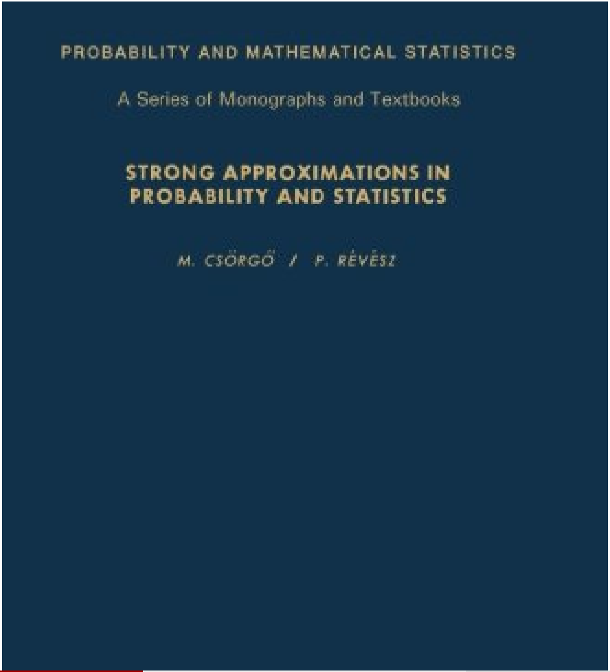
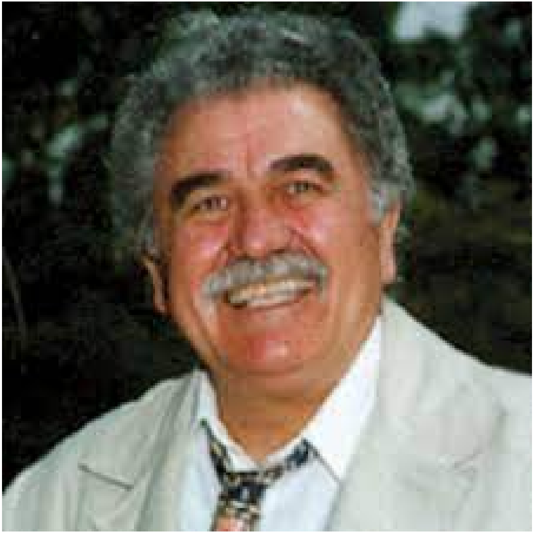
A review on the strong Gaussian approximation of empirical processes and its applications

V. Fakoor
Department of Statistics, Ferdowsi University of Mashhad.
23th Workshop on Applied Stochastic Process; Shiraz University

Contents

- Definition
- Motivation and history
- the Komlós, Major, Tusnády (KMT) approximation
- Extensions
- Application: Nonparametric density estimation


PROBABILITY AND MATHEMATICAL STATISTICS
A Series of Monographs and Textbooks
STRONG APPROXIMATIONS IN PROBABILITY AND STATISTICS
M. CSÖRGŐ / P. RÉVÉSZ



Definitions

Wiener processes

A Gaussian process $\{W(y); 0 \leq y \leq \infty, \}$ if $EW(y) = 0$ and $EW(y_1)W(y_2) = y_1 \wedge y_2$

Brownian bridge

A Gaussian process $\{B(y); 0 \leq y \leq 1, \}$ is called a Brownian bridge if $EB(y) = 0$ and $EB(y_1)B(y_2) = y_1 \wedge y_2 - y_1.y_2$

- $B(0) = B(1) = 0$

Kiefer process

A centered Gaussian process $\{K(y, t); 0 < y < 1, 0 < t < \infty\}$ with covariance function $E(K(s, t)K(s, t)) = (t \wedge t\prime)(s \wedge s\prime - ss\prime)$

Definition

Strong approximations (strong invariance principle)

- Wiener approximation for the partial sums of i.i.d. random variables

$$S_n = W_n + O(\log n) \quad a.s.$$

- Approximation of the empirical process by a Brownian bridge

Motivation and history

Two independent source

• Erdos and Kac (1946), On certain limit theorems of the theory of probability. Bull. Amer. Math. Soc. 52 292-302.
• Doob (1949), entitled "Heuristic approach to the Kolmogorov-Smirnov theorems". Ann. Math. Statist. 20 393-403.

First origin

Classical central limit theorem

Let $X_1, \ldots, X_n$ $i.i.d \sim F(\cdot), EX_i = 0, EX_i^2 = 1, S_n = \sum_{i=1}^n X_i$.
Let $Y_1, \ldots, Y_n$ $i.i.d \sim N(0,1)$, $T_n = \sum_{i=1}^n Y_i$

$$P(\frac{1}{\sqrt{n}} S_n \leq x) \longrightarrow \phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} exp\{-t^2/2\} dt = P(\frac{1}{\sqrt{n}} T_n \leq x)$$

$$P(\frac{1}{\sqrt{n}} S_n \leq x) - P(\frac{1}{\sqrt{n}} T_n \leq x) \longrightarrow 0$$

• as time goes on, S_n forgets about the distribution function F where it has come from.

Erdos and Kac (1946): (weak) invariance principle

$$G_1(y) = \lim_{n\to\infty} P\left(n^{-1/2} \max_{1\leqq k\leqq n} S_k \leqq y\right),$$

$$G_2(y) = \lim_{n\to\infty} P\left(n^{-1/2} \max_{1\leqq k\leqq n} |S_k| \leqq y\right),$$

$$G_3(y) = \lim_{n\to\infty} P\left(n^{-2} \sum_{k=1}^{n} S_k^2 \leqq y\right),$$

$$G_4(y) = \lim_{n\to\infty} P\left(n^{-3/2} \sum_{k=1}^{n} |S_k| \leqq y\right).$$

• They proved that the limit distributions (i)-(iv) exist and they do not depend on the initial distribution of X_1. They called this method of proof the (weak) invariance principle, and their paper has initiated a new methodology for proving limit laws in probability theory.

Donsker (1951)

$$h(S_n(\cdot)) \xrightarrow{\mathcal{W}} h(W(\cdot))$$

for every continuous functional $h : C(0,1) \longrightarrow \mathbb{R}$

$$S_n(t) = \frac{1}{\sqrt{n}}\{S_{[nt]} + X_{[nt]+1}(nt - [nt])\}$$

Strassen (1964): new form of the (strong) invariance principle

Theorem 0.2*.

$$\sup_{0\leq t\leq 1} \frac{|S_n(t) - n^{-1/2}W(nt)|}{\sqrt{\log\log n}} \xrightarrow{\text{a.s.}} 0. \tag{0.9*}$$

Connection

• The precise connection between weak and strong invariance principles was established by Strassen (1965a) (cf. also Dudley (1968) and Wichura (1970)) via the so-called Prohorov distance of probability measures. In fact these results state a kind of equivalence between these two forms of invariance.

Second origin

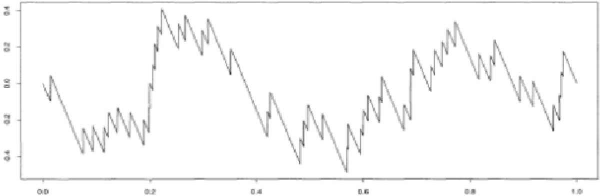
Let $U_1, U_2, \ldots$ be independent uniform $(0,1)$ random variables. Define a uniform empirical distribution function as

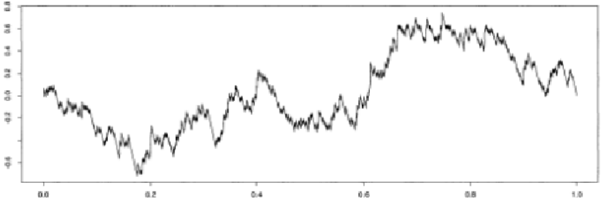
$$F_{U,n}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{U_i \le t}, \quad t \in [0,1]$$

Define a uniform empirical process as

$$\alpha_n(t) = \sqrt{n}\left(F_{U,n}(t) - t\right), \quad t \in [0,1].$$

Uniform, 50, 500





Second origin

- Doob (1949): "Heuristic approach to the Kolmogorov-Smirnov theorems".

Second origin

- Doob (1949): "Heuristic approach to the Kolmogorov-Smirnov theorems".
- Donsker (1952): $\alpha_n \xrightarrow{\mathcal{W}} B(\cdot)$

Second origin

- Doob (1949): "Heuristic approach to the Kolmogorov-Smirnov theorems".
- Donsker (1952): $\alpha_n \xrightarrow{\mathcal{W}} B(\cdot)$
- Prohorov (1956) and Skorohod (1956) $h(\alpha_n) \xrightarrow{\mathcal{W}} h(B(\cdot))$

Second origin

- Doob (1949): "Heuristic approach to the Kolmogorov-Smirnov theorems".
- Donsker (1952): $\alpha_n \xrightarrow{\mathcal{W}} B(\cdot)$
- Prohorov (1956) and Skorohod (1956) $h(\alpha_n) \xrightarrow{\mathcal{W}} h(B(\cdot))$
- Brillinger (1969): first SIP

$$\sup_{0\le t\le 1} |\tilde{\alpha}_n(t) - B_n(t)| = O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}), \quad a.s.$$

Second origin

- Doob (1949): "Heuristic approach to the Kolmogorov-Smirnov theorems".
- Donsker (1952): $\alpha_n \xrightarrow{\mathcal{W}} B(\cdot)$
- Prohorov (1956) and Skorohod (1956) $h(\alpha_n) \xrightarrow{\mathcal{W}} h(B(\cdot))$
- Brillinger (1969): first SIP

$$\sup_{0 \le t \le 1} |\tilde{\alpha_n}(t) - B_n(t)| = O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}), \quad a.s.$$

- Kiefer (1972).

$$\sup_{0 \le y \le 1} |\sqrt{n}\alpha_n(y) - K(y, n)| = O(n^{1/3}(\log n)^{2/3}) \quad a.s.$$

Komlós, Major, Tusnády approximation

• the KMT embedding
• the Hungarian embedding
• Komlós, J., Major, P. Tusnády G. (1975). An approximation of partial sums of independent RV'-s, and the sample DF. I. Z. Wahrsch. Verw. Gebiete 32 111-131.
• sharp bound for the speed of this weak convergence .

the KMT approximation

Theorem

For EP $\{\alpha_n(t); 0 \le t \le 1\}$ there exists a probability space on which one can define a Brownian bridge $\{B_n(t); 0 \le t \le 1\}$ for each n and a Kiefer process $\{K(y,t); 0 \le y \le 1, t \ge 0\}$ such that

$$P\left\{\sup_{0\le t\le 1} |\alpha_n(t) - B_n(t)| > \frac{1}{\sqrt{n}}(C\log n + x)\right\} \le Le^{-\lambda x}$$

$$P\left\{\sup_{0\le k\le n}\sup_{0\le y\le 1} |k^{1/2}\alpha_k(y) - K(y,k)| > \frac{1}{\sqrt{n}}(C\log n + x)\log n\right\} \le Le^{-\lambda x}$$

for all $x \in \mathbb{R}$, where C, L and λ are positive constants

Corollary

$$\sup_{0\leq t\leq 1} |\alpha_n(t) - B_n(t)| = O(\frac{\ln n}{\sqrt{n}}), \quad a.s.$$

$$\sup_{0\leq y\leq 1} |\alpha_n(y) - K(y,n)/n^{1/2}| = O(\frac{(\ln n)^2}{\sqrt{n}}), \quad a.s.$$

Corollary

$$\sup_{0\le t\le 1} |\alpha_n(t) - B_n(t)| = O(\frac{\ln n}{\sqrt{n}}), \quad a.s.$$

$$\sup_{0\le y\le 1} |\alpha_n(y) - K(y,n)/n^{1/2}| = O(\frac{(\ln n)^2}{\sqrt{n}}), \quad a.s.$$

• Bretagnolle and Massart (1989) provided explicit constants $C = 12, L = 2, \lambda = 1/6$.

Results

• Weak limit theorems, as Donsker's invariance principle for the empirical distribution function

Results

- Weak limit theorems, as Donsker's invariance principle for the empirical distribution function
- Almost sure results, as the functional form of the law of the iterated logarithm

Results

- Weak limit theorems, as Donsker's invariance principle for the empirical distribution function
- Almost sure results, as the functional form of the law of the iterated logarithm
- From a statistical point of view, strong approximations with rates allow to construct many statistical procedures

Extensions

• Berkes and Philipp (1977) : Strong approximations for the empirical process with dependent data (strongly mixing),
• Kuelbs (1973), J. Hoffman-Jorgensen-G. Pisier (1976), Garling (1976): higher dimensional Euclidean space or a Banach space
• Random size partial sum
• Csörgő , M. and Revesz , P. (1978). Strong approximation of the quantile process. Ann. Statist. 6 882–894
• Burke, Csörgő and Horváth (1981, 1988): Random censorship model strong approximation for the product-limit process $Z_n(t) := \sqrt{n}[\widehat{F}_n(t) - F(t)]$.

Applications

- Nonparametric density estimation
- nonparametric regression estimation
- characteristic functions
- mean residual lifetime processes
- empiric total-time-on-test,
- Lorenz, concentration, and related processes

can be approximated by appropriate Gaussian processes.

Application: Nonparametric density estimation

• Let $X_1, \ldots, X_n$ iid f, the kernel estimate f_n of f introduced by Rosenblatt (1956) and defined by

$$f_n(t) = \sum_{i=1}^{n} \frac{1}{nh_n} K\left(\frac{t - X_i}{h_n}\right) = \frac{1}{h_n} \int_0^{\infty} K\left(\frac{t - s}{h_n}\right) dF_n(s),$$

K is a kernel function, and h_n is a sequence of (positive) “bandwidths” tending to zero as $n \to \infty$.

Application: Nonparametric density estimation

• Parzen (1962): consistent estimator

• Nadaraya (1965), Schuster (1969) and Van Ryzin (1969): The weak and strong uniform consistency properties of f_n. Condition placed on the bandwidth for strong uniform consistency include $\sum \exp(-cnh_n{}^2) < \infty$ for all positive c.

• Bickel and Rosenblatt (1971) : strong approximation(Brillinger (1969)) predate the development of the KMT approximation,

• Revesz (1976b) and Rosenblatt (1976) used the KMT approximation to derive asymptotically distribution-free confidence bands for the expected value of $f_n(x)$, improving on the earlier results of Bickel Rosenblatt.

• Silverman (1978): strong uniform consistency for $f_n - f$ using the KMT approximation. weak condition

$$\frac{\log n}{nh_n} \longrightarrow 0 \quad as \quad n \longrightarrow \infty$$

Nonparametric density estimation

$$
\begin{aligned}
f_n(t) - Ef_n(t) &= \frac{1}{h_n}\int K(\frac{t-s}{h_n})d[F_n(x) - F(x)] \\
&= \frac{1}{h_n}\int [Fn(x) - F(x)]dK(\frac{t-s}{h_n}) \\
&\overset{a.s.}{=} -\frac{1}{\sqrt{n}}\int_0^\infty B_n(F(x))dK(\frac{t-s}{h_n}) + O\left(\frac{\log n}{nh_n}\right)
\end{aligned}
$$

Random Censorship Model

• Blum and Susarla (1980):

$$f_n(t) = \frac{1}{h_n} \int_0^\infty K(\frac{t-s}{h_n}) d\widehat{F}_n(s), \quad (1)$$

• The properties of the kernel estimator f_n have been examined by Blum and Susarla (1980), Földes, Rejtö and Winter (1981) and Mielniczuk (1986), among others.
• Zhang (1998) established the strong uniform consistency for $f_n - f$ using the strong approximation technique developed by Burke, Csörgő and Horváth (1981, 1988) for the product-limit process $Z_n(t) := \sqrt{n}[\widehat{F}_n(t) - F(t)]$.

Thank you for your attention