$ the following three conditions hold.

1. Every subsequence $\xi_{n'}(\omega)$ of $\xi_n(\omega)$ has a further convergent subsequence $\xi_{n''}(\omega)$.
2. If a subsequence $\xi_{n'}(\omega)$ of $\xi_n(\omega)$ converges then the limit lies in $\mathbb{Y}$.
3. For each element $y\in\mathbb{Y}$ there exists a subsequence $\xi_{n'}(\omega)$ of $\xi_n(\omega)$ depending on the outcome ω such that $\xi_{n'}(\omega)$ converges to y.

By the famous result of Finkelstein (1971) the uniform empirical process $\beta_n(u)$, $0\le u\le 1$, is relative compact in the space $D[0,1]$ with respect to the Skorohod metric. The limit points are those functions $h\in D[0,1]$, which are absolute continuous, vanish at the points 0 and 1, and have Radon–Nikodym derivative $h'(u)$, $0\le u\le 1$, with respect to the Lebesgue measure such that

$$\int_0^1\bigl(h'(u)\bigr)^2\,du\le 1\,.$$

Let $C_\beta\subseteq D[0,1]$ denote the set of limit points, and for a fixed $r=0,1,\dots$ consider

$$\begin{aligned}C_r[a,b] &= \Bigl\{\int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\,dh\bigl(F(x)\bigr)\,,a\le t\le b : h\in C_\beta\Bigr\} \\ &= \Bigl\{\int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}h'\bigl(F(x)\bigr)\,dF(x)\,,a\le t\le b : h\in C_\beta\Bigr\}.\end{aligned}$$

Our result for the process $\gamma_n^{(r)}$ is the following.

Theorem 4.25. *For an arbitrary non-negative valued random variable X and integer $r = 0, 1, \ldots$ the process $\gamma_n^{(r)}(t)$, $a \le t \le b$, is relative compact in the space $C[a,b]$ and the set of limit points is $C_r[a,b]$.*

Proof. We apply the variables $X_1, \ldots, X_n$ with which we worked in the proof of Theorem 4.23. By the referred result of Finkelstein (1971) the sequence β_n constructed on the KMT space is relative compact, and the set of the limit points is D_β. Let Ω_0 stand for the corresponding event in the definition of relative compactness. We prove that for any outcome $\omega \in \Omega_0$ the sequence $\gamma_{r,n}(\omega)$ satisfies all of the three conditions of relative compactness with the limit set $C_r[a,b]$.

First, we show that if an arbitrary subsequence $\beta_{n''}(\omega)$ of $\beta_n(\omega)$ converges to an element h of the limit set C_β, then $\gamma_{r,n''}(\omega)$ converges to

$$\Psi_r(h)(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dh\big(F(x)\big)\,, \qquad a \le t \le b\,.$$

It is important to note that the convergence of $\beta_{n''}(\omega)$ is understood in the Skorohod topology, but since the limit function h is continuous on the interval $[a,b]$ we also have uniform convergence. Introducing the function

$$K_{n''}(x) = \beta_{n''}\big(F(x),\omega\big) - h\big(F(x)\big)\,, \qquad x \in \mathbb{R}\,,$$

we obtain that

$$\sup_{x\in\mathbb{R}} |K_{n''}(x)| \le \sup_{x\in\mathbb{R}} \big|\beta_{n''}(u,\omega) - h(u)\big| \to 0\,, \qquad n'' \to \infty\,.$$

Observe that under the fixed outcome ω the function

$$\gamma_{r,n''}(t,\omega) - \Psi_r(h)(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dK_{n''}(x)\,, \qquad a \le t \le b\,,$$

is not random. Depending on which case of Proposition 4.10 holds for the given variable X and integer r and provides us the interval $[a,b]$, we can use the first or the second part of Theorem 4.7 or Theorem 4.9. We get that

$$\sup_{a\le t\le b} \big|\gamma_{r,n''}(t,\omega) - \Psi_r(h)(t)\big| \le \sup_{x\in\mathbb{R}} |K_{n''}(x)| \to 0\,, \qquad n'' \to \infty\,.$$

Hence, $\gamma_{r,n''}(\omega)$ converges in the space $C[a,b]$ to the element $\Psi_r(h)$.

For the first condition consider a subsequence n' of the positive integers $n = 1, 2, \ldots$ Using the relative compactness of the process β_n there exists a further subsequence n'' such that $\beta_{n''}(\omega)$ converges to some $h \in C_\beta$. Using the previous paragraph we obtain that $\gamma_{r,n''}(\omega)$ also converges.

Now, assume that the subsequence $\gamma_{r,n'}(\omega)$ converges to some function φ in the space $C[a,b]$. Since β_n is relative compact we have a subsequence n'' of n' such that

$\beta_{n''}(\omega)$ converges to an element h of C_β, which implies that $\gamma_{r,n''}(\omega)$ converges to $\Psi_r(h)$. Since n'' is a subsequence of n' we get that $\varphi = \Psi_r(h)$, and hence, φ is an element of $C_r[a,b]$.

Finally, consider any function $\Psi_r(h)$ in the set $C_r[a,b]$. Since h is a limit point of the relative compact process β_n, there is a subsequence n' for which $\beta_{n'}(\omega)$ converges to the function h. With the sequence $n'' = n'$ we obtain the convergence of $\gamma_{r,n'}(\omega)$ to $\Psi_r(h)$. This completes the proof of relative compactness. □

4.8 Confidence bands

The empirical and the theoretical probability generating function based on the first n elements of the sequence of independent variables $X_1, X_2, \ldots$ having only non-negative values and common distribution function $F(x)$, $x \in \mathbb{R}$, were defined as

$$g_n(t) = \frac{1}{n}\sum_{i=1}^{n} t^{X_i} = \int_{\mathbb{R}} t^x \, dF_n(x) \qquad \text{and} \qquad g(t) = E(t^X) = \int_{\mathbb{R}} t^x \, dF(x)\,,$$

where $F_n(x)$ is the corresponding empirical distribution function and X has the same distribution as the X_i's. By the strong law of large numbers the function g_n converges pointwisely to g with probability 1 at every point t where the variable t^X exists and has finite mean. As we saw it in Section 4.3 this condition is satisfied for every $-1 < t < 1$. The strong uniform consistency of the empirical probability generating function, as an estimator of its theoretical counterpart, on the interval $[0,1]$ for any integer valued variable X was already proved by Marques and Pérez-Abreu (1989) and Rémillard and Theodorescu (2000). Note that one can obtain this result also by using the inequality of Theorem 4.9 and the Glivenko–Cantelli theorem. However, thanks to our work in the previous section, we already have a much stronger result for g_n, and also, for its derivatives. Applying Corollary 4.24 we obtain a uniform rate of convergence

$$\sup_{a \le t \le b} \left| g_n^{(r)}(t) - g^{(r)}(t) \right| = n^{-1/2} \sup_{a \le t \le b} \left| \gamma_n^{(r)}(t) \right| = \mathcal{O}\Big(n^{-1/2} (\log\log n)^{1/2} \Big)\,,$$

where the interval $[a,b]$ comes from Proposition 4.10. Note that if X is integer valued and $r = 0$ then we have the option to choose $a = 0$ and $b = 1$.

The main goal if this section is to construct asymptotically correct confidence bands for the probability generating function. For this, motivated by Efron (1979) and Csörgő and Mason (1989), we apply the bootstrap technique. For a fixed n consider a positive integer m_n and the Efron type bootstrapped variables $X_{1,n}^*, \ldots, X_{m_n,n}^*$ based on the sample $X_1, \ldots, X_n$ as described in Section 2.2. That is, choose values among $X_1, \ldots, X_n$ with replacement m_n times such that at each selection each X_i has the same probability of being chosen. Then, we can introduce the bootstrap empirical process

$$\alpha_{m_n,n}^*(x) = m_n^{1/2} \big[F_{m_n,n}^*(x) - F_n(x) \big]\,, \qquad x \in \mathbb{R}\,,$$

by applying the empirical distribution function $F^*_{m_n,n}$ of the bootstrapped sample. Observe that conditionally under the original observations $X_1, \dots, X_n$ the function F_n is the theoretical distribution function of the bootstrapped variables, and hence, $\alpha^*_{m_n,n}$ is the conditional empirical process. Then, the empirical probability generating function and the conditional theoretical probability generating function of the $X^*_{i,n}$'s can be written as

$$g^*_{m_n,n}(t) = \frac{1}{n}\sum_{i=1}^{m_n} t^{X^*_{i,n}} = \int_{\mathbb{R}} t^x \, F^*_{m_n,n}(x) \qquad \text{and} \qquad g_n(t) = \int_{\mathbb{R}} t^x \, dF_n(x)\,. \tag{4.44}$$

By using Propositions 4.2 and 4.8 and Theorem 4.9 for $g^*_{m_n,n}$ just the same way as we applied these results for g_n in Section 4.3, one can show that the function $g^*_{m_n,n}$ is well-defined and has continuous r-th derivative

$$\begin{aligned} g^{*(r)}_{m_n,n}(t) &= \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dF^*_{m_n,n}(x) \\ &= \frac{1}{n}\sum_{i=1}^{m_n} X^*_{i,n}\big(X^*_{i,n}-1\big)\cdots\big(X^*_{i,n}-r+1\big)t^{X^*_{i,n}-r} \end{aligned}$$

on the interval $[a,b]$ provided by Proposition 4.10. Also, we can define the bootstrap empirical probability generating process

$$\gamma^*_{m_n,n}(t) = n^{1/2}\big[g^*_{m_n,n}(t) - g_n(t)\big] = \int_{\mathbb{R}} t^x \, \alpha^*_{m_n,n}(x)\,,$$

and we find that the process exists and has continuous r-th derivative

$$\gamma^{*(r)}_{m_n,n}(t) = n^{1/2}\big[g^{*(r)}_{m_n,n}(t) - g^{(r)}_n(t)\big] = \int_{\mathbb{R}} x(x-r)\cdots(x-r+1)t^x \, \alpha^*_{m_n,n}(x)$$

on the same $[a,b]$. By the bootstrap heuristics we expect that the process $\gamma^{*(r)}_{m_n,n}$ has the same asymptotic behavior, and hence, it has the same weak limit in the space $C[a,b]$ as its non-bootstrapped counterpart $\gamma^{(r)}_n$ has. It turns out in the next result that this is true under a slight condition for the sequence m_n, $n = 1, 2, \dots$

Theorem 4.26. *Consider the distribution function $F(x)$, $x \in \mathbb{R}$, of an arbitrary non-negative valued random variable, and assume that there exist positive constants C_1 and C_2 such that*

$$C_1 < m_n/n < C_2\,, \qquad n = 1, 2, \dots$$

On a sufficiently rich probability space one can define independent random variables $X_1, X_2, \dots$ having common distribution function $F(x)$, $x \in \mathbb{R}$, and bootstrapped sample variables $X^*_{1,n}, \dots, X^*_{m_n,n}$, $n = 1, 2, \dots$ based on the X_i's, and copies $Y^*_{r,1}, Y^*_{r,2}, \dots$ of the process Y_r, such that

$$\sup_{a\le t\le b} \big|\gamma^{*(r)}_{m_n,n}(t) - Y^*_{r,m_n}(t)\big| = \mathcal{O}\big(\max\{l(m_n), l(n)\}\big)\,,$$

with the function $l(n) = n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}$ *and with the interval* $[a, b]$ *provided by Proposition 4.10. Furthermore, the processes* $Y^*_{r,1}, Y^*_{r,2}, \dots$ *are independent from the sequence* $X_1, X_2, \dots$

Proof. Consider the variables $X_1, X_2, \dots$, the bootstrapped samples $X^*_{1,n}, \dots, X^*_{m_n,n}$, $n = 1, 2, \dots$, and the Brownian bridges $B^*_1, B^*_2, \dots$ of Theorem 2.5. Also, consider the bootstrap empirical process $\alpha^*_{m_n,n}(x)$, $x \in \mathbb{R}$, and the bootstrap empirical probability generating process $\gamma^*_{m_n,n}(t)$, $a \le t \le b$, based on these variables. Let

$$Y^*_{r,m}(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dB^*_m\big(F(x)\big)\,, \qquad a \le t \le b\,,$$

and

$$K_n(x) = \alpha^*_{m_n,n}(x) - B^*_{m_n}\big(F(x)\big)\,, \qquad x \in \mathbb{R}\,.$$

Since in Theorem 2.5 the Brownian bridges are independent from the X_i's, the same holds for the processes $Y^*_{r,1}, Y^*_{r,2}, \dots$ Depending on which case of Proposition 4.10 gives us the interval $[a, b]$ the first or the second part of Theorem 4.7 or Theorem 4.9 implies that

$$\sup_{a\le t\le b} \big|\gamma^{*(r)}_{m_n,n}(t) - Y^*_{r,m_n}(t)\big| \le C \sup_{x\in\mathbb{R}} |K_n(x)|\,,$$

with some constant $C = C(a, b, r)$. By Theorem 2.5 the right side has the desired rate of convergence. □

Corollary 4.27. *If there exist positive constants* C_1 *and* C_2 *such that*

$$C_1 < m_n/n < C_2\,, \qquad n = 1, 2, \dots,$$

then the process $\gamma^{*(r)}_{m_n,n}$ *converges weakly to* Y_r *in the space* $C[a, b]$ *as* $n \to \infty$.

To construct confidence bands for the probability generating function we use the method of Csörgő and Mason (1989). We note that they illustrated their receipt by applying it on the infamous horsekick sample of von Bortkiewitz, as well. We say that on a given significance level $0 < \alpha < 1$ the sequence of positive values $c'_n(\alpha)$ provides an asymptotically correct confidence band for g on the interval $[a, b]$ if we have

$$P\Big(g_n(t) - c'_n(\alpha) \le g(t) \le g_n(t) + c'_n(\alpha), a \le t \le b\Big) \to 1 - \alpha\,, \tag{4.45}$$

as the sample size $n \to \infty$. By introducing the functional

$$\psi : C[a, b] \to \mathbb{R}\,, \qquad \psi(h) = \sup_{a\le t\le b} |h(t)|\,,$$

and the sequence $c_n(\alpha) = n^{-1/2}c'_n(\alpha)$, formula (4.45) can be written in the form

$$P\Big(\sup_{a\le t\le b} \big|g_n(t) - g(t)\big| \le c'_n(\alpha)\Big) = P\Big(\psi(\gamma_n) \le c_n(\alpha)\Big) \to 1 - \alpha\,. \tag{4.46}$$

The goal is to provide an estimation of $c_n(\alpha)$ based on the sample variables $X_1, \dots, X_n$. Observe that this is a special case of the problem faced in formula (2.5) and we find a possible solution for it at the end of Section 2.2.

Introduce the variables $\psi^*_{m_n,n} = \psi(\alpha^*_{m_n,n})$ and $\varphi = \psi(Y)$ with the process

$$Y(t) = Y_0(t) = \int_{\mathbb{R}} t^X \, dF(x)\,, \qquad a \le t \le b\,.$$

Since $\alpha^*_{m_n,n}$ converges to Y in distribution and the functional ψ is continuous on the space $C[a,b]$, we immediately get that $\psi^*_{m_n,n}$ converges weakly to φ as $n \to \infty$. However, we can obtain a much stronger result, a strong approximation for $\psi^*_{m_n,n}$, as well. Using the representation of the processes $\gamma^*_{m_n,n}$ and Y^*_n provided by Theorem 4.26 define the copies $\varphi_n = \psi(Y^*_n)$, $n = 1, 2, \dots$ of the variable φ. Since the supremum functional ψ satisfies the Lipschitz condition

$$\big|\psi(h_1) - \psi(h_2)\big| \le \sup_{a \le t \le b} \big|h_1(t) - h_2(t)\big|\,, \qquad h_1, h_2 \in C[a,b]\,,$$

we get that

$$\big|\psi^*_{m_n,n} - \varphi_{m_n}\big| = \big|\psi(\gamma^*_{m_n,n}) - \psi(Y^*_{m_n})\big| \le \sup_{a \le t \le b} \big|\gamma^*_{m_n,n}(t) - Y^*_{m_n}(t)\big| \to 0$$

almost surely as $n \to \infty$. Also, in the current representation the variables $\varphi_1, \varphi_2, \dots$ are independent from the sequence $X_1, X_2, \dots$

Of course, our aim with the representation is to apply Theorem 2.7. For this end we only need to check the last assumption of the theorem, the continuity of the distribution function $F_\varphi(x)$, $x \in \mathbb{R}$, of the variable φ. Observe that the distribution function F_ϕ is continuous on the real line by Proposition 4.21. Using this with Theorem 2.7 and the discussion at the end of Section 2.2, we find that the quantile

$$c^*_n(\alpha) = \inf\Big\{x \in \mathbb{R} : P\big(\psi^*_{m_n,n} \le x \mid X_1, \dots, X_n\big) \ge 1 - \alpha\Big\}$$

can play the role of $c_n(\alpha)$ in (4.46).

From the practical point of view, one can determine $c^*_n(\alpha)$ by direct calculations. Since conditionally under the sample $X_1, \dots, X_n$ the statistics $\psi^*_{m_n,n}$ has finitely many possible values, using combinatorical arguments one can obtain its exact conditional distribution. Also, the quantile $c^*_n(\alpha)$ can be estimated in arbitrary precision by using a computer. Generating a large number of bootstrapped samples and calculating the corresponding statistics $\psi^*_{m_n,n}$ for each sample, one can obtain a good empirical estimation for the conditional distribution of $\psi^*_{m_n,n}$ under the given observations $X_1, \dots, X_n$. Then, the empirical quantiles of $\psi^*_{m_n,n}$ provide a good estimation for $c^*_n(\alpha)$, and hence, for the desired value $c_n(\alpha)$.

4.9 Parameter estimated generating processes

Consider a family $\mathcal{F} = \{F(x,\theta) : x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^d\}$ of non-negative valued univariate distributions, and also, a sample of non-negative values $X_1, \dots, X_n$ having empirical

distribution function $F_n(x)$ and unknown theoretical distribution function $F(x)$, $x \in \mathbb{R}$. Our aim is to test the fit of the sample to the family $\mathcal{F}$, that is, the null-hypotheses

$$\mathcal{H}_0 : \; F(x) = F(x, \theta_0)\,, \; x \in \mathbb{R}\,, \qquad \text{for some fixed } \theta_0 \in \Theta\,.$$

As we sketched it in Section 3.1, one can test $\mathcal{H}_0$ by applying a proper measurable functional $\psi(\hat{\alpha}_n)$ of the parameter estimated empirical process

$$\hat{\alpha}_n(x) = n^{1/2}\big[F_n(x) - F(x, \hat{\theta}_n)\big]\,, \qquad x \in \mathbb{R}\,,$$

where $\hat{\theta}_n$ is a parameter estimation based on the sample $X_1, \ldots, X_n$. The theoretical base of the method is that under $\mathcal{H}_0$ and the regularity conditions of Theorem 3.1 the process $\hat{\alpha}_n$ converges weakly in the space $D[-\infty, \infty]$ to the centered Gaussian process

$$G(x) = B\big(F(x, \theta_0)\big) - \left[\int_{\mathbb{R}} l(x, \theta_0)\, dB\big(F(x, \theta_0)\big)\right] \nabla_\theta F(x, \theta_0)^T\,, \qquad x \in \mathbb{R}\,. \tag{4.47}$$

In this formula $B(u)$, $0 \le u \le 1$, is a Brownian bridge, the function $l(x, \theta_0)$, $x \in \mathbb{R}$, is defined by assumption (a4) in Section 3.2, and

$$\nabla_\theta F(x, \theta) = \left(\frac{\partial}{\partial \theta^{(1)}} F(x, \theta), \ldots, \frac{\partial}{\partial \theta^{(d)}} F(x, \theta)\right), \qquad x \in \mathbb{R}\,. \tag{4.48}$$

Unfortunately, there are distribution families whose parametric distribution function $F(x, \theta)$, $x \in \mathbb{R}$, $\theta \in \Theta \subseteq \mathbb{R}^d$, are not provided in simple forms, and by this reason the application of the parameter estimated empirical process $\hat{\alpha}_n$ can be difficult. However, in many cases the probability generating process can be written in friendly formulas. For example, see the discrete stable, Linnik and Sibuya distributions in Rémillard and Theodorescu (2000). Note that we can face this difficulty also in case of more common distributions. Since the distribution function $F(x, \theta)$ of a discrete variable is a step function, in certain applications the use of the continuous probability generating function can be favorable. By this motivation we define the parameter estimated probability generating process

$$\hat{\gamma}_n(t) = n^{1/2}\big[g_n(t) - g(t, \hat{\theta}_n)\big] = \int_{\mathbb{R}} t^x \, d\hat{\alpha}_n(x)\,, \tag{4.49}$$

where

$$g_n(t) = \frac{1}{n}\sum_{i=1}^{n} t^{X_i} = \int_{\mathbb{R}} t^x \, dF_n(x) \qquad \text{and} \qquad g(t, \theta) = \int_{\mathbb{R}} t^x \, dF(x, \theta)$$

are the empirical probability generating function based on the sample $X_1, \ldots, X_n$ and the parametric generating function of the family $\mathcal{F}$, respectively. Later in this section we show that under some conditions $\hat{\gamma}_n$ converges in distribution to

$$\hat{Y}(t) = \int_{\mathbb{R}} t^x \, dG(x)\,, \tag{4.50}$$

and hence, one can apply the process to test $\mathcal{H}_0$ in just the same way as the parameter estimated empirical process $\hat{\alpha}_n$ can be used.

It is important to observe that the application of the process $\hat{\gamma}_n$ can lead to an other difficulty. As we detailed in Section 3.1, in most cases the critical values corresponding to a test statistics $\psi(\hat{\alpha}_n)$ can not be determined by theoretical calculations, and these statistics usually are not distribution free, either. Since the limit process $\hat{Y}$ is defined as an integral transformation of G, the same problems can arise for a functional $\psi(\hat{\gamma}_n)$. As a possible solution, we introduce the bootstrapped versions of the parameter estimated probability generating process. For this end, consider a bootstrap sample size m_n and parametric or non-parametric bootstrapped variables $X^*_{1,n}, \dots, X^*_{m_n,n}$ as described in Section 2.2. Let $F^*_{m_n,n}$ stand for the empirical distribution function of the bootstrapped sample, and consider the corresponding empirical probability generating function

$$g^*_{m_n,n}(t) = \frac{1}{m_n} \sum_{i=1}^{m_n} t^{X^*_{i,n}} = \int_{\mathbb{R}} t^x \, dF_{m_n,n}(x) \, .$$

Also, consider parameter estimators $\hat{\theta}^*_n$ and $\tilde{\theta}^*_n$ based on the bootstrapped variables in the parametric and the non-parametric bootstrap case, respectively. The parametric bootstrap estimated generating process can be defined as

$$\hat{\gamma}^*_{m_n,n}(t) = n^{1/2} \big[g^*_{m_n,n}(t) - g(t, \hat{\theta}^*_n) \big] = \int_{\mathbb{R}} t^x \, d\hat{\alpha}^*_{m_n,n}(x) \, , \tag{4.51}$$

and its non-parametric variant is

$$\tilde{\gamma}^*_{m_n,n}(t) = n^{1/2} \big[g^*_{m_n,n}(t) - g(t, \tilde{\theta}^*_n) \big] = \int_{\mathbb{R}} t^x \, d\tilde{\alpha}^*_{m_n,n}(x) \, , \tag{4.52}$$

where $\hat{\alpha}^*_{m_n,n}$ and $\tilde{\alpha}^*_{m_n,n}$ are introduced by (3.3) and (3.4) in Section 3.1. If we can show that under $\mathcal{H}_0$ the processes $\hat{\gamma}^*_{m_n,n}$ and/or $\tilde{\gamma}^*_{m_n,n}$ converge to $\hat{Y}$ then we can apply the parametric and/or the non-parametric bootstrap method to obtain critical values for $\psi(\hat{\alpha}_n)$ by using a simple variant of the algorithm presented in Section 3.4.

Consider a sequence of independent variables $X_1, X_2, \dots$ having common distribution function $F(x, \theta_0)$, $x \in \mathbb{R}$, with some fixed $\theta_0 \in \Theta$. First, we answer the question which points $t \in \mathbb{R}$ are the corresponding generating processes defined at. It is clear that the empirical functions $g_n(t)$ and $g^*_{m_n,n}(t)$ exists and are continuous on the positive half-line. Since the distribution function $F(x, \theta)$, $x \in \mathbb{R}$, has bounded increments and finite limit at infinity for every $\theta \in \Theta$, Proposition 4.2 implies that $g(t, \hat{\theta}_n)$, $g(t, \hat{\theta}^*_n)$ and $g(t, \tilde{\theta}^*_n)$ are well-defined and continuous on $[0, 1]$. Then, clearly, the processes $\hat{\gamma}_n(t)$, $\hat{\gamma}^*_{m_n,n}(t)$ and $\tilde{\gamma}^*_{m_n,n}(t)$ exist and are sample-continuous on this interval, as well. Furthermore, if the family $\mathcal{F}$ contains only non-negative integer valued distributions then $g_n(t)$ and $g^*_{m_n,n}(t)$ can be extended to the real line. Since in the integer valued case Proposition 4.8 implies that $g(t, \hat{\theta}_n)$, $g(t, \hat{\theta}^*_n)$ and $g(t, \tilde{\theta}^*_n)$ are well-defined and continuous on the interval $(-1, 1]$, the same holds for the processes $\hat{\gamma}_n(t)$, $\hat{\gamma}^*_{m_n,n}(t)$ and $\tilde{\gamma}^*_{m_n,n}(t)$.

Assume that the components of the vector function $\nabla_\theta F(x,\theta_0)$ are of bounded variation on finite intervals, and let $\nabla_k^+(y,\theta_0)$ and $\nabla_k^-(y,\theta_0)$ be defined as the positive and the negative variation of the k-th component $\partial F(x,\theta)/\partial\theta_k$ at $\theta=\theta_0$ on the interval $(-\infty,y]$. Also, let $l_k(x,\theta)$ denote the k-th component of the function $l(x,\theta)$, $x\in\mathbb{R}$, and consider the d-dimensional normal variable

$$\xi=(\xi_1,\dots,\xi_d)=\int_{\mathbb{R}} l(x,\theta_0)\,dB\big(F(x,\theta_0)\big)\,.$$

Observe that the process

$$A(x)=\xi\nabla_\theta F(x,\theta_0)^T=\sum_{k=1}^d \xi_k\big[\nabla_k^+(x,\theta_0)-\nabla_k^-(x,\theta_0)\big]\,,\qquad x\in\mathbb{R}\,.$$

is of bounded variation on every finite interval, and its positive and negative variation on $(-\infty,y]$ can be written as

$$A^+(x)=\sum_{k=1}^d \xi_k\Big[\mathbb{1}_{\{\xi_k\ge 0\}}\nabla_k^+(x,\theta_0)+\mathbb{1}_{\{\xi_k<0\}}\nabla_k^-(x,\theta_0)\Big]\,,\qquad x\in\mathbb{R}\,,$$

and

$$A^-(x)=\sum_{k=1}^d \xi_k\Big[\mathbb{1}_{\{\xi_k\ge 0\}}\nabla_k^-(x,\theta_0)+\mathbb{1}_{\{\xi_k<0\}}\nabla_k^+(x,\theta_0)\Big]\,,\qquad x\in\mathbb{R}\,,$$

respectively. Because the distribution function $F(x,\theta)$ vanishes on the interval $(-\infty,0)$ for every $\theta\in\Theta$, the derivative $\nabla_\theta F(x,\theta_0)$ and its positive and negative variation $\nabla_k^+(x,\theta_0)$ and $\nabla_k^-(x,\theta_0)$ are constant 0 on the negative half-line. This implies that the processes $A(x)$, $A^+(x)$ and $A^-(x)$ vanish on $(-\infty,0)$. Furthermore, if the components of $\nabla_\theta F(x,\theta_0)$, $x\in\mathbb{R}$, are càdlàg, then clearly $A(x)$, $A^+(x)$ and $A^-(x)$ are càdlàg, too. In the remaining part of the section we work under the following assumption.

Assumption 2. The function $\nabla_\theta F(x,\theta_0)$, $x\in\mathbb{R}$, exists and its components are càdlàg and of bounded variation on finite intervals. Also, with some fixed $\delta>0$ the sequences

$$\nabla_k^+\big((m+1)\delta,\theta_0\big)-\nabla_k^+\big(m\delta,\theta_0\big)\,,\quad \nabla_k^-\big((m+1)\delta,\theta_0\big)-\nabla_k^-\big(m\delta,\theta_0\big)\,,\quad m=0,1,\dots\,,$$

are bounded for every $k=1,\dots,d$.

Note that Assumption 2 holds for any $\delta>0$ if the function $\nabla_\theta F(x,\theta)$, $x\in\mathbb{R}$, has bounded components. The increment of $A^+(x)$ on the interval $(x_1,x_2]$ is

$$\begin{aligned}&A^+(x_2)-A^+(x_1)\\&=\sum_{k=1}^d \xi_k\Big[\mathbb{1}_{\{\xi_k\ge 0\}}\big[\nabla_k^+(x_2,\theta_0)-\nabla_k^+(x_1,\theta_0)\big]+\mathbb{1}_{\{\xi_k<0\}}\big[\nabla_k^-(x_2,\theta_0)-\nabla_k^-(x_1,\theta_0)\big]\Big]\,,\end{aligned}$$

and the related increment of $A^-(x)$ can be written in a similar form. Hence, if the assumption is satisfied then $A^+(x)$ and $A^-(x)$, $x\in\mathbb{R}$, have bounded or slowly growing

increments, and if the assumption holds with $\delta = 1$ then the processes have bounded or slowly growing increments on the unit intervals.

From (4.47) we clearly have

$$\hat{Y}(t) = \int_{\mathbb{R}} t^x \, dB\big(F(x,\theta_0)\big) + \int_{\mathbb{R}} t^x \, dA(x) = Y(t) + I^A(t) \tag{4.53}$$

for every real t where both sides are defined at, and by Proposition 4.10 the process $Y(t)$ exists on the interval $[0,1]$ and has a continuous modification on $(0,1)$. If $A^+(x)$ and $A^-(x)$ have bounded or slowly growing increments, then Proposition 4.10 and formula (4.53) imply that $I^A(t)$ and $\hat{Y}(t)$ are well-defined on the interval $[0,1)$ and has a modification which is sample-continuous on $(0,1)$. In the following the notation $\hat{Y}(t)$, $0 \le t < 1$, stands for this modification. Additionally, if $\nabla_k^+(x,\theta_0)$ and $\nabla_k^-(x,\theta_0)$ have finite limit at infinity for every $k = 1,\dots,d$, then the processes $A^+(x)$ and $A^-(x)$ have finite 0-th moment, and $I^A(t)$ and $\hat{Y}(t)$ exist at the point $t = 1$, as well.

If the law corresponding to the distribution function $F(x,\theta_0)$ is non-negative integer valued, then $\nabla_\theta F(x,\theta_0)$ and the processes $B(F(x,\theta_0))$, $A^+(x)$ and $A^-(x)$, $x \in \mathbb{R}$, are constant on the intervals $[m, m+1)$, $m = 0, 1, \dots$ In this case the integral $Y(t)$ can be considered also in Lebesgue–Stieltjes sense, and by Proposition 4.10 the process Y is well-defined and sample-continuous on $(-1,1)$. If $A^+(x)$ and $A^-(x)$ have bounded or slowly growing increments on the unit intervals, then Propositions 4.9 and equation (4.53) imply that $I^A(t)$ and $\hat{Y}(t)$ can be defined in Lebesgue–Stieltjes sense on $(-1,1)$ and they are continuous on this interval. Note that the investigation presented in Section 4.3 highlighted that the integral $Y(1)$ may be not defined in Lebesgue–Stieltjes sense, and hence, $\hat{Y}(1)$ do not exist for every possible background distribution $F(x,\theta_0)$, $x \in \mathbb{R}$, in this sense.

We summarize what we found in the next result.

Proposition 4.28. *Under Assumptions 2 the process $\hat{Y}$ is defined and has a sample-continuous modification on the interval $[a', b']$, where $[a', b']$ can be chosen as*

- *$[0, 1-\varepsilon]$ with any $0 < \varepsilon < 1$, if $\mathcal{F}$ is an arbitrary non-negative valued family;*
- *$[0,1]$, if $\nabla_k^+(x,\theta_0)$ and $\nabla_k^-(x,\theta_0)$ have finite limit as $x \to \infty$ for $k = 1,\dots,d$;*
- *$[-\tau, \tau]$ with any $0 < \tau < 1$, if $\mathcal{F}$ is non-negative integer valued and Assumption 2 holds with $\delta = 1$.*

Also, $\hat{\gamma}_n$, $\hat{\gamma}^*_{m_n,n}$ and $\tilde{\gamma}^*_{m_n,n}$ exists and continuous on $[a', b']$ with probability 1.

In our following theorem we prove approximations for certain representations of the parameter estimated generating processes $\hat{\gamma}_n$, $\hat{\gamma}^*_{m_n,n}$ and $\tilde{\gamma}^*_{m_n,n}$. For this end consider the independent variables $X_1, X_2, \dots$ provided by Theorem 3.1. The parameter estimated probability generating process based on the sample $X_1,\dots,X_n$ can be written in the form (4.49) with the estimated empirical process $\hat{\alpha}_n$ of the referred theorem. Let

$$\hat{Y}_n(t) = \int_{\mathbb{R}} t^x \, dG_n(x)\,, \qquad n = 1, 2, \dots \tag{4.54}$$

be copies of $\hat{Y}$ being defined by using the Gaussian processes $G_1(x), G_2(x), \ldots$, $x \in \mathbb{R}$, of Theorem 3.1. Also, using the bootstrap samples of Theorem 3.2 the parametric bootstrap estimated generating process $\hat{\gamma}^*_{m_n,n}$ can be represented by the equation (4.51) with the bootstrap empirical process $\hat{\alpha}^*_{m_n,n}$ of the theorem. Similarly, the non-parametric bootstrap generating process $\tilde{\gamma}^*_{m_n,n}$ based on the variables provided by Theorem 3.3 is in (4.52) with the corresponding empirical process $\tilde{\alpha}^*_{m_n,n}$. Finally, let $\hat{Y}^*_n(t)$ and $\tilde{Y}^*_n(t)$ be defined with the integral (4.54) by applying the processes $G_1, G_2, \ldots$ of Theorems 3.2 and 3.3 instead of those of Theorem 3.1, respectively.

Theorem 4.29. *Suppose that Assumption 2 is satisfied. Under the conditions of Theorem 3.1 we have*

$$\sup_{0\le t<1} \left|\hat{\gamma}_n(t) - \hat{Y}_n(t)\right| \xrightarrow{P} 0\,, \qquad n \to \infty\,.$$

Also, if the assumptions of Theorem 3.2 and Theorem 3.3 hold then we have

$$\sup_{0\le t<1} \left|\hat{\gamma}^*_{m_n,n}(t) - \hat{Y}^*_{m_n}\right| \xrightarrow{P} 0 \qquad \textit{and} \qquad \sup_{0\le t<1} \left|\tilde{\gamma}^*_{m_n,n}(t) - \hat{\gamma}_n(t) - \tilde{Y}^*_{m_n}\right| \xrightarrow{P} 0$$

as $n \to \infty$, respectively. Additionally, if $\nabla^+_k(x,\theta_0)$ and $\nabla^-_k(x,\theta_0)$, $x \in \mathbb{R}$, have finite limit at infinity for $k = 1, \ldots, d$, then the approximations hold also at the point $t = 1$. Furthermore, suppose that the all distributions in the family $\mathcal{F}$ are non-negative integer valued and Assumption 2 is valid with $\delta = 1$, and consider any value $0 < \tau < 1$. If the conditions of Theorem 3.1 are satisfied then

$$\sup_{-\tau\le t\le\tau} \left|\hat{\gamma}_n(t) - \hat{Y}_n(t)\right| \xrightarrow{P} 0\,, \qquad n \to \infty\,,$$

and we have the approximations

$$\sup_{-\tau\le t\le\tau} \left|\hat{\gamma}^*_{m_n,n}(t) - \hat{Y}^*_{m_n}\right| \xrightarrow{P} 0 \qquad \textit{and} \qquad \sup_{-\tau\le t\le\tau} \left|\tilde{\gamma}^*_{m_n,n}(t) - \hat{\gamma}_n(t) - \tilde{Y}^*_{m_n}(t)\right| \xrightarrow{P} 0$$

as $n \to \infty$ under the conditions of Theorem 3.2 and Theorem 3.3, respectively.

Proof. Observe that under the conditions all processes are defined on the intervals, where the supremums are taken, and consider the process

$$K_n(x) = \hat{\alpha}_n(x) - G_n(x)\,, \qquad x \in \mathbb{R}\,,$$

with the sequence $G_1, G_2, \ldots$ provided by Theorem 3.1. Applying the first inequality of Theorem 4.7 and the approximation of Theorem 3.1 we obtain that

$$\sup_{0\le t<1} \left|\hat{\gamma}_n(t) - \hat{Y}_n(t)\right| = \sup_{0\le t<1} \left|\int_{\mathbb{R}} t^x \, dK_n(x)\right| \le \sup_{x\in\mathbb{R}} |K_n(x)| \xrightarrow{P} 0\,, \qquad n \to \infty\,.$$

Furthermore, by Theorem 4.7 the supremum can be extended also to $t = 1$ if $\hat{\gamma}_n$ and $\hat{Y}_n$ exist at this point, which additional condition is satisfied if the functions $\nabla^+_k(x,\theta_0)$

and $\nabla_k^-(x,\theta_0)$, $x\in\mathbb{R}$, have finite limit at infinity for $k=1,\dots,d$. In the integer valued case Theorems 4.9 and 3.1 imply that

$$\sup_{-\tau\le t\le\tau}\left|\hat\gamma_n(t)-\hat Y_n(t)\right| = \sup_{-\tau\le t\le\tau}\left|\int_{\mathbb{R}} t^x\,dK_n(x)\right| \le C_2(\tau,0)\sup_{x\in\mathbb{R}}|K_n(x)| \xrightarrow{P} 0\,,\quad n\to\infty\,.$$

The corresponding statements for the parametric and the non-parametric bootstrap processes can be proved similarly by applying the approximations of Theorems 3.2 and 3.3, respectively. □

Since the processes are sample-continuous on the interval $[a',b']$ of Proposition 4.28, Theorem 4.29 has the following consequence.

Corollary 4.30. *Suppose that Assumption 2 holds, and consider the interval $[a',b']$ provided by Proposition 4.28. Under the conditions of Theorem 3.1 the process $\hat\gamma_n$ converges to $\hat Y$ in distribution in the space $C[a',b']$. Also, $\hat\gamma^*_{m_n,n}$ and $\tilde\gamma^*_{m_n,n}-\hat\gamma_n$ have the same weak limit in $C[a',b']$ if the assumptions of Theorems 3.2 and 3.3 are satisfied, respectively.*

Based on the results presented in Theorem 4.29 and Corollary 4.30 one can test the fit of a given non-negative valued sample $X_1,\dots,X_n$ to the family $\mathcal{F}$ by using the bootstrap algorithm of Section 3.4. Assume that we have the approximations

$$\sup_{a'\le t\le b'}\left|\hat\gamma_n(t)-\hat Y_n(t)\right|\xrightarrow{P}0 \qquad\text{and}\qquad \sup_{a'\le t\le b'}\left|\gamma^*_{m_n,n}(t)-Y^*_{m_n}(t)\right|\xrightarrow{P}0\,,$$

on some interval $[a',b']$, where $\gamma^*_{m_n,n}$ stands for $\hat\gamma^*_{m_n,n}$ in the parametric and for $\tilde\gamma^*_{m_n,n}$ in the non-parametric bootstrap case, and $Y^*_{m_n}$ is the corresponding copy of $\hat Y$ provided by Theorem 4.29. Also, suppose that the limit process $\hat Y(t)$ is not degenerate at every point t on the interval $[a',b']$, because we can not obtain any statistical result based on a constant limit distribution. Consider a continuous functional $\psi: C[a',b']\to\mathbb{R}$ such that the statistics

$$\psi^*_{m_n,n}=\psi(\gamma^*_{m_n,n}) \qquad\text{and}\qquad \varphi_n=\psi(Y^*_n)\,,\qquad n=1,2,\dots$$

satisfy the assumptions of Theorem 2.7, that is, the common distribution function of the φ_n's is continuous, and

$$\left|\psi^*_{m_n,n}-\varphi_n\right|\xrightarrow{P}0\,,\qquad n\to\infty\,. \tag{4.55}$$

Then, all conditions of the bootstrap algorithm hold, and we can apply the method in the same way as in Section 3.4.

It can be easily seen that we have (4.55) if ψ is Lipschitzian, that is, if

$$\left|\psi(h_1)-\psi(h_2)\right|\le M\sup_{a\le t\le b}\left|h_1(t)-h_2(t)\right|\,,\qquad h_1,h_2\in C[a',b']\,,$$

with some constant M, because using the assumptions we obtain the convergence

$$\left|\psi^*_{m_n,n} - \varphi_n\right| \le \sup_{a' \le t \le b'} \left|\gamma^*_{m_n,n}(t) - Y^*_{m_n}(t)\right| \xrightarrow{P} 0\,.$$

Since the supremum functional is Lipschitzian, the Kolmogorov–Smirnov type statistics of the processes satisfy (4.55). The other condition, the continuity of the distribution function F_φ on the whole real line or, at least, at certain points $x \in \mathbb{R}$, requires different considerations for different functionals. We investigate only the supremum functional

$$\psi(h) = \sup_{a' \le t \le b'} |h(t)|\,, \qquad h \in C[a', b']\,.$$

Observe that under the regularity conditions of Proposition 4.29 we can apply the first or the second inequality of Theorem 4.7 or the inequality of Theorem 4.9 on the interval $[a', b']$ for the process $\hat{Y}$ defined by formula (4.50), and it follows that

$$\psi(\hat{Y}) = \sup_{a' \le t \le b'} |\hat{Y}(t)| \le \sup_{x \in \mathbb{R}} |G(x)|\,. \tag{4.56}$$

Since the right side of the inequality is finite with probability 1 by our results in Section 3.4, we obtain that $\psi(\hat{Y})$ is almost surely a finite variable. Because the process $\hat{Y}$ is Gaussian and continuous on $[a', b']$, and it is not degenerate at every point on this interval, Theorem 2.10 implies that the distribution function $F_\varphi(x)$, $x \in \mathbb{R}$, of variable

$$\varphi = \psi(\hat{Y}) = \sup_{a' \le t \le b'} |\hat{Y}(t)| = \sup\left\{\hat{Y}(t), -\hat{Y}(t) : a' \le t \le b', t \in \mathbb{Q}\right\}$$

is continuous on the interval (s_0, ∞), where

$$s_0 = \inf\left\{x \in \mathbb{R} : F_\varphi(x) > 0\right\} \in [0, \infty)$$

is the left endpoint of the support of F_φ.

We suggest that s_0 is equal to 0 and the function F_φ is continuous at s_0, as well, but unfortunately, we can not prove our conjecture. Recall that we had a similar unproven conjecture in Section 3.4 for the supremum of the process $G(x)$, $x \in \mathbb{R}$. Let

$$s_1 = \inf\left\{x \in \mathbb{R} : P\Big(\sup_{x \in \mathbb{R}} |G(x)| \le x\Big) > 0\right\},$$

and note that (4.56) implies the inequality $s_0 \le s_1$. If one can show that $s_1 = 0$, then we immediately get that $s_0 = 0$, and using the assumption that the Gaussian process $\hat{Y}$ is not degenerate at every point t on the interval $[a', b']$, the continuity of F_φ at s_0 follows. Hence, F_φ is continuous on the whole real line, and we can apply the bootstrap method without any restriction.

Summary

Introduction

In the thesis we investigate the asymptotic behavior of some empirical processes based on independent and identically distributed random variables. In most cases we apply the approximation method which means that on a suitable probability space we construct a representation of the underlying variables and empirical processes, and also, a sequence of copies of a Gaussian process such that the distance between the empirical process and the Gaussian processes converges to zero in almost sure or stochastic sense as the sample size goes to infinity. In this way, using the properties of the applied Gaussian process, we can obtain information on the examined empirical process.

The thesis is organized as follows. In Chapter 2 we introduce some basic tools which will be applied in our research. In Chapter 3 we investigate the parametric and the nonparametric bootstrap versions of the parameter estimated empirical process defined on a parametric family of distributions, and we demonstrate the bootstrap technique in a simulation study. Finally, in Chapter 4 we provide an effective and flexible background for the study of empirical processes based on probability generating functions of nonnegative valued variables. Using this framework we prove asymptotic results for the empirical probability generating process and its derivatives, and for the corresponding bootstrapped and/or parameter estimated versions.

Some basic concepts

In the chapter we introduce three concepts. The first one is the Hungarian construction or so-called KMT approximation for the uniform empirical process $\beta_n(u)$, $0 \le u \le 1$, based on independent variables distributed uniformly on the interval $[0,1]$. By the construction one can define the uniform variables on a suitable probability space carrying a sequence of Brownian bridges $B_1, B_2, \dots$ such that we have

$$\sup_{0\le u\le 1} \big|\beta_n(u) - B_n(u)\big| = \mathcal{O}\big(n^{-1/2}\log n\big)\,, \qquad n\to\infty\,, \qquad \text{a.s.}$$

This construction of Komlós, Major and Tusnády will be essential in our research.

The second tool is Efron's bootstrap method for estimating the distribution of some statistics τ_n based on a given sample having n elements. Note that such an estimation

can not be obtained by using only the standard statistical techniques, because having only one set of sample variables we have only one observation for the variable τ_n. By the bootstrap heuristics if we estimate the unknown distribution function $F(x)$ of the sample variables with a function $\hat{F}_n(x)$, $x \in \mathbb{R}$, and consider conditionally independent variables $X^*_{1,n}, \dots, X^*_{m_n,n}$ having conditional distribution function $\hat{F}_n$ with respect to the original observations, then the corresponding statistics $\tau^*_{m_n,n} = \tau_{m_n}(X^*_{1,n}, \dots, X^*_{m_n,n})$ has "similar" distribution as τ_n has. Since the distribution of $\tau^*_{m_n,n}$ can be obtained in arbitrary precision by direct calculations or by applying Monte Carlo simulation, we can get a better or worse estimation for the distribution of τ_n. We apply two versions of the bootstrap technique, the parametric and the non-parametric bootstrap.

Finally, we need to extend the standard theory of stochastic integration on a finite interval with respect to a locally square integrable martingale to stochastic integration on the whole real line. We provide a condition for the existence of the integral, and we examine the distribution of processes defined as the integrals of bivariate functions with respect to the one-dimensional standard Wiener process.

Bootstrap parameter estimated empirical processes

Consider a parametric collection of distributions $\mathcal{F} = \{F(x,\theta) : x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^d\}$, and independent variables $X_1, X_2, \dots$ having common distribution function $F(x,\theta_0)$, $x \in \mathbb{R}$, with a fixed $\theta_0 \in \Theta$. If $F_n(x)$, $x \in \mathbb{R}$, stands for the empirical distribution function of the first n elements of the sequence and $\hat{\theta}_n$ is an estimator of θ_0 based on this sample, then the corresponding parameter estimated empirical process is

$$\hat{\alpha}_n(x) = n^{1/2}\big[F_n(x) - F(x,\hat{\theta}_n)\big], \qquad x \in \mathbb{R}.$$

Since Durbin proved the weak convergence of $\hat{\alpha}_n$ to a Gaussian process $G(x)$, $x \in \mathbb{R}$, as the sample size n goes to infinity, the parameter estimated empirical process became a widely used tool to test goodness-of-fit to parametric distribution families. In general, statistical methods based on the process are not distribution free, and the critical values can not be calculated in theoretical way. However, one can avoid these difficulties by applying the parametric or the non-parametric bootstrap technique.

Consider bootstrapped sample variables $X^*_{1,n}, \dots, X^*_{m_n,n}$ based on $X_1, \dots, X_n$, and let $F^*_{m_n,n}(x)$, $x \in \mathbb{R}$, stand for the empirical distribution function of the bootstrapped variables, and let θ^*_n be a parameter estimator based on the bootstrapped sample. The bootstrapped parameter estimated empirical process can be defined as the parameter estimated empirical process based on the bootstrapped sample, that is, by the form

$$\bar{\alpha}^*_{m_n,n}(x) = n^{1/2}\big[F^*_{m_n,n}(x) - F(x,\theta^*_n)\big], \qquad x \in \mathbb{R}.$$

The process is denoted by $\hat{\alpha}^*_{m_n,n}$ in the parametric and by $\tilde{\alpha}^*_{m_n,n}$ in the non-parametric bootstrap case. The heuristics of the bootstrap method is that if $\hat{\alpha}^*_{m_n,n}$ and/or $\tilde{\alpha}^*_{m_n,n}$ converges in distribution to the same weak limit as $\hat{\alpha}_n$ does, then the critical values of a test statistic $\psi(\hat{\alpha}_n)$ can be estimated by the empirical quantiles of the corresponding

functional $\psi(\hat{\alpha}^*_{m_n,n})$ and/or $\psi(\tilde{\alpha}^*_{m_n,n})$. It turns out that the convergence of $\hat{\alpha}^*_{m_n,n}$ is in fact true, but $\tilde{\alpha}^*_{m_n,n}$ requires bias correction.

In Sections 3.2 and 3.3 we show that, under certain conditions on the distribution family $\mathcal{F}$ and the parameter estimation method, on a suitable probability space, one can construct a representation of the original sequence and the bootstrapped variables and copies $G_1, G_2, \dots$ of the limiting process G such that we have the approximation

$$\sup_{x\in\mathbb{R}} \left|\hat{\alpha}^*_{m_n,n}(x) - G_{m_n}(x)\right| \xrightarrow{P} 0\,, \qquad n\to\infty\,,$$

in the parametric and

$$\sup_{x\in\mathbb{R}} \left|\tilde{\alpha}^*_{m_n,n}(x) - \left(\frac{m_n}{n}\right)^{1/2}\hat{\alpha}_n(x) - G_{m_n}(x)\right| \xrightarrow{P} 0\,, \qquad n\to\infty\,,$$

in the non-parametric case. As a direct consequence we obtain the weak convergence of the processes

$$\hat{\alpha}^*_{m_n,n}(x) \qquad \text{and} \qquad \tilde{\alpha}^*_{m_n,n}(x) - \left(\frac{m_n}{n}\right)^{1/2}\hat{\alpha}_n(x)\,, \qquad x\in\mathbb{R}\,,$$

as $n\to\infty$ to the limit of $\hat{\alpha}_n$, that is, to the Gaussian process G.

We present the bootstrap testing algorithm in Section 3.4, and we prove that the method can applied for the Kolmogorov–Smirnov type supremum functionals on the empirical processes. In Section 3.5 we discuss on the regularity conditions of the results, and we check the validity of these assumptions for the Poisson and the normal distribution family endowed with the maximum likelihood parameter estimation method. To demonstrate the bootstrap technique in an application in the last section we report on simulation studies. Using the parametric and the non-parametric bootstrap method we test the fit of negative binomial variables having various parameters to the Poisson distribution, and also, the fit of location and scale contaminated normal samples to the normal family.

Empirical probability generating processes

Let $X, X_1, X_2, \dots$ be a sequence of independent and identically distributed nonnegative valued random variables having distribution function $F(x)$, $x\in\mathbb{R}$. Let

$$g(t) = Et^X = \int_{\mathbb{R}} t^x dF(x) \qquad \text{and} \qquad g_n(t) = \frac{1}{n}\sum_{j=1}^{n} t^{X_j}\,, \qquad 0\le t\le 1\,,$$

be the common probability generating function and its empirical counterpart based on the first n observations. Throughout this chapter the symbol 0^0 is interpreted as 1, because we will need the continuity of the function t^x in variable x. Then the empirical probability generating process can be defined by

$$\gamma_n(t) = n^{1/2}\big[g_n(t) - g(t)\big]\,, \qquad 0\le t\le 1\,.$$

The idea of the application of generating functions to solve various statistical problems is not unusual, similar transformed processes based on empirical characteristic and moment generating functions are well-known. (See Csörgő (1981) and Csörgő, Csörgő, Horváth and Mason (1986).) In each case the theoretical basis of the method is the fact, that under appropriate conditions the transformed processes converge in distribution in some function space. In the case of the empirical probability generating process, Rémillard and Theodorescu (2000) state that γ_n converges in distribution in $C[0,1]$ to the process

$$Y(t) = \int_{\mathbb{R}} t^x \, dB\big(F(x)\big)\,, \qquad 0 \le t \le 1\,,$$

for every non-negative integer valued variable X. Unfortunately, there is an oversight in their proof, but we show that their basic idea is good, and the proof can be corrected.

The aim of the chapter is to present a general approach to convergence problems for probability generating functions and processes and their derivatives, and for the bootstrapped and/or parameter estimated versions of the empirical probability generating process. Our results are general in the other sense, as well, that they hold not only for an integer valued variable, but for an arbitrary non-negative valued X.

In Section 4.2 we investigate processes defined by the integral

$$I_r(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dK(x)\,,$$

where the function $K(x)$ can be represented by the sum of a locally square integrable martingale $M(x)$ and a process $A(x)$ being of bounded variation on finite intervals, $x \in \mathbb{R}$. Also, we assume that M and A vanish on the negative half-line $(-\infty, 0)$ and have càdlàg trajectories. In Propositions 4.2, 4.3 and 4.8 we provide conditions under which the process I_r exists on certain subsets $[a,b]$ of the interval $(-1,1]$. The main results of the section are Theorems 4.7 and 4.9 where we prove inequalities in the form

$$\sup_{a\le t\le b} |Y_r(t)| \le C \sup_{x\in\mathbb{R}} |K(x)|$$

with a constant $C = C(r,a,b)$ not depending on the process K. In the following sections these inequalities trivialize the investigation of probability generating processes, because we can obtain asymptotic results for them simply by applying the asymptotic properties of the corresponding empirical processes.

In Section 4.3 we prove that the empirical probability generating process γ_n has r-th ($r = 0, 1, \dots$) derivative

$$\gamma_n^{(r)}(t) = n^{1/2}\big[g_n^{(r)}(t) - g^{(r)}(t)\big] = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, d\alpha_n(x)\,,$$

where $g_n^{(r)}$ and $g^{(r)}$ are the r-th derivative of the empirical and the theoretical probability generating function of the variables $X_1, \dots, X_n$, and $\alpha_n(x)$, $x \in \mathbb{R}$, is the empirical process corresponding to the sample. Also, we investigate the process

$$Y_r(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dB\big(F(x)\big)\,,$$

with the Brownian bridge B. In Proposition 4.10 we found that $\gamma_n^{(r)}$ is continuous and Y_r has a sample-continuous modification on certain subintervals $[a, b]$ of $(-1, 1]$. In Section 4.11 we show that Y_r is Gaussian with mean zero and continuous covariance function. Also, we prove that the Kolmogorov–Smirnov type supremum functionals and the Cramér–von Mises type integral functional of Y_r have bounded density function.

As we specified earlier, Rémillard and Theodorescu (2000) showed the convergence of γ_n in distribution to Y for non-negative integer valued random variables, but there is an oversight in their proof. However, the justification method is interesting, and we correct it in Section 4.5. In Section 4.6 we improve this result by proving uniform strong approximations for the empirical probability generating process γ_n and its derivatives not only in the integer valued case but in case of an arbitrary non-negative valued X. We provide a rate of convergence, as well, and we show that representing the processes γ_n and Y_r on a suitable probability space we have

$$\sup_{a \le t \le b} \left| \gamma_n^{(r)}(t) - Y_{r,n}(t) \right| = \mathcal{O}\big(n^{-1/2} \log n\big)\,, \qquad n \to \infty\,, \qquad \text{a.s.}$$

The approximation immediately implies the weak convergence of $\gamma_n^{(r)}$ to Y_r in the space $C[a, b]$. Additionally, we obtain the uniform convergence of the distribution functions of the Kolmogorov–Smirnov type supremum functionals and the Cramér–von Mises type integral functional of the process $\gamma_n^{(r)}$ with rates of convergence.

In Section 4.7 we prove the law of the iterated logarithm

$$\limsup_{n \to \infty} \frac{\sup_{a \le t \le b} |\gamma_n^{(r)}(t)|}{(\log \log n)^{1/2}} \le \frac{C}{2^{1/2}} \qquad \text{a.s.}$$

with a constant $C = C(a, b, r)$ not depending on the distribution of the variable X. Furthermore, we show that the process $\gamma_n^{(r)}$ is relative compact in the space $C[a, b]$, and we determine the set of limit functions.

In Section 4.8 we investigate the bootstrapped version of the probability generating process $\gamma_n^{(r)}$, and we prove a strong approximation for it. Based on this result we show how to construct confidence bands for the unknown probability generating function $g(t)$, $a \le t \le b$, of the variable X. Finally, based on a parametric family of distributions we define the parameter estimated probability generating process and its parametric and non-parametric bootstrap versions in Section 4.9. Applying the result of Chapter 3 we prove weak approximations for the processes, and we explain their application for testing goodness-of-fit.

Összefoglalás

Bevezetés

A disszertációban független és azonos eloszlású változók segítségével felírt bizonyos empirikus folyamatok aszimptotikus viselkedését vizsgáljuk. A legtöbb esetben az approximációs módszert fogjuk alkalmazni, ami azt jelenti, hogy a vizsgált változókat és a kapcsolatos empirikus folyamatokat egy megfelelően választott valószínűségi mezőn konstruáljuk meg, valamint ezen a mezőn definiáljuk egy megfelelő Gauss folyamat reprezentánsait. Mindezt olyan módon tesszük, hogy az empirikus folyamat és a Gauss folyamatok távolsága nullához konvergál, amint a mintaméret tart a végtelenbe. Ezáltal az alkalmazott Gauss folyamat tulajdonságai alapján leírhatjuk a vizsgált empirikus folyamat viselkedését.

A disszertáció a következőképpen épül fel. A 2. fejezetben ismertetünk néhány alapvető eszközt, melyeket alkalmazni fogunk a vizsgálataink során. A 3. fejezetben a paraméteres eloszláscsaládokon definiált becsült paraméteres empirikus folyamat paraméteres és nemparaméteres bootstrap változatát tanulmányozzuk, valamint egy szimulációs tanulmány segítségével bemutatjuk a módszer gyakorlati alkalmazását. Végül, a 4. fejezetben megteremtünk egy hatékony és rugalmas elméleti hátteret a nemnegatív értékű valószínűségi változók valószínűségi generátorfüggvényei alapján felírt empirikus folyamatok vizsgálatához. Ennek segítségével aszimptotikus tételeket bizonyítunk az empirikus valószínűségi generátor folyamatra és deriváltjaira, továbbá a kapcsolatos bootstrap és/vagy paraméterbecsült változatokra.

Néhány alapvető fogalom

A fejezetben három elméleti fogalmat mutatunk be. Az első a független, a $[0,1]$ intervallumon egyenletes eloszlású változók segítségével felírt $\beta_n(u)$, $0 \le u \le 1$, egyenletes empirikus folyamatra vonatkozó KMT approximáció. Eszerint egy alkalmasan választott valószínűségi mezőn a valószínűségi változók definiálhatóak olyan módon, hogy Brown hidaknak egy alkalmas $B_1, B_2, \ldots$ sorozatára

$$\sup_{0\le u\le 1} \left|\beta_n(u) - B_n(u)\right| = \mathcal{O}\left(n^{-1/2}\log n\right), \qquad n\to\infty, \qquad \text{m.b.}$$

A konstrukció Komlós Jánostól, Major Pétertől and Tusnády Gábortól származik, és alapvető lesz a munkánk során.

A második eszköz Efron bootstrap módszere, melynek segítségével megbecsülhetjük egy n elemű mintára felírt τ_n statisztika eloszlását. Vegyük észre, hogy ilyen becslés a szokásos statisztikai technikákkal nem adható, ugyanis egyetlen minta birtokában a τ_n változóra csupán egy megfigyelés áll rendekésünkre. A bootstrap alapötlete az, hogy ha a mintaelemek ismeretlen $F(x)$ eloszlásfüggvényét egy $\hat{F}_n(x)$, $x \in \mathbb{R}$, függvénnyel becsüljük, és tekintünk az eredeti mintára nézve feltételesen független $X^*_{1,n}, \ldots, X^*_{m_n,n}$ változókat, melyek feltételes eloszlásfüggvénye $\hat{F}_n$, akkor $\tau^*_{m_n,n} = \tau_{m_n}(X^*_{1,n}, \ldots, X^*_{m_n,n})$ eloszlása „hasonlít” τ_n eloszlásához. Mivel elméleti számítások vagy Minte Carlo szimuláció révén $\tau^*_{m_n,n}$ eloszlása tetszőleges pontossággal megkapható, ilyen módon egy jobb vagy rosszabb becslést nyerhetünk τ_n eloszlására. Munkánk során a bootstrap módszer két változatát alkalmazzuk, a paraméteres és a nemparaméteres bootstrapet.

Végül, ki kell terjesztenünk a véges intervallumon értelmezett lokálisan négyzetesen integrálható martingálokra vett sztochasztikus integrált az egész valós egyenesen vett sztochasztikus integrálra. Bizonyítunk egy állítást az integrál létezésére, és leírjuk az olyan folyamatok eloszlását, melyek bizonyos kétváltozós függvényeknek a standard Wiener folyamatra vett integráljaként állnak elő.

Bootstrap becsült paraméteres empirikus folyamatok

Tekintsük eloszlásoknak egy $\mathcal{F} = \{F(x,\theta) : x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^d\}$ paraméterezett családját, továbbá független $X_1, X_2, \ldots$ változókat közös $F(x,\theta_0)$, $x \in \mathbb{R}$, eloszlásfüggvénnyel, ahol $\theta_0 \in \Theta$. Jelölje $F_n(x)$, $x \in \mathbb{R}$, a sorozat első n elemének, mint mintának az empirikus eloszlásfüggvényét, és legyen $\hat{\theta}_n$ az ismeretlen θ_0 paraméter egy becslése. Ekkor a becsült paraméteres empirikus folyamat

$$\hat{\alpha}_n(x) = n^{1/2}\big[F_n(x) - F(x,\hat{\theta}_n)\big]\,, \qquad x \in \mathbb{R}\,.$$

Mióta Durbin bebizonyította, hogy $\hat{\alpha}_n(x)$ gyengén konvergál egy $G(x)$, $x \in \mathbb{R}$, Gauss folyamathoz, amint a mintaméret tart a végtelenbe, a becsült paraméteres empirikus folyamat széles körben használt eszköz összetett illeszkedési hipotézisek tesztelésére. Sajnos a folyamatra épülő statisztikai módszerek általában nem eloszlásmentesek, és a kapcsolatos kritikus értékeket nem lehet elméleti úton meghatározni. Szerencsére ezen nehézségek kiküszöbölhetőek a paraméteres vagy a nemparaméteres bootstrap módszer alkalmazásával.

Tekintsünk $X^*_{1,n}, \ldots, X^*_{m_n,n}$ bootstrap mintaelemeket az $X_1, \ldots, X_n$ megfigyelések alapján. Legyen $F^*_{m_n,n}(x)$, $x \in \mathbb{R}$, a bootstrap változók empirikus eloszlásfüggvénye, és legyen θ^*_n paraméterbecslés a bootstrap mintaelemek segítségével. A bootstrap becsült paraméteres empirikus folyamat a bootstrap mintaelemekre felírt becsült paraméteres empirikus folyamat, tehát

$$\bar{\alpha}^*_{m_n,n}(x) = n^{1/2}\big[F^*_{m_n,n}(x) - F(x,\theta^*_n)\big]\,, \qquad x \in \mathbb{R}\,.$$

A folyamatra a $\hat{\alpha}^*_{m_n,n}$ jelölést használjuk a paraméteres és a $\tilde{\alpha}^*_{m_n,n}$ jelölést a nemparaméteres bootstrap esetben. A bootstrap alkalmazásának motivációja az az ötlet, hogy

ha $\hat{\alpha}^*_{m_n,n}$ és/vagy $\hat{\alpha}^*_{m_n,n}$ eloszlásban konvergál ugyanazon határfolyamathoz, mint $\hat{\alpha}_n$, akkor egy $\psi(\hat{\alpha}_n)$ statisztika kritikus értékei becsülhetőek, mint a kapcsolatos $\psi(\hat{\alpha}^*_{m_n,n})$ és/vagy $\psi(\tilde{\alpha}^*_{m_n,n})$ funkcionál empirikus kvantilisei. Vizsgálataink során kiderül, hogy $\hat{\alpha}^*_{m_n,n}$ valóban konvergál, de a $\tilde{\alpha}^*_{m_n,n}$ folyamat esetében bias korrekcióra van szükség.

Tegyük fel, hogy teljesülnek bizonyos, az $\mathcal{F}$ családra és az alkalmazott paraméterbecslő eljárásra vonatkozó feltételek. A 3.2. és a 3.3. fejezetben megmutatjuk, hogy egy megfelelő valószínűségi mezőn konstruálható az eredeti és a bootstrappelt valószínűségi változóknak olyan reprezentációja, hogy

$$\sup_{x\in\mathbb{R}} \left|\hat{\alpha}^*_{m_n,n}(x) - G_{m_n}(x)\right| \xrightarrow{P} 0\,, \qquad n\to\infty\,,$$

a paraméteres, és

$$\sup_{x\in\mathbb{R}} \left|\tilde{\alpha}^*_{m_n,n}(x) - \left(\frac{m_n}{n}\right)^{1/2}\hat{\alpha}_n(x) - G_{m_n}(x)\right| \xrightarrow{P} 0\,, \qquad n\to\infty\,,$$

a nemparaméteres bootstrap esetben. A $G_1, G_2, \ldots$ folyamatok a G Gauss folyamat reprezentációi. Ezen eredményekből közvetlenül kapjuk a

$$\hat{\alpha}^*_{m_n,n}(x) \qquad \text{és a} \qquad \tilde{\alpha}^*_{m_n,n}(x) - \left(\frac{m_n}{n}\right)^{1/2}\hat{\alpha}_n(x)\,, \qquad x\in\mathbb{R}\,,$$

folyamat eloszlásbeli konvergenciáját a G Gauss folyamathoz.

A bootstrap tesztelő algoritmus a 3.4. fejezetben található. Ugyanitt megmutatjuk, hogy a módszer alkalmazható a vizsgált empirikus folyamatok Kolmogorov–Szmirnov típusú szuprémum funkcionáljaira. A 3.5. fejezetben körbejárjuk a tételek regularitási feltételeit, és bebizonyítjuk, hogy ezen feltételek teljesülnek a Poisson és a normális eloszlásra és a maximum likelihood becslésre. Hogy bemutassuk a bootstrap technikát egy gyakorlati alkalmazáson keresztül, az utolsó fejezetben ismertetjük egy szimulációs tanulmány eredményeit. A paraméteres és a nemparaméteres bootstrap alkalmazásával teszteljük negatív binomiális minták illeszkedését a Poisson családhoz, valamint lokáció és skála kontaminált normális változók illeszkedését a normális eloszláshoz.

Empirikus valószínűségi generátor folyamatok

Legyen $X, X_1, X_2, \ldots$ nemnegatív értékű független és azonos eloszlású valószínűségi változó $F(x)$, $x\in\mathbb{R}$, eloszlásfüggvénnyel, és legyen

$$g(t) = Et^X = \int_{\mathbb{R}} t^x dF(x) \qquad \text{és} \qquad g_n(t) = \frac{1}{n}\sum_{j=1}^{n} t^{X_j}\,, \qquad 0\le t\le 1\,,$$

az első n elem valószínűségi generátorfüggvénye és a generátorfüggvény empirikus változata. A fejezetben a 0^0 szimbólum 1-nek van definiálva, ugyanis szükségünk lesz a t^x függvény az x változóban való folytonosságára. Ekkor az empirikus valószínűségi generátor folyamat

$$\gamma_n(t) = n^{1/2}\big[g_n(t) - g(t)\big]\,, \qquad 0\le t\le 1\,.$$

A generátorfüggvények használata statisztikai problémák megoldására nem új ötlet, a karakterisztikus és a momentumgeneráló függvényen alapuló hasonló transzformált folyamatok ismertek. (Például, Csörgő (1981) és Csörgő, Csörgő, Horváth and Mason (1986).) Ezen módszerek elméleti alapja az, hogy bizonyos feltételek teljesülése esetén a transzformált folyamatok eloszlásban konvergálnak valamilyen függvénytérben. A valószínűségi generátor folyamat esetében Rémillard and Theodorescu (2000) mondta ki, hogy γ_n eloszlásban konvergál a $C[0,1]$ téren a

$$Y(t) = \int_{\mathbb{R}} t^x \, dB\big(F(x)\big)\,, \qquad 0 \le t \le 1\,,$$

folyamathoz minden nemnegatív egész értékű X változóra. Sajnos a bizonyításukba belecsúszott egy hiba, de az alapötlet jó, és a bizonyítás javítható.

A fejezet célja kidolgozni egy olyan általános és rugalmas eszköztárat, melynek segítségével vizsgálhatjuk az empirikus generátor folyamat és deriváltjai, valamint a bootstrap és/vagy becsült paraméteres változatok aszimptotikus viselkedését. Az eredmények abban az értelemben is általánosak, hogy nem csak az egész értékű esetben alkalmazhatóak, hanem tetszőleges nemnegatív értékű változóra.

A 4.2. fejezetben olyan folyamatokat vizsgálunk, melyeket az

$$I_r(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dK(x)$$

integrál definiál, ahol a $K(x)$ függvény előáll egy lokálisan négyzetesen integrálható $M(x)$ martingál és egy $A(x)$ korlátos változású folyamat összegeként, $x \in \mathbb{R}$. Emellett feltesszük, hogy M és A eltűnik a $(-\infty, 0)$ negatív félegyenesen, és a folyamatoknak càdlàg trajektóriái vannak. A 4.2., 4.3. és 4.8. Állításban megmutatjuk, hogy bizonyos feltételek mellett az I_r folyamat definiált a $(-1,1]$ intervallum valamely $[a,b]$ részhalmazain. Az általános rész fő eredménye a 4.7 és a 4.9 Tétel. Ezekben az I_r folyamatra vonatkozó egyenlőtlenségeket bizonyítunk

$$\sup_{a\le t\le b} |Y_r(t)| \le C \sup_{x\in\mathbb{R}} |K(x)|$$

alakban, ahol $C = C(r,a,b)$ a K folyamattól független konstans. Ezen egyenlőtlenségek alkalmazásával a következő fejezetekben a generátorfolyamatokra vonatkozó problémák visszavezethetőek a kapcsolatos empirikus folyamatok aszimptotikus tulajdonságaira.

A 4.3. fejezetben megmutatjuk, hogy a γ_n generátor folyamat r. deriváltja

$$\gamma_n^{(r)}(t) = n^{1/2}\big[g_n^{(r)}(t) - g^{(r)}(t)\big] = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, d\alpha_n(x)$$

alakban írható fel, $r = 0, 1, \ldots$, ahol $g_n^{(r)}$ és $g^{(r)}$ az empirikus és az elméleti valószínűségi generátorfüggvény r. deriváltja az $X_1, \ldots, X_n$ minta alapján, továbbá $\alpha_n(x)$, $x \in \mathbb{R}$, a kapcsolatos empirikus folyamat. Emellett definiáljuk az

$$Y_r(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\, dB\big(F(x)\big)$$

folyamatot, ahol B a Brown híd. A 4.10. Állításban megmutatjuk, hogy $\gamma_n^{(r)}$ folytonos és az Y_r folyamatnak létezik mintafolytonos modifikációja a $(-1,1]$ bizonyos $[a,b]$ részintervallumain. A 4.11. fejezetben belátjuk, hogy Y_r Gauss folyamat nulla várható értékkel és folytonos kovariancia függvénnyel. Továbbá bizonyítjuk, hogy a folyamat Kolmogorov–Szmirnov típusú szuprémum funkcionáljai és Cramér–von Mises típusú integrál funkcionálja abszolút folytonos változók korlátos sűrűségfüggvénnyel.

Mint azt korábban már említettük, Rémillard and Theodorescu (2000) megmutatta, hogy γ_n eloszlásban konvergál az Y folyamathoz tetszőleges nemnegatív egész értékű változó esetén, de van egy kisebb hiba a bizonyításban. Mindazonáltal az alapötletük érdekes, és a 4.5. fejezetben sikerül kijavítanunk a bizonyítást. A következő fejezetben továbbfejlesztjük ezt az eredmény, és bizonyítunk egy erős approximációs tételt a γ_n empirikus valószínűségi generátor folyamatra és deriváltjaira. Ez azt jelenti, hogy egy alkalmas valószínűségi mezőn megkonstruáljuk a γ_n és a Y_r folyamatok olyan reprezentánsait, melyekre

$$\sup_{a\le t\le b}\left|\gamma_n^{(r)}(t)-Y_{r,n}(t)\right| = \mathcal{O}\big(n^{-1/2}\log n\big)\,, \qquad n\to\infty\,, \qquad \text{m.b.}$$

Az approximációból azonnal jön, hogy $\gamma_n^{(r)}$ eloszlásban konvergál az Y_r folyamathoz a $C[a,b]$ térben. Emellett bebizonyítjuk, hogy a $\gamma_n^{(r)}$ deriválton értelmezett Kolmogorov–Szmirnov típusú szuprémum és Cramér–von Mises típusú integrál funkcionálok eloszlásfüggvényei egyenletesen konvergálnak egy meghatározott rátával.

A 4.7. fejezetben igazolunk egy iterált logaritmustételt a $\gamma_n^{(r)}$ folyamatra, megmutatjuk, hogy

$$\limsup_{n\to\infty}\frac{\sup_{a\le t\le b}|\gamma_n^{(r)}(t)|}{(\log\log n)^{1/2}} \le \frac{C}{2^{1/2}} \qquad \text{m.b.}$$

valamely $C=C(a,b,r)$ konstanssal, mely nem függ az X változó eloszlásától. Emellett megmutatjuk, hogy $\gamma_n^{(r)}$ relatív kompakt a $C[a,b]$ térben, valamint meghatározzuk a határfüggvények halmazát.

A 4.8. fejezetben az empirikus generátor folyamat bootstrap változatát vizsgáljuk, és bizonyítunk egy erős approximációs tételt a folyamatra. Ezen eredmény segítségével megmutatjuk, hogyan lehet konfidenciasávot konstruálni az X változó ismeretlen $g(t)$, $a\le t\le b$, valószínűségi generátorfüggvényéhez. Az utolsó fejezetben, egy paraméteres eloszláscsaládot alapul véve, definiáljuk a becsült paraméteres valószínűségi generátor folyamatot, valamint ennek paraméteres és nemparaméteres változatát. A 3. fejezet eredményeit alkalmazva gyenge approximációt bizonyítunk a folyamatokra, és vázoljuk, hogy a folyamatok hogyan alkalmazhatóak illeszkedési hipotézisek tesztelésére.

Acknowledgments

I am heartily thankful to my late supervisor Professor Sándor Csörgő for introducing me to probability theory and for guiding me in mathematical research at the beginning of my academic career. He was an example to me both as a great person and talented mathematician. I like to thank to my colleague Professor Gyula Pap for lecturing the earlier versions of my thesis. He highlighted many mathematical inaccuracies and grammatical mistakes in the manuscript. Special thanks to my mother for her love and encouragement. She supports my plans and provides me peaceful background for more than thirty years. Finally, many thanks to my friends for the happy hours at movie screenings, sport events and in smoky alehouses. Without them this thesis would be completed much earlier, but my life would be a dreadful bore.

Bibliography

Ash, R.B. and Gardner, M.F. (1975) Topics in Stochastic Processes. Academic Press, New York.

Babu, G.J. and Rao, C.R. (2004) Goodness-of-fit tests when parameters are estimated. *Sankhyā: The Indian Journal of Statistics* **66** 63–74.

Babu, G.J., Singh, K. (1984) Asymptotic representations related to jackknifing and bootstrapping L-statistics. *Sankhyā: The Indian Journal of Statistics* **46** 195–206.

Baringhaus, L., Gürtler, N. and Henze, N. (2000) Weighted integral test statistics and components of smooth tests of it. *Australian & New Zealand Journal of Statistics* **42** 179–192.

Baringhaus, L. and Henze, N. (1992) A goodness of fit test for the Poisson distribution based on the empirical generating function. *Statistics & Probability Letters* **13** 269–274.

Bartko, J.J. (1966) Approximating the negative binomial. *Technometrics* **8** 345–350.

Bickel, P.J., Freedman, D.A. (1981) Some asymptotic theory for the bootstrap. *The Annals of Statistics* **9** 1196–1217.

Billingsley, P. (1968) Convergence of Probability Measures. John Wiley & Sons, New York.

Bretagnolle, J. and Massart, P. (1989) Hungarian constructions from the nonasymptotic viewpoint. *The Annals of Probability* **17** 239–256.

Brillinger, D.R. (1969) An asymptotic representation of the sample distribution function. *Bulletin of the American Mathematical Society* **75** 545–547.

Burke, M.D., Csörgő, M., Csörgő, S. and Révész, P. (1979) Approximations of the empirical process when parameters are estimated. *The Annals of Probability* **7** 790–810.

Burke, M.D. and Gombay, E. (1988) On goodness-of-fit and the bootstrap. *Statistics & Probability Letters* **6** 287–293.

Burke, M.D. and Gombay, E. (1991) The bootstrapped maximum likelihood estimator with an application. *Statistics & Probability Letters* 12: 421–427.

Castelle, N. and Laurent-Bonvalot, F. (1998) Strong approximations of bivariate uniform empirical processes. *Annales de l'Institut Henri Poincaré: Probabilités at Statistique* **34** 425–480.

Csörgő, M., Csörgő, S., Horváth, L. and Mason, D.M. (1986) Supnorm convergence of the empirical process indexed by functions and applications. *Probability and Mathematical Statistics* **7** 13–26.

Csörgő, M. and Révész, P. (1975) A strong approximation of the multivariate empirical process. *Studia Scientiarum Mathematicarum Hungarica* **10** 427–434.

Csörgő, M. and Révész, P. (1981) Strong Approximation in Probability and Statistics. Academic Press and Akadémiai Kiadó, New York and Budapest.

Csörgő, S. (1976) On an asymptotic expansion for the Mises ω^2 statistic. *Acta Scientarium Mathematicarum* **38** 45–67.

Csörgő, S. (1981) Limit behaviour of the empirical characteristic function. *The Annals of Probability* **9** 130–144.

Csörgő, S. and Mason, D.M. (1989) Bootstraping empirical functions. *The Annals of Statistics* **17** 1447–1471.

Csörgő, S. and Rosalsky, A. (2003) A survey of limit laws for bootstrapped sums. *International Journal of Mathematics and Mathematical Sciences* **45** 2835–2861.

Dowling, M.M. and Nakamura, M. (1997) Estimating parameters for discrete distributions via the empirical probability generating function. *Communications in Statistics: Simulation and Computation* **26** 301–313.

Durbin, J. (1973) Weak convergence of the sample distribution, when parameters are estimated. *The Annals of Statistics* **1** 279–290.

Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* **7** 1–26.

Efron, B., Tibshirani, R.J. (1993) An Introduction to the Bootstrap. Chapman & Hall, New York.

Finkelstein, H. (1971) The law of the iterated logarithm for empirical processes. *The Annals of Mathematical Statistics* **42** 607–615.

Gan F.F. and Koehler K.J. (1990) Goodness-of-fit tests based on P-P probability plots. *Technometrics* **32** 289–303.

Genz, M. and Haeusler, E. (2006) Empirical processes with estimated parameters under auxiliary information. *Journal of Computational and Applied Mathematics* **186** 191–217

Gürtler, N. and Henze., N. (2000) Recent and classical goodness-of-fit tests for the Poisson distribution. *Journal of Statistical Planning and Inference* **90** 207–225.

Hall, P. (1992) The bootstrap and Edgeworth expansion. Springer-Verlag, New York

Henze, N. (1996) Empirical-distribution-function goodness-of-fit tests for discrete models. *The Canadian Journal of Statistics* **24** 81–93.

Jacod, J. and Shiryaev, A.N. (2003) Limit Theorems for Stochastic Processes. Springer-Verlag, Berlin.

Karatzas, I. and Shreve, E. (1988) Brownian Motion and Stochastic Calculus. Springer-Verlag, New York.

Kemp, C.D. and Kemp, A.W. (1988) Rapid estimation for discrete distributions. *The Statistician* **37** 243–255.

Kiefer, J. (1972) Skorohod embedding of multivariate R.V.'s and the sample D.F. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **24** 1–35.

Komlós, J., Major, P. and Tusnády, G. (1975) An approximation of partial sums of independent R.V.'s, and the sample D.F.I. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **32** 111–131.

Komlós, J., Major, P. and Tusnády, G. (1976) An approximation of partial sums of independent R.V.'s, and the sample D.F.II. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **34** 33–58.

Kocherlakota, S. and Kocherlakota, K. (1986) Goodness of fit tests for discrete distributions. *Communications in Statistics: Theory and Methods* **15** 815–829.

Major, P. (1988) On the tail behaviour of the distribution function of multiple stochastic integrals. *Probability Theory and Related spaces* **78** 419–435.

Marques, M.S. and Pérez-Abreu, V. (1989) Law of large numbers and limit theorem for the empirical probability generating function of stationary random sequences and processes. *Aportaciones Matematicas* **4** 100–109.

Nakamura, M. and Pérez-Abreu, V. (1993) Exploratory data analysis for counts using the empirical probability generating function. *Communications in Statistics: Theory and Methods* **22** 827–842.

Nakamura, M. and Pérez-Abreu, V. (1993) Use of the empirical probability generating function for testing a Poisson Model. *The Canadian Journal of Statistics* **21** 149–156.

Philipp, W. and Pinzur, L. (1980) Almost sure approximation theorems for the multivariate empirical process. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **54** 1–13.

Pollard, D. (1984) Convergence of stochastic processes. Springer-Verlag, New York.

Rémillard, B. and Theodorescu, R. (2000) Inference based on the empirical probability generating function for mixtures of Poisson distributions. *Statistics & Decisions* **18** 349–366.

Rueda, R., Pérez-Abreu, V. and O'Reilly, F. (1991) Goodness of fit for the Poisson distribution based on the probability generating function. *Communications in Statistics: Theory and Methods* **20** 3093-3110. [Corrigendum: (1992) **21** 3623.]

Shorack, G.R., Wellner, J.A. (1986) Empirical Processes with Applications to Statistics. John Wiley & Sons, New York.

Szűcs, G. (2005) Approximations of empirical probability generating processes. *Statistics & Decisions* **23** 67–80.

Szűcs, G. (2008) Parametric bootstrap tests for discrete distributions. *Metrika* **67** 63–81.

Szűcs, G. (20??) Non-parametric bootstrap tests for parametric distribution families. *Acta Scientarium Mathematicarum*, accepted.

Stute, W., Manteiga, W.G. and Quindimil, M.P. (1993) Bootstrap based goodness-of-fit-tests. *Metrica* **40** 243–256.

Tsirel'son, V.S. (1975) The density of the maximum of a Gaussian process. *Theory of Probability and its Applications* **20** 817–856.

van der Vaart, A.W. (1998) Asymptotic Statistics. Cambridge University Press, Cambridge.