Accuracy of approximations in limit theorems

VYDAS ČEKANAVIČIUS
Vilnius University
Email: vydas.cekanavicius@maf.vu.lt

July 15, 2003

Introduction

Approximation

In short, the aim of this course is to learn how to estimate the closeness of distribution of the sum of random variables and its approximation:

$$P(S_n \in B) - W\{B\} \approx ???.$$

In this course, the general principles of construction of approximations (which might be distributions or signed measures) are given. However, comparative analysis of approximations is beyond the scopes of this course. The main emphasis is on the methods needed for estimating and their practical usage. You'll be given a 'toolkit' which can be applied in various fields (for example, when writing master or PhD thesis).

The standard three stages in investigation of the behavior of S_n:

1. Finding of the limiting distribution for S_n, say D.
2. Establishing of the rate of convergency of $\mathcal{L}(S)$ to D.
3. Investigation what approximations to use.

Note: the best approximation unnecessary coincides with the limit distribution.
We shall learn methods useful in the second and third situations.

Accuracy

The usual estimates of the accuracy are in the form:

1. Of the type $O(n^{-s})$.
2. Of the type Cn^{-s}.

Methods

The main methods of this course are related to the characteristic function method. In general, this method means that we use

$$|\,P(S_n \in B) - W\{B\}\,| \leqslant C \int\limits_{t\in A} f(t)|\,\widehat{F}(t) - \widehat{W}(t)\,|\,\mathrm{d}t + U(F, W, A).$$

In the right-hand side we have the difference of the characteristic functions (Fourier-Stieltjes transforms), $f(t)$ is an additional (smoothing or truncating) factor allowing estimation in the neighborhood of zero and U is the remainder. Characteristic functions have many nice properties.

Incomplete list of references

1. Arak, T.V. (1980) Approximation of n-fold convolutions of distributions, having a nonnegative characteristic function, with accompanying laws. *Theory Prob. Appl.* **25,** 221–243.

2. Arak, T.V.(1981) On the convergence rate in Kolmogorov's uniform limit theorem I,II *Theory Prob. Appl.* **26**, 219–239; 437–451.

3. Arak, T.V. and Zaĭtsev, A.Yu. (1988) Uniform limit theorems for sums of independent random variables. *Proc. Steklov Inst. Math.* **174,** 1–222.

4. Barbour, A.D. and Jensen, J.L. (1989) Local and tail approximations near the Poisson limit. *Scand. J. Statist.* **16**, 75-87.

5. Barbour, A.D. Holst, L. and Janson, S. (1992). *Poisson approximations.* Oxford. Clarendon Press.

6. Barbour, A.D. and Xia, A. (1999) Poisson perturbations. *ESAIM: P&S* **3,** 131–150.

7. Barbour, A.D. and Chryssaphinou, O. (2001) Compound Poisson approximation: a user's guide. *Ann. Appl. Prob.* **11**, 964–1002.

8. Barbour, A.D. and Čekanavičius, V. (2002) Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.* **30**, no 2, 509–545.

9. Bergström, H. (1951) On asymptotic expansion of probability functions. *Skand. Aktuar.* **1,** 1–34.

10. Brown, T.C. and Xia, A. (2001) Stein's method and birth-death processes. *Ann. Probab.* **29**, 1373-1403.

11. Čekanavičius, V. (1993) Non-uniform theorems for discrete measures. *Lithuanian Math. J.* **33**, 114–126.

12. Čekanavičius, V. (1995) On smoothing properties of Compound Poisson distibutions. *Lithuanian Math. J.* **35**, 121–135.

13. Čekanavičius, V. (1997) Asymptotic expansions for compound Poisson measures. *Lithuanian Math. J.* **37**, 1997, 426–447.

14. Čekanavičius, V. (1998) Estimates in total variation for convolutions of compound distributions. *J. London Math. Soc. (2)* **58**, 748–760.

15. Čekanavičius, V. and Kruopis, J. (2000) Signed Poisson approximation: a possible alternative to normal and Poisson laws. *Bernoulli*, **6(4)**, 591–606.

16. Čekanavičius, V. and Wang, Y.H. (2003) Compound Poisson approximations for sums of discrete non-lattice variables. *Adv. Appl. Prob.* **35**(1), 228–250.

17. Chen, L.H.Y. (1975) Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534-545.

18. Chen, L.H.Y. (1998) Stein's method: some perspectives with applications. *Probability towards 2000* (Eds. L. Accardi and C. Heyde), p.p. 97-122. Lecture Notes in Statistics No. 128.

19. Deheuvels, P. and Pfeifer, D. (1986) A semigroup approach to Poisson approximation. *Ann. Prob.* **14,** 663–676.

20. Deheuvels, P., Karr, A., Pfeifer, D. and Serfling, R. (1988) Poisson approximations in selected metrics by coupling and semigroup methods with applications. *J. Statist. Plann. Inference*, **22**, 1–22.

21. Franken, P. (1964) Approximation der Verteilungen von Summen unabhängiger nicht-negativen ganzzahliger Zufallsgrösen durch Poissonsche Verteilungen. *Math. Nachr.* **27,** 303–340.

22. Hipp, C. (1986) Improved approximations for the aggregate claims distribution in the individual model. *ASTIN Bull.* **16,** 89–100.

23. Kolmogorov A.N. (1963) On the approximation of distributions of sums of independent summands by infinitely divisible distributions. *Sankhya, Ser.A*, **25**, 159–174.

24. Kerstan, J. (1964) Verallgemeinerung eines Satzes von Prochorov und Le Cam. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **2**, 173–179.

25. Kruopis, J. (1986a) Precision of approximations of the generalized Binomial distribution by convolutions of Poisson measures. *Lithuanian Math. J.* **26**, 37–49.

26. Kruopis, J. (1986b) Approximations for distributions of sums of lattice random variables I. *Lithuanian Math. J.* **26,** 234–244.

27. Le Cam, L. (1965) On the distribution of sums of independent random variables. In: *Bernoulli, Bayes, Laplace (Anniversary volume).* Springer-Verlag, Berlin, Heidelberg,New York. 179–202.

28. Michel, R. (1988) An improved error bound for the compound Poisson approximation of a nearly homogeneous portfolio. *ASTIN Bull.* **17,** 165–169.

29. Petrov, V.V. (1995) *Limit Theorems of Probability Theory.* Clarendon Press, Oxford.

30. Presman, E.L. (1983) Approximation of binomial distributions by infinitely divisible ones. *Theory Prob. Appl.* **28,** 393–403.

31. Prohorov, Yu.V. (1953) Asymptotic behaviour of the binomial distribution. (Russian), *Uspekhi Matem. Nauk.* **8,** 3, 135–142. Engl. transl. in *Select. Transl. Math. Statist. and probability.***1**, 87–95 (1961).

32. Prokhorov, Yu.V. (1955) On the sums of identically distributed random variables. *Doklady Akad.Nauk SSSR* **105**, 645–647 (Russian).

33. Roos, B. (1995) A semigroup approach to Poisson approximation with respect to the point metric. *Statist. Probab. Letters*, **24**, 305-314.

34. Roos, B. (1999) Asymptotics and sharp bounds in the Poisson approximation to the Poisson-binomial distribution. *Bernoulli* **5**(6), 1021–1034.

35. Roos, B. (2002a) Kerstan's method in the multivariate Poisson approximation: an expansion in the exponent. *Teorya Veroyatn. Primen.* **47**(2), 397–401.

36. Roos, B. (2002b) Kerstan's method for compound Poisson approximation. *Preprint No 2002-04, Schwerpunkt Mathematische Statistik und Stochastik Prozesse, Universitaet Hamburg.*(Accepted by Ann. Probab.)

37. Šiaulys, J. and Čekanavičius, V. (1988) Approximation of distributions of integer-valued additive functions by discrete charges. I. *Lithuanian Math. J.* **28**, 392–401.

38. Studnev, Yu.P. (1960) An approximation to the distribution of sums by infinitely divisible laws. *Theory Probab. Appl.* , **5**, 421-425.

39. Tsaregradskii, I.P. (1958) On uniform approximation of a binomial distribution by infinitely divisible laws. *Teorya Veroyatn. Primen.* **3**, 470–474 (Russian).

40. Zaĭtsev, A.Yu. (1984). On the accuracy of approximation of distributions of sums of independent random variables-which are nonzero with a small probability by accompanying laws. *Theory Prob. Appl.* **28,** 657–669.

41. Zaĭtsev, A.Yu. (1992) On the approximation of convolutions of multidimensional symmetric distributions by accompanying laws. *J. Soviet. Math.* **61,** 1859–1872

42. Zolotarev, V. (1965)On the closeness of the distributions of two sums of independent random variables *Teorya Veroyatn. Primen.* **10**, 519–526 (Russian).

(To be continued)

1 Definitions and preliminary facts

1.1 Useful identities and inequalities

Bergström identity for commutative objects:

$$a^n \;=\; b^n + \sum_{m=1}^{s} \binom{n}{m} b^{n-m}(a-b)^m \;+\; r_n(s+1). \tag{1.1}$$

Here

$$r_n(s+1) \;=\; \sum_{m=s+1}^{n} \binom{m-1}{s} a^{n-m}(a-b)^{s+1} b^{m-s-1}. \tag{1.2}$$

Note that

$$\sum_{m=s+1}^{n} \binom{m-1}{s} = \binom{n}{s+1}.$$

Bergström identity can be extended to products. Particularly,

$$\prod_{k=1}^{n} a_k - \prod_{k=1}^{n} b_k = \sum_{k=1}^{n} (a_k - b_k) \prod_{j=1}^{k-1} a_j \prod_{j=k+1}^{n} b_j, \tag{1.3}$$

$$\prod_{k=1}^{n} a_k - \prod_{k=1}^{n} b_k - \sum_{k=1}^{n} (a_k - b_k) \prod_{j\neq k}^{n} b_j = \sum_{l=2}^{n} (a_l - b_l) \prod_{j=l+1}^{n} a_l \sum_{k=1}^{l-1} (a_k - b_k) \prod_{j=1, j\neq k}^{l-1} b_j. \tag{1.4}$$

Taylor's expansion:

$$f(z) = f(0) + f'(0)z + f''(0)\frac{z^2}{2!} + \ldots + f^{(k)}(0)\frac{z^k}{k!} + \frac{z^{k+1}}{k!}\int_0^1 f^{(k+1)}(\tau z)(1-\tau)^k \,\mathrm{d}\tau \tag{1.5}$$

and its special case:

$$\mathrm{e}^M = 1 + M + \frac{M^2}{2!} + \ldots + \frac{M^k}{k!} + \frac{M^{k+1}}{k!}\int_0^1 \mathrm{e}^{\tau M}(1-\tau)^k \,\mathrm{d}\tau. \tag{1.6}$$

We give one example of how (1.6) can be applied.

Example 1.1 *Let Re $a < 0$. Then*

$$|\,\mathrm{e}^a - 1\,| \leqslant |\,a\,|.$$

Indeed, we have

$$|\,\mathrm{e}^a - 1\,| = \Big|\, a\int_0^1 \mathrm{e}^{\tau a}\,\mathrm{d}t \Big| \leqslant |\,a\,| \int_0^1 |\,\mathrm{e}^{\tau a}\,|\,\mathrm{d}t = |\,a\,| \int_0^1 \mathrm{e}^{\tau Re\ a}\,\mathrm{d}t \leqslant |\,a\,|.$$

Abel's partial summation formula:

$$\sum_{k=M}^{N} a_k b_k = A_N b_N - \sum_{k=M}^{N-1} A_k (b_{k+1} - b_k). \tag{1.7}$$

Here

$$A_k = \sum_{m=M}^{k} a_m.$$

Some other facts. Let $k = 1, 2, \ldots,\ \lambda > 0$. Then

$$|\,\mathrm{e}^{\mathrm{i}t} - 1\,| = 2|\,\sin\frac{t}{2}\,|, \quad |\,\sin\frac{kt}{2}\,| \leqslant k\,|\,\sin\frac{t}{2}\,|,$$

$$(\mathrm{e}^{\mathrm{i}t} - 1) + (\mathrm{e}^{-\mathrm{i}t} - 1) = -(\mathrm{e}^{\mathrm{i}t} - 1)(\mathrm{e}^{-\mathrm{i}t} - 1),$$

$$|\,(\mathrm{e}^{-\mathrm{i}t} - 1)^k - (-1)^k(\mathrm{e}^{\mathrm{i}t} - 1)^k\,| \leqslant k|\,\mathrm{e}^{\mathrm{i}t} - 1\,|^{k+1},$$

$$\int_{-\pi}^{\pi} |\,\sin\frac{t}{2}|^k \exp\Big\{-2\lambda\sin^2\frac{t}{2}\Big\}\,\mathrm{d}t \leqslant 2\sqrt{\mathrm{e}}\Big(1 + \sqrt{\frac{\pi}{2}}\Big)\pi\Big(\frac{\max(k, 1/3)}{4\lambda\mathrm{e}}\Big)^{(k+1)/2} \leqslant C(k)\lambda^{-(k+1)/2}. \tag{1.8}$$

If $|\,t\,| \leqslant \pi$, then $|\,\sin(t/2)\,| \geqslant |\,t\,|/\pi$. We shall repeatedly use the following simple estimates:

$$|\,a^n - b^n\,| \leqslant n|\,a - b\,|\max\{|\,a\,|^{n-1}, |\,b\,|^{n-1}\}, \tag{1.9}$$

$$\sum_{k\in\mathbb{Z}} \Big(1 + \Big(\frac{k-a}{b}\Big)^2\Big)^{-1} \leqslant 1 + b\pi, \tag{1.10}$$

$$\sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}(k+1)^{-1/r} \leqslant p^{-1/r}(n+1)^{-1/r}, \quad (0 < p < 1,\ n \geqslant 1,\ r \in [1, \infty]), \tag{1.11}$$

$$x^\alpha \mathrm{e}^{-x} \leqslant \alpha^\alpha \mathrm{e}^{-\alpha} \quad (x > 0, \alpha > 0). \tag{1.12}$$

Finally, note that $1 + x \leqslant \mathrm{e}^x$, $(x \in \mathbb{R})$.

1.2 Distributions and measures

Let $\mathbb{R}$ denote the set of real numbers, $\mathbb{Z}$ denote the set of all integers, $\mathbb{Z}_+ = \mathbb{Z} \cup \{0\}$, and let $\mathbb{N} = \{1, 2, \ldots\}$. The symbol C denotes (in general different) absolute positive constants. Similarly, by $C(\cdot)$ we denote constants depending on the indicated argument only. We always use θ to denote a quantity satisfying $|\,\theta\,| \leqslant 1$.

Let $\mathcal{M}$ denote the set of all finite signed measures defined on σ- field $\mathcal{B}$ of one-dimensional Borel subsets; $\mathcal{M}_Z \subset \mathcal{M}$ denote the set of all finite signed measures concentrated on $\mathbb{Z}$; $\mathcal{F} \subset \mathcal{M}$ denote the set of all distributions; $\mathcal{F}_+$ denotes the set of all distributions having nonnegative characteristic functions, $\mathcal{F}_Z$ denote the set of all distributions concentrated on $\mathbb{Z}$; $E_a \in \mathcal{F}$ denote the distribution concentrated at a point $a \in \mathbb{R}$, with $E \equiv E_0$. Let $U \in \mathcal{F}$. Then $U^{(-)}$ denotes distribution, for any Borel set X, satisfying $U^{(-)}\{X\} = U\{-X\}$. Similarly $U^{(2)}\{X\} = U\{X/2\}$. By $[X]_\tau$ we denote a closed τ - neighborhood of the set X. For the distribution of random variable ξ we also use notation $\mathcal{L}(\xi)$.

1.2.1 Convolutions

All products and powers of signed measures are defined **in the convolution sense**, that is,

$$FG\{A\} = \int_{\mathbb{R}} F\{A-x\}\, G\{\mathrm{d}x\}, \qquad F^0 \equiv E.$$

Note that $(E_1)^k = E_k$.

Convolution of distributions is distribution of the sum of independent random variables. For example, F^n is distribution of $\xi_1+\xi_2+\ldots+\xi_n$, where all ξ_j are independent and have the same distribution F.

Let $0 \leqslant p \leqslant 1$, $F, G \in \mathcal{F}$. Then $pF+(1-p)G \in \mathcal{F}$. This property can be extended to the case of more than two distributions.

By the exponential of $W \in \mathcal{M}$ we call

$$\exp\{W\} = \sum_{m=0}^{\infty} \frac{W^m}{m!}. \tag{1.13}$$

Exponential measures have some nice properties, for example, $\exp\{W\}\exp\{V\} = \exp\{V+W\}$, $\exp\{aW\}\exp\{bW\} = \exp\{(a+b)W\}$.

Convolutions allow to write some distributions in a convenient way. Fore example, the Binomial distribution $Bi(n,p)$ can be written as:

$$((1-p)E+pE_1)^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^k (E_1)^k = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^k E_k.$$

Similarly, the Poisson distribution with parameter λ can be written in the following way:

$$\exp\{\lambda(E_1-E)\} = \sum_{k=0}^{\infty} \frac{\lambda^k (E_1-E)^k}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} E_k.$$

Special attention will be played to discrete measures concentrated on $\mathbb{Z}$. Let $W, V \in \mathcal{M}_Z$. Then

$$W = \sum_{j=-\infty}^{\infty} W\{j\}E_j = \sum_{j=-\infty}^{\infty} W\{j\}E_1^j; \quad WV\{k\} = \sum_{j=-\infty}^{\infty} W\{k-j\}V\{j\}. \tag{1.14}$$

1.2.2 Compound distributions

Nowadays Compound distributions attract a lot of attention, mainly due to the needs of actuarial mathematics. In our course, we shall pay a special interest to compound distributions and principles of their approximation. By **compound** (signed) measure we understand

$$\varphi(F) = \sum_{m=0}^{\infty} p_m F^m, \quad \text{where} \quad F \in \mathcal{F}, \quad \sum_{m=0}^{\infty} |\, p_m\,| < \infty. \tag{1.15}$$

If $p_0+p_1+p_2+p_3+\ldots = 1$, $0 \leqslant p_j \leqslant 1$, then $\varphi(F)$ is compound distribution. Any compound distribution corresponds to the *random* sum of independent random variables $\xi_1+\xi_2+\ldots+\xi_\eta$, when all ξ_j have the same distribution F and η is independent of ξ_j and has distribution $P(\eta=k)=p_k$, $k=0,1,\ldots$.

Example 1.2 *Let us consider a simplified version of the insurance claims occurrence.Let us assume that claim occurs with the probability p and the amount of claim is determined by the distribution B. Then the aggregate claims distribution of n individuals is equal to $((1-p)E+pB)^n$.*

Similarly, assuming that probabilities for claims occurrence and distributions of claims differ from individual to individual, we get the aggregate claims distribution equal to $\prod_{i=1}^{n} H_i$, where $H_i = (1-p_i)E + p_iB_i$ and, H_i is the distribution of risk i, p_i is the probability that risk i produces a claim, B_i is distribution of the claim in risk i, given the claim occurrence in risk i.

Example 1.3 *Examples of compound distributions:*

a)Compound Poisson distribution. Let $F \in \mathcal{F}$, $\lambda \geqslant 0$, then Compound Poisson distribution is defined by

$$\exp\{\lambda(F-E)\} = \sum_{m=0}^{\infty} \frac{\lambda}{m!} \mathrm{e}^{-\lambda} F^m = \sum_{m=o}^{\infty} \frac{\lambda^m (F-E)^m}{m!}. \tag{1.16}$$

Note that $\exp\{\lambda(F-E)\}$ *is a direct generalization of the Poisson law.*

b) Compound Geometric distribution. Let $F \in \mathcal{F}$, $0 < p \leqslant 1$ then Compound Geometric is defined by

$$(p/(E-(1-p)F)) = \sum_{m=0}^{\infty} p(1-p)^m F^m. \tag{1.17}$$

1.2.3 Schemes of series and sequences

Methods for estimating of the accuracy of approximation can depend on the chosen scheme of summation.

In this course, we distinguish between the scheme of sequences and scheme of series. The difference between two schemes can be explained in terms of random variables. By the scheme of sequences we mean a sequence of independent random variables: $\xi_1, \xi_2, \ldots$;

$$S_n = \xi_1 + \xi_2 + \ldots + \xi_n = S_{n-1} + \xi_n.$$

The more general scheme of series means that random variables form a triangular array. That is, for any n, we have (possibly different) set of random variables $\xi_{1n}, \xi_{2n}, \ldots, \xi_{nn}$,

$$S_{n-1} = \xi_{1,n-1} + \ldots + \xi_{n-1,n-1}, \quad S_n = \xi_{1n} + \ldots + \xi_{nn}.$$

In terms of distributions, the scheme of series means that the distribution of k'th summand in the n'th series may depend on n.

1.3 Fourier-Stieltjes transforms

1.3.1 General properties

Let $W \in \mathcal{M}$. Then its Fourier-Stieltjes transform is defined in the following way:

$$\widehat{W}(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx}\, W\{\mathrm{d}x\}.$$

It is easy to check that

$$\widehat{\exp\{W\}}(t) = \exp\{\widehat{W}(t)\}, \quad \widehat{WV}(t) = \widehat{W}(t)\widehat{V}(t), \quad \widehat{E}_a(t) = \mathrm{e}^{\mathrm{i}ta}, \quad \widehat{E}(t) = 1.$$

If $F \in F$, i.e. F is the distribution of random variable ξ, then $\widehat{F}$ is the *characteristic function* of ξ. We recall that, in this case,

$$\widehat{F}(t) = \mathbb{E}\mathrm{e}^{\mathrm{i}\xi t}.$$

If $\widehat{F}(t)$ is the characteristic function then

1. $\widehat{F}(-t)$, $Re\,\widehat{F}(t)$ and $|\,\widehat{F}(t)\,|^2$ also are characteristic functions. Here $Re\widehat{F}(t)$ denotes the real part of $\widehat{F}(t)$.
2. $|\,\widehat{F}(t)\,| \leqslant 1$, $\widehat{F}(0) = 1$.
3. The following estimate holds:

$$|\,1 - \widehat{F}(t)\,|^2 \leqslant 2|\,1 - Re\,\widehat{F}(t)\,|. \tag{1.18}$$

4. Let $\mathbb{E}|\,\xi\,|^s < \infty$. Then

$$\widehat{F}(t) = 1 + (\mathrm{i}t)\,\mathbb{E}\xi + \frac{(\mathrm{i}t)^2}{2}\,\mathbb{E}\xi^2 + \frac{(\mathrm{i}t)^3}{3!}\,\mathbb{E}\xi^3 + \ldots + \frac{(\mathrm{i}t)^{s-1}}{(s-1)!}\,\mathbb{E}\xi^{s-1} + \theta\frac{|\,t\,|^s}{s!}\,\mathbb{E}|\,\xi\,|^s. \tag{1.19}$$

and

$$\widehat{F}'(t) = \mathrm{i}\left\{\mathbb{E}\xi + (\mathrm{i}t)\,\mathbb{E}\xi^2 + \frac{(\mathrm{i}t)^2}{2!}\,\mathbb{E}\xi^3 + \ldots + \frac{(\mathrm{i}t)^{s-2}}{(s-2)!}\,\mathbb{E}\xi^{s-1} + \theta\frac{|\,t\,|^{s-1}}{(s-1)!}\,\mathbb{E}|\,\xi\,|^s\right\}. \tag{1.20}$$

Expansion (1.19) (expansion in moments) plays the main role in the classical approximation theory.

1.3.2 Integer-valued random variables

Let $F \in \mathcal{F}$ be concentrated on $0, 1, 2, \ldots$. Then

$$\widehat{F}(t) = \sum_{k=0}^{\infty} \mathrm{e}^{\mathrm{i}tk} F\{k\}.$$

Sometimes it is more natural to use expansion of $\widehat{F}(t)$ in powers of $(\mathrm{e}^{\mathrm{i}t}-1)$. The kth **factorial moment** of F is defined in the following way:

$$\nu_k = \sum_{j=1}^{\infty} j(j-1)(j-2)\dots(j-k+1)F\{j\}. \tag{1.21}$$

Lemma 1.1 *Let $F \in \mathcal{F}$ be concentrated on non-negative integers and let , for some integer $s \geqslant 1$, $\nu_s < \infty$. Then*

$$\widehat{F}(t) = \sum_{j=0}^{\infty} F\{j\}\mathrm{e}^{\mathrm{i}tj} =$$

$$1 + (\mathrm{e}^{\mathrm{i}t}-1)\,\nu_1 + \frac{(\mathrm{e}^{\mathrm{i}t}-1)^2}{2!}\nu_2 + \ldots + \frac{(\mathrm{e}^{\mathrm{i}t}-1)^{s-1}}{(s-1)!}\nu_{s-1} + \theta\frac{|\,\mathrm{e}^{\mathrm{i}t}-1\,|^s}{s!}\nu_s; \tag{1.22}$$

$$\widehat{F}'(t) = \mathrm{i}\sum_{j=1}^{\infty} F\{j\}j\mathrm{e}^{\mathrm{i}tj} =$$

$$\mathrm{i}\mathrm{e}^{\mathrm{i}t}\left(\nu_1 + (\mathrm{e}^{\mathrm{i}t}-1)\,\nu_2 + \frac{(\mathrm{e}^{\mathrm{i}t}-1)^2}{2!}\nu_3 + \ldots + \frac{(\mathrm{e}^{\mathrm{i}t}-1)^{s-2}}{(s-2)!}\nu_{s-1} + \theta\frac{|\,\mathrm{e}^{\mathrm{i}t}-1\,|^{s-1}}{(s-1)!}\nu_s\right). \tag{1.23}$$

If F is concentrated on non-negative integers, then we can express ν_k through moments and vice versa. However, the magnitudes of ν_k and moments are different.

Example 1.4 *Let ξ be Bernoulli variable, that is*

$$P(\xi = 1) = p = 1 - P(\xi = 0).$$

We can write the distribution of ξ in the following way:

$$\mathcal{L}(\xi) = (1-p)E + pE_1.$$

It is obvious, that the characteristic function of ξ can be expanded in two different ways:

$$(1-p) + p\,\mathrm{e}^{\mathrm{i}t} = 1 + p(\mathrm{e}^{\mathrm{i}t}-1) = 1 + (\mathrm{i}t)p + \frac{(\mathrm{i}t)^2}{2}p + \frac{(\mathrm{i}t)^3}{3!}p + \ldots$$

Thus, expansion in factorial moments is much shorter.

In what follows, the essential role will play estimates of the absolute values of characteristic functions. We shall prove one estimate.

Lemma 1.2 *Let $F \in \mathcal{F}$ be concentrated on non-negative integers, $\lambda(F) = \nu_1 - \nu_1^2 - \nu_2 > 0$. Then*

$$|\widehat{F}(t)| \leqslant \exp\Big\{-2\lambda(F)\sin^2\frac{t}{2}\Big\}. \tag{1.24}$$

Proof. Note that, for $0 \leqslant u \leqslant 1$, $\sqrt{1-u} \leqslant 1 - u/2$ and

$$|\mathrm{e}^{\mathrm{i}t} - 1|^2 = (\cos t - 1)^2 + \sin^2 t = 2 - 2\cos t = 4\sin^2\frac{t}{2}.$$

Consequently,

$$|\widehat{F}(t)| \leqslant |1 + \nu_1(\mathrm{e}^{\mathrm{i}t} - 1)| + \frac{\nu_2}{2}|\mathrm{e}^{\mathrm{i}t} - 1|^2 =$$

$$|(1 + \nu_1(\cos t - 1))^2 + \nu_1^2\sin^2 t|^{1/2} + 2\nu_2\sin^2\frac{t}{2} =$$

$$|(1 + \nu_1^2(\cos^2 t - 2\cos t + \sin^2 t) + 2\nu_1(\cos t - 1)|^{1/2} + 2\nu_2\sin^2\frac{t}{2} =$$

$$|1 - 4(\nu_1 - \nu_1^2)\sin^2\frac{t}{2}|^{1/2} + 2\nu_2\sin^2\frac{t}{2} \leqslant 1 - 2\lambda(F)\sin^2\frac{t}{2} \leqslant \exp\Big\{-2\lambda(F)\sin^2\frac{t}{2}\Big\}.$$

□

1.3.3 Fourier transforms of integrable functions

We say that $f : \mathbb{B} \to \mathbb{R}$ belongs to the space $L_1(\mathbb{R})$ if

$$\|f\|_1 = \int_{-\infty}^{\infty} |f(x)|\,\mathrm{d}x \leqslant \infty.$$

The Fourier transform for $f \in L_1(\mathbb{R})$ is defined by

$$\widehat{f}(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx} f(x)\,\mathrm{d}x.$$

If f is continuous on $\mathbb{R}$ and $f, \widehat{f} \in L_1(\mathbb{R})$, then

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}(t)\mathrm{e}^{-\mathrm{i}tx}\,\mathrm{d}t. \tag{1.25}$$

Moreover, for $F \in \mathcal{F}$,

$$\int_{-\infty}^{\infty} f(x)\,F\{\mathrm{d}x\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}(t)\widehat{F}(t)\,\mathrm{d}t. \tag{1.26}$$

The last relation is known as Parseval's identity. Similarly version of Parseval's identity for $F, G \in \mathcal{F}$ states that

$$\int_{-\infty}^{\infty} \widehat{F}(x)\, G\{\mathrm{d}x\} = \int_{-\infty}^{\infty} \widehat{G}(x)\, F\{\mathrm{d}x\} = \mathbb{E}\mathrm{e}^{\mathrm{i}\,\xi\eta}. \tag{1.27}$$

Here ξ and η are independent random variables, having distributions F and G, respectively.

1.4 Concentration function

Let $F \in \mathcal{F}$, $h \geqslant 0$. The Levy concentration function is defined in the following way

$$Q(F,h) = \sup_x F\{[x, x+h]\}, \quad Q(F,0) = \max_x F\{x\}.$$

It is obvious, that $Q(F,h) \leqslant 1$. Moreover, let $F, G \in \mathcal{F}$, $h > 0$, $a > 0$. Then

$$Q(FG,h) \leqslant \min\{Q(F,h), Q(G,h)\}, \quad Q(F,h) \leqslant (h/a+1)Q(F,a); \tag{1.28}$$

the Kolmogorov-Rogozin inequality:

$$Q(F^n,h) \leqslant \frac{C}{\sqrt{n(1-Q(F,h))}}; \tag{1.29}$$

the Le Cam inequality:

$$Q(\exp\{a(F-E)\},h) \leqslant \frac{C}{\sqrt{aF\{x : |\,x\,| > h\}}}; \tag{1.30}$$

and two inequalities (in principle, due to Esseen) establishing the relation between the concentration and characteristic functions:

$$Q(F,h) \leqslant Ch \int_{|t|<1/h} |\,\widehat{F}(t)\,|\,\mathrm{d}t; \tag{1.31}$$

$$h \int_{|t|<1/h} |\,\widehat{F}(t)\,|^2\,\mathrm{d}t \leqslant CQ(\tilde{F},h). \tag{1.32}$$

Here $\tilde{F}$ denotes distribution with the characteristic function $\widehat{F}(t)\widehat{F}(-t) = |\,\widehat{F}(t)\,|^2$.

The following Lemma gives another relation between the concentration and characteristic functions.

Lemma 1.3 *Let $F \in \mathcal{F}_+$ and let g be measurable bounded function. Then, for any $h > 0$, the following inequality holds*

$$\left|\int_{-\infty}^{\infty} g(t)\widehat{F}(t)\,\mathrm{d}t\right| \leqslant 13Q(F,h)\left(\frac{\sup_t |\,g(t)\,|}{h} + \int_0^{\infty} \sup_{s:|s|\geqslant|t|} |\,g(s)\,|\,\mathrm{d}t\right). \tag{1.33}$$

1.5 Metrics and their properties

1.5.1 Total variation norm

The **total variation norm** of $W \in \mathcal{M}$ is denoted by $\| W \|$. More precisely,

$$\| W \| = W^+\{\mathbb{R}\} + W^-\{\mathbb{R}\}.$$

Here $W = W^+ - W^-$ denotes the Jordan – Hahn decomposition of W. Note that, in probability theory, the total variation distance between $F, G \in \mathcal{F}$ is defined as a supremum over all Borel subsets, i.e.

$$d_{TV}(F,G) = \sup_{A \in \mathcal{B}} | F\{A\} - G\{A\} |.$$

The total variation norm is equivalent to $d_{TV}(\cdot,\cdot)$ in a sense that, for $F, G \in \mathcal{F}$, we have

$$\| F - G \| = 2 d_{TV}(F,G). \tag{1.34}$$

Moreover, for any $W, V \in \mathcal{M}$, $F \in \mathcal{F}$

$$\| W \|/2 \leqslant \sup_{A \in \mathcal{B}} | W\{A\} | \leqslant \| W \|,$$

$$\| F \| = 1, \quad \| WV \| \leqslant \| W \| \, \| V \|, \quad \| E_a W \| = \| W \|.$$

Let W be concentrated on $\mathbb{Z}$. Then

$$\| W \| = \sum_{k=-\infty}^{\infty} | W\{k\} |. \tag{1.35}$$

We shall use the total variation norm because it is more convenient to write $\| F(G - E)^2 \|$ instead of $d_{TV}(FG^2 + F, 2FG)$.

1.5.2 Uniform distance

The analogue of **uniform** Kolmogorov distance is defined in the following way

$$| W | = \sup_x | W\{(-\infty, x)\} |.$$

Let $V, W \in \mathcal{M}$, $F \in \mathcal{F}$, $a \in \mathbb{R}$, then

$$| W | \leqslant \| W \|, \quad | VW | \leqslant | V | \, \| W \|, \quad | E_a V | = | V |, \quad | F | = 1. \tag{1.36}$$

1.5.3 Local distance

Let $W \in \mathcal{M}_Z$, i.e. let W be concentrated on $\mathbb{Z}$. The analogue of the **local** distance is defined in the following way

$$| W |_\infty = \sup_j | W\{j\} |.$$

If $W, V \in \mathcal{M}_Z$ then

$$| WV |_\infty \leqslant \| W \| \, | V |_\infty, \quad | W |_\infty \leqslant \| W \|. \tag{1.37}$$

1.5.4 Fortet - Mourier distance

Let $W \in \mathcal{M}_Z$. The analogue of Fortet - Mourier distance (also known as Kantorovich, Dudley or Wasserstein distance) is defined by

$$| W |_{FM} = \sum_{k=-\infty}^{\infty} | W\{(-\infty, k)\} |. \tag{1.38}$$

Note that Fortet-Mourier distance can be defined for $W \in \mathcal{M}$. However then it is more usual to call it the Wasserstein or Kantorovich distance.

Exercises

1. To prove that, for $k = 1, 2, \ldots,$

$$| \sin \frac{kt}{2} | \leqslant k | \sin \frac{t}{2} |.$$

2. Let $F_i, G_i \in \mathcal{F}$ and let

$$| F_1 - G_1 | \leqslant a_1, \quad | F_2 - G_2 | \leqslant a_2.$$

To prove that

$$| F_1F_2 - G_1G_2 | \leqslant a_1 + a_2.$$

3. To prove (1.37).

4. To prove (1.22).

5. To prove (1.18).

6. Let $W \in \mathcal{M}$. To prove that

$$\| \mathrm{e}^W - E \| \leqslant \exp\{\| W \|\} - 1 \leqslant \| W \| \mathrm{e}^{\| W \|}. \tag{1.39}$$

7. Let $F \in \mathcal{F}$, $\lambda > 0$. To prove that

$$\| \exp\{\lambda(F - E)\} - F \| \leqslant \| (F - E) \|^2/2. \tag{1.40}$$

8. Let $W \in \mathcal{M}$ and $W\{\mathbb{R}\} = 0$. To prove that then

$$\| W \| = 2 \sup_{A \in \mathcal{B}} | W\{A\} |.$$

Bibliographical notes

Most of the material of this Section is quite standard and can be found in any textbook on probability. Factorial moments and their generalizations are comprehensively studied in Franken (1964), Kruopis (1986b),Šiaulys and Čekanavičius (1988). Properties of the concentration function can be found in Arak and Zaĭtsev (1988)or Petrov (1995). Metrics and distances are discussed in Pfeifer et al. (1988), or Barbour et al. (1992). Bergström identity is proved in Bergström (1951) and generalized in Čekanavičius (1998).

2 Direct application of the properties of total variation norm

In this section, we show how simple application of the properties of total variation norm can be used when estimating the accuracy of compound approximation. Frequently, such direct approach is too rough. However, many classical results were obtained in this way. Note that, in principle, all results and methods of this section can be applied to the multidimensional measures or commutative operators.

For convenience, we repeat the definition of the compound measure:

$$\varphi(F) = \sum_{m=0}^{\infty} p_m F^m, \quad \text{where} \quad F \in \mathcal{F}, \quad \sum_{m=0}^{\infty} |\, p_m \,| < \infty. \tag{2.1}$$

2.1 Reduction to lattice case

One of the best known and most frequently used the properties of total variation is the following one.

$$\|\, \varphi(F) \,\| \leqslant \|\, \varphi(E_1) \,\|. \tag{2.2}$$

Indeed, taking into account the fact that m'th convolution of F is a distribution and its norm is equal to 1, we get

$$\|\, \varphi(F) \,\| \leqslant \sum_{m=0}^{\infty} |\, p_m \,|\, \|\, F^m \,\| = \sum_{m=0}^{\infty} |\, p_m \,| = \left\| \sum_{m=0}^{\infty} p_m E_m \right\| = \left\| \sum_{m=0}^{\infty} p_m E_1^m \right\| = \|\, \varphi(E_1) \,\|.$$

It is interesting to note that from (2.2) we get

$$\sup_{F \in \mathcal{F}} \|\, \varphi(F) \,\| = \|\, \varphi(E_1) \,\|. \tag{2.3}$$

In many cases estimate (2.2) allows reducing of the general compound approximation to the case of integer-valued distributions.

Example 2.1 *For $B \in \mathcal{F}$, $0 < p < 1$,*

$$\|\, ((1-p)E + pB)^n - \exp\{np(B-E)\} \,\| \leqslant \|\, ((1-p)E + pE_1)^n - \exp\{np(E_1 - E)\} \,\| =$$

$$\sum_{k=0}^{\infty} \left| \binom{n}{k} p^k (1-p)^{n-k} - \frac{(np)^k}{k!} \mathrm{e}^{-np} \right|.$$

Thus, we reduced the general estimate to the case of the difference between binomial and Poisson distributions.

2.2 Expansion in factorial moments

Characteristic function of an integer-valued random variable can be expanded in factorial moments, see Lemma 1.1. We prove that analogous expansion holds for compound measures (as well as for their Fourier-Stieltjes transforms), which may be preferable if one does not apply the characteristic function method. To make our notation shorter we shall write

$$\alpha_k(\varphi) = \sum_{j=0}^{\infty} j(j-1)\cdots(j-k+1)p_j, \tag{2.4}$$

$$\beta_k(\varphi) = \sum_{j=0}^{\infty} j(j-1)\cdots(j-k+1)|\,p_j\,|. \tag{2.5}$$

It is evident that $\alpha_k(\varphi)$ and $\beta_k(\varphi)$ are very similar to ν_k defined in section *Characteristic functions*. Indeed, if all $p_j \in (0,1)$, $p_0+p_1+\ldots=1$ (i.e. if $\varphi(F)$ is compound *distribution*) then $\alpha_k(\varphi) = \beta_k(\varphi)$ equals to the kth factorial moment for $\varphi(E_1)$. In this case, we can give probabilistic interpretation. Let us recall the fact that compound distribution can be viewed as a distribution of random sum of random variables. Then $\alpha_k(\varphi)$ is kth factorial moment of the number of summands.

Lemma 2.1 *Let $\varphi(F)$ be defined by (2.1), $p_1+p_2+\ldots=1$, and let $\beta_{s+1}(\varphi)<\infty$ for some $s \geqslant 1$. Then*

$$\varphi(F) \;=\; E+\sum_{m=1}^{s}\alpha_m(\varphi)\frac{(F-E)^m}{m!}+\Theta(s)(F-E)^{s+1}. \tag{2.6}$$

Here $\Theta(s)\in\mathcal{M}$, $\|\,\Theta(s)\,\| \leqslant \beta_{s+1}(\varphi)/(s+1)!$.

Proof. Applying (1.1)–(1.2) we get

$$\varphi(F) = \sum_{m=0}^{\infty} p_m\left(F^m - \sum_{j=0}^{s}\binom{m}{j}(F-E)^j + \sum_{j=0}^{s}\binom{m}{j}(F-E)^j\right) =$$

$$\sum_{m=0}^{\infty} p_m \sum_{j=0}^{s}\binom{m}{j}(F-E)^j + (F-E)^{s+1}\sum_{m=s+1}^{\infty} p_m \sum_{j=s+1}^{m}\binom{j-1}{s}F^{m-j} \;=\; I_1+I_2.$$

But

$$I_1 = E+\sum_{m=0}^{\infty} p_m \sum_{j=1}^{s}\binom{m}{j}(F-E)^j = E+\sum_{j=1}^{s}(F-E)^j\alpha_j(\varphi)/j!,$$

$$\|\,I_2\,\| \leqslant \left\| \sum_{m=s+1}^{\infty} p_m \sum_{j=s+1}^{m}\binom{j-1}{s}F^{m-j}\right\| \leqslant \sum_{m=s+1}^{\infty}|\,p_m\,|\sum_{j=s+1}^{m}\binom{j-1}{s} =$$

$$\sum_{m=s+1}^{\infty}|\,p_m\,|\binom{m}{s+1} = \beta_{s+1}(\varphi)/(s+1)!.$$

□.

Corollary 2.1 *Let $\lambda > 0$, $0 < p \leqslant 1$, $F \in \mathcal{F}$, $s \in \{0, 1, ...\}$. Then*

$$\exp\{\lambda(F - E)\} = E + \sum_{m=1}^{s} \frac{\lambda^m (F-E)^m}{m!} + \Theta_1(s)(F - E)^{s+1} \frac{\lambda^{s+1}}{(s+1)!} \tag{2.7}$$

Here $\| \Theta_1(s) \| \leqslant 1$.

2.3 Estimation of the accuracy of approximation

It evident, that the most natural approach in applying compound approximations is to match as much of factorial moments as possible. We begin from the classical result which is usually associated with the names of Khintchine or Le Cam.

2.3.1 The Le Cam inequality

Let $0 \leqslant p_k \leqslant 1$, $B_k \in \mathcal{F}$ $(k = 1, 2, \dots, n)$. Then

$$\Big\| \prod_{k=1}^{n} \Big((1-p_k)E + p_k B_k \Big) - \exp\Big\{ \sum_{k=1}^{n} p_k (B_k - E) \Big\} \Big\| \leqslant 2 \sum_{k=1}^{n} p_k^2. \tag{2.8}$$

The proof of (2.8) is based on the application of Corollary 2.1 or (1.40). To make things more transparent let us introduce auxiliary notation:

$$H_k = (1 - p_k)E - p_k B_k, \quad D_k = \exp\{p_k(B_k - E)\}.$$

Then we obtain

$$\| D_k - H_k \| = \| \exp\{p_k(B_k - E)\} - E - p_k(B_k - E) \| \leqslant$$

$$\| p_k(B_k - E) \|^2/2 \leqslant p_k^2 (\| B_k \| + \| E \|)^2/2 = 2p_k^2.$$

Now, from (1.3) it follows that

$$\Big\| \prod_{k=1}^{n} H_k - \prod_{k=1}^{n} D_k \Big\| \leqslant$$

$$\sum_{k=1}^{n} \| H_k - D_k \| \Big\| \prod_{j=1}^{k-1} H_k \prod_{j=k+1}^{n} D_k \Big\| = \sum_{k=1}^{n} \| H_k - D_k \| \leqslant 2 \sum_{k=1}^{n} p_k^2.$$

Note that there are many various other methods for proving of (2.8).

2.3.2 The Hipp inequality

Let us write a formal equality

$$1 + a = \exp\{\ln(1+a)\} = \exp\Big\{a - \frac{a^2}{2} + \frac{a^3}{3} - \dots\Big\}$$

We shall restrict ourselves to a partial case of Hipp's inequality only. Let $0 < p < 1/2$, $B \in \mathcal{F}$. Then

$$(1-p)E + pB = \exp\Big\{\sum_{k=0}^{\infty}(-1)^k \frac{p^k}{k}(B-E)^k\Big\}. \tag{2.9}$$

Simplified version of Hipp's inequality can be formulated in the following way:

$$\Big\| ((1-p)E + pB)^n - \exp\Big\{n\sum_{k=1}^{s}(-1)^k \frac{p^k}{k}(B-E)^k\Big\}\Big\| \leqslant$$

$$n\frac{(2p)^{s+1}}{1-2p}\exp\Big\{n\frac{(2p)^{s+1}}{1-2p}\Big\}. \tag{2.10}$$

The proof of (2.10) is based on the properties of total variation norm. For the sake of convenience, let us denote

$$H = ((1-p)E + pB)^n, \quad W = n\sum_{k=s+1}^{\infty}(-1)^k \frac{p^k}{k}(B-E)^k,$$

$$D = \exp\Big\{n\sum_{k=1}^{s}(-1)^k \frac{p^k}{k}(B-E)^k\Big\}.$$

Then

$$H = D\exp\{W\}, \quad D = H\exp\{-W\}.$$

Note that $H \in \mathcal{F}$ and $\| H \| = 1$ and we can apply (1.39). Consequently,

$$\| H - D \| = \| H - H\exp\{-W\} \| = \| E - \exp\{-W\} \| \leqslant \| W \| \exp\{\| W \|\}.$$

To end the proof it suffices to note that

$$\| W \| \leqslant n\sum_{k=s+1}^{\infty} p^k \frac{\| B - E \|^k}{k} \leqslant n\sum_{k=s+1}^{\infty}(2p)^k = n(2p)^{s+1}(1-2p)^{-1}.$$

Exercises

1. Let $a > 0$, $b > 0$, $F \in \mathcal{F}$. To prove that

$$\| \exp\{a(F-E)\} - \exp\{b(F-E)\} \| \leqslant 2|\, a - b\,|.$$

2. To prove the Hipp (1986) theorem. Let $s \in \{1, 2, \dots\}$,

$$0 \leqslant p_i \leqslant 1/2, \quad B_i \in \mathcal{F}, \quad (i = 1, 2, ...n),$$

$$H_i = (1 - p_i)E + p_i B_i, \quad D(s) = \exp\Big\{\sum_{i=1}^{n}\sum_{j=1}^{s}(-1)^{j+1}(1/j)(H_i - E)^j\Big\},$$

$$u_i = \frac{(2p_i)^{s+1}}{(1 - 2p_i)(s+1)}.$$

Then

$$\Big\| \prod_{i=1}^{n} H_i - D(s) \Big\| \leqslant \exp\Big\{\sum_{i=1}^{n} u_i\Big\} - 1.$$

3. Let $n \in \mathbb{N}$,

$$H = \sum_{m=0}^{\infty}\left(\frac{1}{2}\right)^{m+1} F^m.$$

To prove that

$$\| H^n - F^n \| \leqslant n \| F - E \|.$$

4. Let $W \in \mathcal{M}$. To obtain the estimate for

$$\Big\| E + W - \mathrm{e}^W \Big(E - \frac{W^2}{2}\Big) \Big\|.$$

Bibliographical notes

The properties of total variation norm are discussed in detail in Le Cam (1965), Hipp (1986) and Čekanavičius (1998). The estimate (2.2) is widely used; see, for example, Michel (1988), Roos (2003).

3 Local estimates

We begin to study the characteristic function method. In general, the characteristic function method means that we estimate differences between distributions (measures)through differences between their characteristic functions (Fourier-Stieltjes transforms). Though the method is usually associated with Esseen and CLT, it is much simpler to use it for integer-valued distributions.

Throughout the following few sections, when estimating the difference of two characteristic functions $f^n - g^n$ we frequently use the same general scheme:

1. The following rough estimate:
$$|\, f^n - g^n \,| \leqslant n \max\{|\, f\,|^{n-1}, |\, g\,|^{n-1}\} |\, f - g\,|.$$
2. Estimate of the difference:
$$|\, f - g\,| \leqslant CU(t).$$
3. Estimates for the characteristic functions:
$$|\, f\,|^{n-1} \leqslant CV^n(t), \quad |\, g\,|^{n-1} \leqslant CV^n(t).$$
4. Collecting of all estimates and application of trivial inequality $x^a \mathrm{e}^{-x} \leqslant C(a)$, $x, a > 0$.

In this section, we consider local estimates, for $W \in \mathcal{M}_Z$. By definition
$$\widehat{W}(t) = \sum_{k=-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tk} W\{k\}.$$
It is easy to verify that
$$\int\limits_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}t(k-m)}\,\mathrm{d}t = \begin{cases} 0, & \text{if } k \neq m, \\ 2\pi, & \text{if } k = m. \end{cases}$$

Therefore, the following formula of inversion holds, for $k \in \mathbb{Z}$,:
$$W\{k\} = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}tk} \widehat{W}(t)\,\mathrm{d}t. \tag{3.1}$$
Now it is very simple to get the estimate in terms of Fourier-Stieltjes transforms.
$$|\, W\,|_\infty \leqslant \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} |\, \widehat{W}(t)\,|\,\mathrm{d}t. \tag{3.2}$$

We shall examine some applications of (3.2).

3.1 Local Poisson approximation to the Binomial law

Let us estimate the difference between the Binomial and Poisson distributions. From Le Cam's estimate we have that

$$|\,((1-p)E+pE_1)^n-\exp\{np(E_1-E)\}\,|_\infty \leqslant \|\,((1-p)E+pE_1)^n-\exp\{np(E_1-E)\}\,\| \leqslant 2np^2.$$

However, the above estimate can be very rough. Therefore, we shall use (3.2). We have

$$|\,((1-p)E+pE_1)^n-\exp\{np(E_1-E)\}\,|_\infty = \sup_k \left| \binom{n}{k} p^k(1-p)^{n-k} - \frac{(np)^k}{k!}\mathrm{e}^{-np} \right| \leqslant$$

$$\frac{1}{2\pi}\int\limits_{-\pi}^{\pi} |\,(1-p+p\mathrm{e}^{\mathrm{i}t})^n - \exp\{np(\mathrm{e}^{\mathrm{i}t}-1)\}\,|\,\mathrm{d}t. \tag{3.3}$$

It is easy to check directly that

$$|\exp\{p(\mathrm{e}^{\mathrm{i}t}-1)\}\,| = \exp\{p\mathit{Re}\,(\mathrm{e}^{\mathrm{i}t}-1)\} = \exp\Big\{-2p\sin^2\frac{t}{2}\Big\}. \tag{3.4}$$

We need a similar estimate for the Binomial distribution. Noting that, for the Bernoulli random variable $\nu_1 = p$, $\nu_2 = 0$ (see Example 1.4), by (1.24) we get

$$|\,(1-p)+p\mathrm{e}^{\mathrm{i}t}\,| \leqslant \exp\Big\{-2p(1-p)\sin^2\frac{t}{2}\Big\}. \tag{3.5}$$

In this course, we do not seek to minimize the absolute constants. Therefore, we shall use the following rough estimate, replacing $n-1$ by n:

$$\max\{|\,(1-p)+p\mathrm{e}^{\mathrm{i}t}\,|^{n-1}, |\exp\{p(\mathrm{e}^{\mathrm{i}t}-1)\}\,|^{n-1}\} \leqslant \mathrm{e}^2\exp\Big\{-2np(1-p)\sin^2\frac{t}{2}\Big\} \tag{3.6}$$

For the estimate of the difference of characteristic functions we shall use Taylor expansion (1.6). Note that $\mathit{Re}\,(\mathrm{e}^{\mathrm{i}t}-1) = -2\sin^2(t/2) \leqslant 0$. Consequently,

$$|\exp\{p(\mathrm{e}^{\mathrm{i}t}-1)\} - 1 - p(\mathrm{e}^{\mathrm{i}t}-1)\,| = \left| p^2(\mathrm{e}^{\mathrm{i}t}-1)^2\int_0^1 \exp\{\tau p(\mathrm{e}^{\mathrm{i}t}-1)\}(1-\tau)\,\mathrm{d}\tau \right| \leqslant$$

$$p^2|\,\mathrm{e}^{\mathrm{i}t}-1\,|^2\int_0^1(1-\tau)\,\mathrm{d}\tau = 4p^2\sin^2\frac{t}{2}. \tag{3.7}$$

Collecting (3.4),(3.5) and (3.6) from (1.9) we get the estimate for the difference of characteristic functions. Substituting the obtained result it into (3.3) and applying (1.8) we get the following estimate

$$|\,((1-p)E+pE_1)^n-\exp\{np(E_1-E)\}\,|_\infty \leqslant \frac{4\mathrm{e}^2}{2\pi}\int\limits_{-\pi}^{\pi} np^2\sin^2\frac{t}{2}\exp\Big\{-2np(1-p)\sin^2\frac{t}{2}\Big\}\,\mathrm{d}t \leqslant$$

$$C\min\Big\{np^2,\,(1-p)^{-3/2}\sqrt{\frac{p}{n}}\Big\}. \tag{3.8}$$

It is easy to see, that (3.8) provides better accuracy for large n. Indeed, we see that the local Poisson approximation is tending to zero, even for $p = \mathit{constant}$.

3.2 General case

We shall generalize previous example for the case of arbitrary chosen integer-valued random variable. Let $F \in \mathcal{F}_Z$ and $F\{k\} = q_k$, that is

$$\widehat{F}(t) = \sum_{k=-\infty}^{\infty} q_k \mathrm{e}^{\mathrm{i}tk}. \tag{3.9}$$

Note that all $q_k \in [0, 1]$ and their sum equal 1. For approximation we shall use a special case of compound distributions — the so called accompanying distribution $\exp\{F - E\}$. Note that

$$\exp\{F - E\} = \exp\Big\{ \sum_{k=-\infty}^{\infty} q_k (E_k - E) \Big\},$$

i.e., we have convolution of various Poisson distributions concentrated on different sublatices of $\mathbb{Z}$. Applying (1.9) we get

$$| \widehat{F}^n(t) - \exp\{n(\widehat{F}(t) - 1)\} | \leqslant$$

$$n \max\{| \widehat{F}(t) |^{n-1}, | \exp\{\widehat{F}(t) - 1\} |^{n-1}\} | \widehat{F}(t) - \exp\{\widehat{F}(t) - 1\} |. \tag{3.10}$$

If $q_0 = 1$ then $F = \exp\{F - E\} = E$ and the difference of distributions equal to 0. Therefore, further we assume that $q_0 < 1$. It is easy to check that

$$| \exp\{\widehat{F}(t) - 1\} | = \exp\{Re\, \widehat{F}(t) - 1\}.$$

We need similar estimate for $\widehat{F}(t)$.

Lemma 3.1 *Let $\widehat{F}(t)$ be defined by (3.9). Then, for all t,*

$$| \widehat{F}(t) | \leqslant \exp\{q_0 (Re\, \widehat{F}(t) - 1)\}. \tag{3.11}$$

Proof. The essential step is the following expression

$$\widehat{F}(t) = q_0 + (1 - q_0) \widehat{V}(t). \tag{3.12}$$

Here

$$\widehat{V}(t) = \sum_{\substack{k \in \mathbb{Z} \\ k \neq 0}} \mathrm{e}^{\mathrm{i}tk} \frac{q_k}{1 - q_0}.$$

It is obvious, that $\widehat{V}(t)$ is characteristic function. But all characteristic functions satisfy inequality

$$| \widehat{V}(t) | \leqslant 1.$$

Consequently,

$$(Re\, \widehat{V}(t))^2 + (Im\, \widehat{V}(t))^2 \leqslant 1.$$

For the proof of Lemma we use the same idea as in (1.24). We have

$$| \widehat{F}(t) |^2 = | q_0 + (1 - q_0) Re\, \widehat{V}(t) + \mathrm{i}(1 - q_0) Im\, \widehat{V}(t) |^2 =$$

$$q_0^2 + 2q_0(1-q_0)Re\,\widehat{V}(t) + (1-q_0)^2\Big((Re\,\widehat{V}(t))^2 + (Im\,\widehat{V}(t))^2\Big) \leqslant$$
$$q_0^2 + 2q_0(1-q_0)Re\,\widehat{V}(t) + (1-q_0)^2 = 1 + 2q_0(1-q_0)(Re\,\widehat{V}(t)-1) \leqslant$$
$$\exp\{2q_0(1-q_0)(Re\,\widehat{V}(t)-1)\} = \exp\{2q_0(Re\,\widehat{F}(t)-1)\}.$$

For the last step, one should note that from (3.12) it follows that

$$(1-q_0)\widehat{(V(t)-1)} = \widehat{F}(t)-1, \quad (1-q_0)(Re\,\widehat{V}(t)-1) = Re\,\widehat{F}(t)-1.$$

□.

Similarly to (3.7) we have

$$|\,\widehat{F}(t) - \exp\{\widehat{F}(t)-1\}\,| \leqslant C|\,\widehat{F}(t)-1\,|^2.$$

Submitting all estimates into (3.10), we get

$$|\,\widehat{F}^n(t) - \exp\{n(\widehat{F}(t)-1)\}\,| \leqslant Cn|\,\widehat{F}(t)-1\,|^2\exp\{nq_0(Re\,\widehat{F}(t)-1)\} =$$
$$Cn(1-q_0)^2|\,\widehat{V}(t)-1\,|^2\exp\Big\{n\frac{q_0}{2}(1-q_0)(Re\,\widehat{V}(t)-1)\Big\}\exp\Big\{n\frac{q_0}{2}(Re\,\widehat{F}(t)-1)\Big\} \leqslant$$
$$Cn(1-q_0)^2|\,Re\,\widehat{V}(t)-1\,|\exp\Big\{n\frac{q_0}{2}(1-q_0)(Re\,\widehat{V}(t)-1)\Big\}\exp\Big\{n\frac{q_0}{2}(Re\,\widehat{F}(t)-1)\Big\} \leqslant$$
$$C\min\Big\{n(1-q_0)^2, \frac{(1-q_0)}{q_0}\Big\}\exp\Big\{n\frac{q_0}{2}(Re\,\widehat{F}(t)-1)\Big\}. \tag{3.13}$$

In above we used (1.18) and (1.12). Consequently, from (3.2) we get

$$|\,F^n - \exp\{n(F-E)\}\,|_\infty \leqslant$$
$$C\min\Big\{n(1-q_0)^2, \frac{(1-q_0)}{q_0}\Big\}\int\limits_{-\pi}^{\pi}\exp\Big\{n\frac{q_0}{2}(Re\,\widehat{F}(t)-1)\Big\}\,\mathrm{d}t. \tag{3.14}$$

Obviously, we can estimate the integral by constant. However, sometimes a more precise estimate is needed. The following estimate is useful in quite many cases.

Lemma 3.2 *Let F be defined by (3.9). Then*

$$\int\limits_{-\pi}^{\pi}\exp\Big\{\lambda(Re\,\widehat{F}(t)-1)\Big\}\,\mathrm{d}t = \int\limits_{-\pi}^{\pi}\exp\Big\{-2\lambda\sum_{j=-\infty}^{\infty}q_j\sin^2\frac{tk}{2}\Big\}\,\mathrm{d}t \leqslant \frac{C}{\sqrt{\lambda(1-q_0)}}. \tag{3.15}$$

Proof. Proof of (3.15) is based on the properties of concentration function. First, note that

$$\lambda(Re\,\widehat{F}(t)-1) = \lambda(1-q_0)(Re\,\widehat{V}(t)-1)$$

and $Re\,\widehat{V}(t)$ is non-negative characteristic function of symmetric distribution $\tilde{V}$, which is concentrated on $\mathbb{Z}\setminus\{0\}$. Therefore, by (1.32) and (1.30) we get

$$\int_{-\pi}^{\pi} \exp\Big\{\lambda(Re\,\widehat{F}(t)-1)\Big\}\,\mathrm{d}t = \int_{-\pi}^{\pi} \exp\Big\{\lambda(1-q_0)(Re\,\widehat{V}(t)-1)\Big\}\,\mathrm{d}t \leqslant$$

$$CQ(\tilde{V},\pi^{-1}) \leqslant \frac{C}{\sqrt{\lambda(1-q_0)\tilde{V}\{x:|\,x\,|>\pi^{-1}\}}} = \frac{C}{\sqrt{\lambda(1-q_0)}};$$

□.

Applying (3.15) to (3.14) we get the final estimate: if F is defined by (3.9), then

$$|\,F^n - \exp\{n(F-E)\}\,|_\infty \leqslant\ C\min\left\{n(1-q_0)^2, q_0^{-3/2}\sqrt{\frac{1-q_0}{n}}\right\}. \qquad (3.16)$$

If F has the Binomial distribution, then $q_0 = 1-p$ and the general estimate (3.16) has the same order of accuracy as (3.8).

3.3 Local Franken-type estimate

Previous example was a direct generalization of the Poisson approximation to the Binomial law. Now we shall consider another generalization of the same inequality. Let us assume that $F, G \in \mathcal{F}$ are concentrated on non-negative integers and have s finite moments ($s>1$ some fixed integer). The corresponding factorial moments (see (1.21)) we denote by $\nu_k(F)$ and $\nu_k(G)$, respectively. Let

$$\nu(F)_k = \nu(G)_k, \quad (k=1,2,\dots,s-1) \qquad (3.17)$$

$$\lambda = \min\{\nu_1(F)-\nu_1^2(F)-\nu_2(F), \nu_1(G)-\nu_1^2(G)-\nu_2(G)\} > 0. \qquad (3.18)$$

Condition (3.18) is known as Franken's condition. In general, it is quite restrictive one. For example, it requires for both distributions F and G to have means less than 1.

We shall prove that if (3.17) and (3.18) are satisfied, then

$$|\,F^n - G^n\,|_\infty \leqslant C(s)(\nu_s(F)+\nu_s(G))\lambda^{-(s+1)/2})n^{-(s-1)/2}. \qquad (3.19)$$

.

Indeed, from expansion in factorial moments (1.22) we have that

$$|\,\widehat{F}(t)-\widehat{G}(t)\,| \leqslant (\nu_s(F)+\nu_s(G))|\,\mathrm{e}^{\mathrm{i}t}-1\,|^s/s! = (\nu_s(F)+\nu_s(G))4^s|\,\sin\frac{t}{2}\,|^s.$$

Moreover, by (1.24),

$$\max\{|\,\widehat{F}(t)\,|, |\,\widehat{G}(t)\,|\} \leqslant \exp\Big\{-2\lambda\sin^2\frac{t}{2}\Big\}.$$

Consequently, collecting both estimates and substituting them into (1.9) and (3.2) we get

$$| F^n - G^n |_\infty \leqslant C(s)n(\nu_s(F) + \nu_s(G)) \int\limits_{-\pi}^{\pi} \exp\Big\{-2\lambda \sin^2 \frac{t}{2}\Big\} \,\mathrm{d}t.$$

The desired estimate follows from (1.8).

We shall check that (3.19) indeed is some generalization of the Poisson approximation to the Binomial distribution with $s = 2$. Let F be the Binomial distribution with parameters p and n. Then

$$F = (1-p)E + pE_1 = E + p(E_1 - E) \quad \nu_1(F) = p, \quad \nu_2(F) = 0.$$

Similarly, if G is corresponding Poisson law, then

$$G = \exp\{p(E_1 - E)\}, \quad \nu_1(G) = p, \quad \nu_2(G) = p^2.$$

It is easy to check that, $\nu_2(F) + \nu_2(G) = p^2$. Now comes the important step. Formally, we should take $\lambda = \nu_1(G) - \nu_1^2(G) - \nu_2(G) = p - 2p^2$. Substituting these estimates into (3.19) we shall get a slightly different estimate than in (3.8). On the other hand, we already know that characteristic function of the Poisson law can be estimated through $\lambda = p(1-p)$, i.e. better than in (1.24).

Exercises

1. Let F and G have $s - 1$ coinciding factorial moment and finite factorial moments of the sth order. To prove that then

$$| F^n - G^n |_\infty \leqslant C(s)n(\nu_s(F) + \nu_s(G)). \tag{3.20}$$

2. Let $W \in \mathcal{M}_Z$. To prove that

$$\sum_{k=-\infty}^{\infty} | W\{k\} |^2 = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} | \widehat{W}(t) |^2 \,\mathrm{d}t. \tag{3.21}$$

Hint. To use the inversion formula for the convolution of F and distribution with characteristic function $\widehat{F}(-t)$.

3. Let F be defined by (3.9) with $q_0 = 1/5$. To prove that then

$$| (F - E)^k \exp\{n(F - E)\} |_\infty \leqslant C(k)n^{-(k+1)/2}.$$

4. Let $n \in \mathbb{N}$, F be defined by (3.9) with $q_0 = const < 1$,

$$H = \sum_{m=0}^{\infty} \left(\frac{1}{2}\right)^{m+1} F^m.$$

To get the estimate for

$$| H^n - F^n |_\infty.$$

5. Let $F \in \mathcal{F}_Z$ and $a > 0$, $b > 0$. To prove that

$$|\exp\{a(F-E)\} - \exp\{b(F-E)\}|_\infty \leqslant C\frac{|a-b|}{b}.$$

Hint. Consider three cases $b \leqslant a$, $a < b \leqslant 2a$ and $2a \leqslant b$. For the last case apply

$$\frac{1}{2} \leqslant \frac{|a-b|}{b}.$$

Bibliographical notes

Formula of inversion (3.1) is well known and can be found in numerous papers. The characteristic function method for lattice local estimates under Franken's condition was applied by Franken (1964), Kruopis (1986b),Šiaulys and Čekanavičius (1988).

4 The Tsaregradskii inequality

One of the most popular methods for the uniform estimates of $W \in \mathcal{M}_Z$ is the Tsaregradskii inequality. It can be written in the following way. Let $W \in \mathcal{M}_Z$, then

$$|\,W\,| \leqslant \frac{1}{4\pi}\int\limits_{-\pi}^{\pi} \frac{|\,\widehat{W}(t)\,|}{|\,\sin\frac{t}{2}\,|}\,\mathrm{d}t. \tag{4.1}$$

Proof of (4.1). The estimate is trivial if the right-hand of (4.1)is infinite. Therefore, we shall assume that it is finite. For two integers $s < m$ by formula of inversion (3.1) we get

$$\sum_{k=s}^{m-1} W\{k\} = \sum_{k=s}^{m-1} \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}tk}\widehat{W}(t)\,\mathrm{d}t =$$

$$\frac{1}{2\pi}\int\limits_{-\pi}^{\pi} \widehat{W}(t)\sum_{k=s}^{m-1}\mathrm{e}^{-itk}\,\mathrm{d}t = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\widehat{W}(t)\frac{\mathrm{e}^{-itm}-\mathrm{e}^{-its}}{1-\mathrm{e}^{-\mathrm{i}t}}\,\mathrm{d}t. \tag{4.2}$$

The Riemann-Lebesgue theorem states that, for absolutely integrable function $g(t)$,

$$\lim_{y\to\pm\infty}\int_{-\infty}^{\infty} \mathrm{e}^{ity}g(t)\,\mathrm{d}t = 0.$$

Therefore, the limit of (4.2) when $s \to -\infty$ gives the following formula of inversion

$$W\{(-\infty, m)\} = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\widehat{W}(t)\frac{\mathrm{e}^{-itm}}{\mathrm{e}^{-\mathrm{i}t}-1}\,\mathrm{d}t. \tag{4.3}$$

Consequently,

$$|\,W\{(-\infty, m)\}\,| \leqslant \frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\frac{|\,\widehat{W}(t)\,|}{|\,\sin\frac{t}{2}\,|}\,\mathrm{d}t.$$

The Tsaregradskii inequality is obtained by taking supremum over all $m \in \mathbb{Z}$.

Note that, in the literature, a slightly more rough version of (4.1) is used. It is obtained from (4.1) replacing $|\,\sin\frac{t}{2}\,|$ by its estimate $|\,t\,|/\pi$:

$$|\,W\,| \leqslant \frac{1}{4}\int\limits_{-\pi}^{\pi}\frac{|\,\widehat{W}(t)\,|}{|\,t\,|}\,\mathrm{d}t. \tag{4.4}$$

Version (4.4) is more convenient when using expansions in moments. Meanwhile (4.1) is more convenient when using expansions in factorial moments.

4.1 Integral Poisson approximation to the Binomial law

Taking into account estimates (3.4)–(3.7) we can write

$$|\,((1-p)+p\mathrm{e}^{\mathrm{i}t})^n - \exp\{np(\mathrm{e}^{\mathrm{i}t}-1)\}\,| \leqslant Cnp^2\sin^2\frac{t}{2}\exp\{-2np(1-p)\sin^2\frac{t}{2}\}. \qquad (4.5)$$

Therefore, by (4.1) and (1.8) we get

$$|\,((1-p)E + pE_1)^n - \exp\{np(E_1-E)\}\,| \leqslant C(1-p)^{-1}p. \qquad (4.6)$$

Note that from Le Cam's estimate (2.8) we already know that the accuracy of Poisson approximation is less than $2np^2$. The same order np^2 can be easily obtained from the Tsaregradkii inequality with, however, larger constant. As can be seen from (4.6) Poisson approximation is accurate only for small p.

4.2 Integral Franken type estimate

Without any difficulty we can get the integral estimate for the difference of two distributions, satisfying Frankens condition. Let us assume that $F, G \in \mathcal{F}$ are concentrated on non-negative integers and have s finite moments ($s > 1$ some fixed integer) and let

$$\nu(F)_k = \nu(G)_k, \quad (k = 1, 2, \ldots, s-1) \qquad (4.7)$$

$$\lambda = \min\{\nu_1(F) - \nu_1^2(F) - \nu_2(F), \nu_1(G) - \nu_1^2(G) - \nu_2(G)\} > 0. \qquad (4.8)$$

Then, as proved in section for local estimates

$$|\,\widehat{F}^n(t) - \widehat{G}^n(t)\,| \leqslant Cn(\nu_s(F) + \nu_s(G))4^s|\,\sin\frac{t}{2}\,|^s.$$

Consequently, applying (1.8) and (4.1) we get

$$|\,F^n - G^n\,| \leqslant C\lambda^{-s/2}n^{-(s-2)/2}. \qquad (4.9)$$

4.3 Asymptotic expansion for the Poisson approximation

For the sake of brevity, we shall use notation $z = \mathrm{e}^{\mathrm{i}t} - 1$. Formally, we can write the following series:

$$(1+pz)^n = \exp\Big\{npz - n\frac{(pz)^2}{2} + n\frac{(pz)^3}{3} - \ldots\Big\} = \mathrm{e}^{npz}\Big(1 - n\frac{(pz)^2}{2} + \ldots\Big).$$

Consequently, it is natural to construct short asymptotic expansion for the Poisson approximation to the Binomial law in the following way:

$$D = \exp\{np(E_1-E)\}\Big(E - \frac{np^2}{2}(E_1-E)^2\Big). \qquad (4.10)$$

Let us assume that $0 \leqslant p \leqslant 1/2$. We seek to apply the Tsaregradkii inequality. Therefore, we need an estimate for the difference of Fourier-Stieltjes transforms of corresponding measures.

The following technical trick might be quite effective, especially when minimizing constants (which we will not do). We introduce an auxiliary τ in the following way:

$$\Big|((1-p)+p\mathrm{e}^{\mathrm{i}t})^n-\widehat{D}(t)\Big| = \Big|(1+pz)^n-\mathrm{e}^{npz}\Big(1-n\frac{(pz)^2}{2}\Big)\Big| =$$

$$|\,\mathrm{e}^{npz}\,|\Big|\Big((1+p\tau z)^n\mathrm{e}^{-np\tau z}-\Big(1-n\frac{(p\tau z)^2}{2}\Big)\Big)\Big|_0^1\Big| =$$

$$|\,\mathrm{e}^{npz}\,|\Big|\int_0^1\frac{\partial}{\partial\tau}\Big((1+p\tau z)^n\mathrm{e}^{-np\tau z}-\Big(1-n\frac{(p\tau z)^2}{2}\Big)\Big)\,\mathrm{d}\tau\Big| =$$

$$|\,\mathrm{e}^{npz}\,|\Big|\int_0^1\Big(n(1+p\tau z)^{n-1}pz\mathrm{e}^{-np\tau z}-npz(1+pz\tau)n\mathrm{e}^{-np\tau z}+np^2z^2\tau\Big)\Big)\,\mathrm{d}\tau\Big| \leqslant$$

$$|\,\mathrm{e}^{npz}\,|\int_0^1 np^2|\,z\,|^2\tau\Big|(1+p\tau z)^{n-1}pz\mathrm{e}^{-np\tau z}-1\Big|\,\mathrm{d}\tau. \tag{4.11}$$

Similarly to (3.5) we obtain

$$|\,1+pz\tau\,|\leqslant\exp\Big\{-2p\tau(1-p\tau)\sin^2\frac{t}{2}\Big\}$$

and

$$|\,\mathrm{e}^{-pz\tau}\,| = \exp\{-p\tau(\cos t-1)\} = \exp\Big\{2p\tau\sin^2\frac{t}{2}\Big\}.$$

Consequently,

$$|\,1+pz\tau\,|\,|\,\mathrm{e}^{-pz\tau}\,|\leqslant\exp\Big\{2p^2z\tau\sin^2\frac{t}{2}\Big\}.$$

Note that $0\leqslant\tau\leqslant 1$. Therefore,

$$\Big|(1+p\tau z)^{n-1}pz\mathrm{e}^{-np\tau z}-1\Big|\leqslant$$

$$\Big|(1+p\tau z)^{n-1}pz\mathrm{e}^{-np\tau z}-(1+pz\tau)^n\mathrm{e}^{-np\tau z}\Big|+\Big|(1+p\tau z)^npz\mathrm{e}^{-np\tau z}-1\Big|\leqslant$$

$$|\,1+pz\tau\,|^{n-1}|\,\mathrm{e}^{-npz\tau}\,||\,z\,|p\tau+n\max\{1,|\,1+pz\tau\,||\,\mathrm{e}^{-npz\tau}\,|\}|\,(1+pz\tau)\mathrm{e}^{-pz\tau}-1\,|\leqslant$$

$$C(|\,z\,|p\tau+np^2\tau^2|\,z\,|^2)\exp\Big\{2np^2\tau^2\sin^2\frac{t}{2}\Big\}\leqslant C(|\,z\,|p\tau+np^2\tau^2|\,z\,|^2)\exp\Big\{2np^2\sin^2\frac{t}{2}\Big\}.$$

Combining the last estimate with (4.11) we get

$$\Big|((1-p)+p\mathrm{e}^{\mathrm{i}t})^n-\widehat{D}(t)\Big|\leqslant C(|\,z\,|^3np^3+n^2p^4|\,z\,|^4)\exp\Big\{2np^2\sin^2\frac{t}{2}\Big\}\exp\Big\{-2np\sin^2\frac{t}{2}\Big\} = .$$

$$C(|\,z\,|^3np^3+n^2p^4|\,z\,|^4)\exp\Big\{-2np(1-p)\sin^2\frac{t}{2}\Big\}. \tag{4.12}$$

From (4.12) and Tsaregradskii inequality we finally get, for $0\leqslant p\leqslant 1/2$,

$$\Big|((1-p)E+pE_1)^n-\exp\{np(E_1-E)\}\Big(E-\frac{np^2}{2}(E_1-E)^2\Big)\Big|\leqslant C\min\{np^3,p^2\}. \tag{4.13}$$

Remark 4.1 *There other methods for estimation of the difference of Fourier-Stieltjes transforms. For example, one can use (1.5) obtaining:*

$$\Big|((1-p)+p\mathrm{e}^{\mathrm{i}t})^n-\widehat{D}(t)\Big| = \mathrm{e}^{npz}\int_0^1\frac{(1-\tau)^2}{2}\frac{\partial^3}{\partial\tau^3}\Big(\mathrm{e}^{-npz\tau}(1+pz\tau)^n\Big)\,\mathrm{d}\tau.$$

4.4 Taking into account symmetry

Let $F = p_0E + p_1E_1 + p_2E_{-1}$, $G = \exp\{F - E\}$. For approximation of $|\,F^n - G^n\,|$ we use the Tsaregradskii inequality. Note that by Lemma 3.1

$$\max\{|\,\widehat{F}(t)\,|, |\,\widehat{G}(t)\,|\} \leqslant \exp\{p_0((Re\,\widehat{F}(t) - 1))\} = \exp\Big\{-2p_0(1-p_0)\sin^2\frac{t}{2}\Big\}.$$

Moreover,

$$|\,\widehat{F}(t) - \widehat{G}(t)\,| \leqslant C|\,\widehat{F}(t) - 1\,|^2 = C|\,p_1(\mathrm{e}^{\mathrm{i}t} - 1) + p_2(\mathrm{e}^{-\mathrm{i}t} - 1)\,|^2.$$

Of course, we can use the estimate

$$|\,p_1(\mathrm{e}^{\mathrm{i}t} - 1) + p_2(\mathrm{e}^{-\mathrm{i}t} - 1)\,| \leqslant p_1|\,\mathrm{e}^{\mathrm{i}t} - 1\,| + p_2|\,\mathrm{e}^{-\mathrm{i}t} - 1\,| = 2(1-p_0)\sin^2\frac{t}{2}.$$

However, such an estimate might be too rough, because we do not take into account possible symmetry of distribution F. It is possible to take into account the effects of symmetry expanding $\mathrm{e}^{\mathrm{i}t}$ and $\mathrm{e}^{-\mathrm{i}t}$ in the powers of $(\mathrm{i}t)$. We shall choose another route. From the relation between $(\mathrm{e}^{\mathrm{i}t} - 1)$ and $(\mathrm{e}^{-\mathrm{i}t} - 1)$ given at the beginning of the course we obtain

$$|\,p_1(\mathrm{e}^{\mathrm{i}t} - 1) + p_2(\mathrm{e}^{-\mathrm{i}t} - 1)\,| = |\,p_1(\mathrm{e}^{\mathrm{i}t} - 1) - p_2(\mathrm{e}^{\mathrm{i}t} - 1) - p_2(\mathrm{e}^{\mathrm{i}t} - 1)(\mathrm{e}^{-\mathrm{i}t} - 1)\,| \leqslant$$
$$|\,p_1 - p_2\,|\,\Big|\sin\frac{t}{2}\Big| + p_2\sin^2\frac{t}{2}. \tag{4.14}$$

We can see that, in (4.14), the smallness of the difference between p_1 and p_2 is taken into account. Applying Tsaregradkii inequality, we now get the estimate

$$|\,(p_0E + p_1E_1 + p_2E_{-1})^n - \exp\{np_1(E_1 - E) + np_2(E_{-1} - E)\}\,| \leqslant$$
$$Cp_0^{-1}(1-p_0)^{-1}|\,p_1 - p_2\,| + p_0^{-2}(1-p_0)^{-2}p_2^2n^{-1}. \tag{4.15}$$

If $p_1 = p_2$ and $p_0 = const$ then the accuracy is of the order n^{-1}.

Exercises

1. Let F be the Binomial distribution with parameters n and p. To prove that
$$|\,F^n - F^{n+1}\,| \leqslant C\sqrt{\frac{p}{n(1-p)}}.$$
2. Using expansion in moment to get the estimate for $|\,F^n - G^n\,|$, when $F \in \mathcal{F}_Z$ and $G \in \mathcal{F}_Z$, both satisfy (4.7) and (4.8) and, for some $s > 1$ have finite absolute moments of order s and coinciding moments of lesser order.
3. To get an estimate for $|\,F^n - \exp\{n(F - E)\}\,|$, when $F \in \mathcal{F}_Z$.
4. Let $0 \leqslant p \leqslant 1$. To prove that
$$\Big|\,((1-p)E + pE_1)^n - \exp\Big\{np(E_1 - E) - \frac{np^2}{2}(E_1 - E)^2\Big\}\,\Big| \leqslant Cp\sqrt{\frac{p}{n}}.$$

Bibliographical notes

Tsaregradkii inequality is used in so numerous papers that it become a standard technique. We note only Tsaregradkii (1958), Franken (1964) and Kruopis (1986a,b).

5 Estimating total variation

We recall that, for $W \in \mathcal{M}$, $|W| \leqslant \|W\|$. In this section, we continue investigation of measures concentrated on non-negative integers. One of our aims is showing that many estimates, obtained for the uniform distance, have the same order of accuracy when the total variation norm is used.

5.1 The characteristic function method

As in previous section we need a formula of inversion. Once again, it can be obtained quite easily.

Lemma 5.1 *Let $W \in \mathcal{M}_Z$, $\sum_{k\in\mathbb{Z}} |k||W\{k\}| < \infty$. Then, for any $a \in \mathbb{R}$, $b > 0$ the following inequality holds*

$$\|W\| \leqslant (1+b\pi)^{1/2}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\left|\widehat{W}(t)\right|^2 + \frac{1}{b^2}\left|\left(\mathrm{e}^{-\mathrm{i}ta}\widehat{W}(t)\right)'\right|^2\right)\mathrm{d}t\right)^{1/2}. \tag{5.1}$$

Proof. We begin from the following identity

$$\sum_{k=-\infty}^{\infty}(k-a)^2|W\{k\}|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|(\mathrm{e}^{-\mathrm{i}a}\widehat{W}(t))'|^2\,\mathrm{d}t. \tag{5.2}$$

Indeed, due to Lemma's assumptions we can take the derivative of $\widehat{W}(t)$:

$$\left(\widehat{W}(t)\mathrm{e}^{-\mathrm{i}ta}\right)' = \sum_{k\in\mathbb{Z}} W\{k\}\left(\mathrm{e}^{\mathrm{i}t(k-a)}\right)' = \mathrm{i}\sum_{k\in\mathbb{Z}} W\{k\}(k-a)\mathrm{e}^{\mathrm{i}t(k-a)}.$$

Consequently, defining $V\{k\} = (k-a)W\{k\}$, we have

$$\widehat{V}(t) = -i\mathrm{e}^{\mathrm{i}ta}\left(\widehat{W}(t)\mathrm{e}^{-\mathrm{i}ta}\right)'$$

and can apply (3.21).

Applying (1.10) and (3.21) – (5.2) we obtain

$$\|W\|^2 = \left(\sum_{k\in\mathbb{Z}}|W\{k\}|\left(1+\left(\frac{k-a}{b}\right)^2\right)^{1/2}\left(1+\left(\frac{k-a}{b}\right)^2\right)^{-1/2}\right)^2 \leqslant$$

$$\sum_{k\in\mathbb{Z}}|W\{k\}|^2\left(1+\left(\frac{k-a}{b}\right)^2\right)\sum_{k\in\mathbb{Z}}\left(1+\left(\frac{k-a}{b}\right)^2\right)^{-1} \leqslant$$

$$(1+b\pi)\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\left|\widehat{W}(t)\right|^2 + \frac{1}{b^2}\left|\left(\mathrm{e}^{-\mathrm{i}ta}\widehat{W}(t)\right)'\right|^2\right)\mathrm{d}t.$$

What role in (5.1) play a and b? We shall give some heuristic answers. The characteristic function method for lattice measures means that, estimating the Fourier-Stieltjes transform of the measure, we get rough impression about the uniform estimate. Moreover, in quite a lot of cases, we encounter the following principal scheme: if $|\widehat{W}(t)| \approx A\lambda^{-n}$, then $|W| \approx A\lambda^{-n}$ and

$$\int_{-\pi}^{\pi} |\widehat{W}(t)|\,\mathrm{d}t \approx A\lambda^{-n-1/2} \text{ and } \int_{-\pi}^{\pi} |\widehat{W}(t)|/|t|\,\mathrm{d}t \approx A\lambda^{-n}.$$

The estimate in total variation can not be better than the uniform estimate. Now let us study (5.1). We see, that estimate contains two parts: integral of the Fourier-Stieltjes transform and integral of its derivative. In both integrals b acts differently. Taking $b = \lambda^{1/2}$ we will preserve the order λ^{-n} for the first integral and improve the order of the second integral.

The role of a can be explained heuristically in the following way. Any additional factor $|\sin(t/2)|$ improves the accuracy of approximation by $\lambda^{-1/2}$. As can be seen from the following example, suitable centering, can radically improve the accuracy. Indeed, multiplying $|\widehat{W}(t)|$ by

$$\left|\left(\mathrm{e}^{p(\mathrm{e}^{\mathrm{i}t}-1)}\right)'\right| = \left|\mathrm{e}^{p(\mathrm{e}^{\mathrm{i}t}-1)}\mathrm{i}p\mathrm{e}^{\mathrm{i}t}\right| = p\mathrm{e}^{-2p\sin^2\frac{t}{2}}$$

we get no additional improvement. On the other hand, multiplying $|\widehat{W}(t)|$ by

$$\left|\left(\mathrm{e}^{-\mathrm{i}tp}\mathrm{e}^{p(\mathrm{e}^{\mathrm{i}t}-1)}\right)'\right| = \left|\mathrm{e}^{p(\mathrm{e}^{\mathrm{i}t}-1-\mathrm{i}t)}\mathrm{i}p(\mathrm{e}^{\mathrm{i}t}-1)\right| = 2p\left|\sin^2\frac{t}{2}\right|\mathrm{e}^{-2p\sin^2\frac{t}{2}} \qquad (5.3)$$

we get improvement of the order $\lambda^{-1/2}$. As a rule, a equals to the mean of approximated distribution.

We demonstrate how to obtain the total variation estimates on a couple of examples.

Poisson approximation to the Binomial law

Let $np \geqslant 1$, $p \leqslant 1/4$. and let

$$W = ((1-p)E + pE_1)^n - \exp\{np(E_1 - E)\}.$$

We already established (see (4.5)) that

$$|\widehat{W}(t)| = |((1-p)+p\mathrm{e}^{\mathrm{i}t})^n - \exp\{np(\mathrm{e}^{\mathrm{i}t}-1)\}| \leqslant Cnp^2\sin^2\frac{t}{2}\exp\{-2np(1-p)\sin^2\frac{t}{2}\}. \quad (5.4)$$

Let $a = np$, i.e. a equals to the mean of Binomial distribution. Now we can estimate the derivative part.

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{e}^{-\mathrm{i}ta}\widehat{W}(t)\right) = -\mathrm{i}a\mathrm{e}^{-\mathrm{i}ta}\widehat{W}(t) + \mathrm{e}^{-\mathrm{i}ta}\widehat{W}'(t) =$$

$$\mathrm{e}^{-\mathrm{i}ta}\Big(- \mathrm{i}np\widehat{W}(t) + \mathrm{i}np\mathrm{e}^{\mathrm{i}t}(1-p+p\mathrm{e}^{\mathrm{i}t})^{n-1} - \mathrm{i}np\mathrm{e}^{\mathrm{i}t}\exp\{np(\mathrm{e}^{\mathrm{i}t}-1)\}\Big) =$$

$$-\mathrm{e}^{-\mathrm{i}ta}\mathrm{i}np\Big(((1-p)+p\mathrm{e}^{\mathrm{i}t})^{n-1}\mathrm{e}^{\mathrm{i}t} - (1-p+p\mathrm{e}^{\mathrm{i}t})^n + \exp\{np(\mathrm{e}^{\mathrm{i}t}-1)\} - \mathrm{e}^{\mathrm{i}t}\exp\{np(\mathrm{e}^{\mathrm{i}t}-1)\}\Big) =$$

$$-\mathrm{e}^{-ita} inp\Big(((1-p)+p\mathrm{e}^{\mathrm{i}t})^{n-1}(\mathrm{e}^{\mathrm{i}t}-1+p-p\mathrm{e}^{\mathrm{i}t})+(1-\mathrm{e}^{\mathrm{i}t})\exp\{np(\mathrm{e}^{\mathrm{i}t}-1)\}\Big)=$$

$$-\mathrm{e}^{-ita} inp(\mathrm{e}^{\mathrm{i}t}-1)\Big(((1-p)+p\mathrm{e}^{\mathrm{i}t})^{n-1}(1-p)-\exp\{np(\mathrm{e}^{\mathrm{i}t}-1)\}\Big)=$$

$$-\mathrm{e}^{-ita} inp(\mathrm{e}^{\mathrm{i}t}-1)\Big(((1-p)+p\mathrm{e}^{\mathrm{i}t})^{n-1}(1-p)-(1-p+p\mathrm{e}^{\mathrm{i}t})^{n}+\widehat{W}(t)\Big)=$$

$$-\mathrm{e}^{-ita} inp(\mathrm{e}^{\mathrm{i}t}-1)\Big(((1-p)+p\mathrm{e}^{\mathrm{i}t})^{n-1}(-p\mathrm{e}^{\mathrm{i}t})+\widehat{W}(t)\Big).$$

Consequently,

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\Big(\mathrm{e}^{-ita}\widehat{W}(t)\Big)\right|\leqslant Cnp^2\Big|\sin\frac{t}{2}\Big|\exp\Big\{-2np(1-p)\sin^2\frac{t}{2}\Big\}. \tag{5.5}$$

From the elementary inequality $x\exp\{-x\}\leqslant 1$ we get

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\Big(\mathrm{e}^{-ita}\widehat{W}(t)\Big)\right|\leqslant C\sqrt{n}p^{3/2}\exp\Big\{-2np(1-p)\sin^2\frac{t}{2}\Big\},$$

$$|\,\widehat{W}(t)\,|\leqslant Cp\exp\Big\{-2np(1-p)\sin^2\frac{t}{2}\Big\}.$$

Let $b=\sqrt{np}$. Taking into account that $b\geqslant 1$ we can rewrite (5.1) in the following way

$$\|\,W\,\|^2\leqslant C\int\limits_{-\pi}^{\pi}\left(b\Big|\,\widehat{W}(t)\Big|^2+\frac{1}{b}\Big|\,\Big(\mathrm{e}^{-ita}\widehat{W}(t)\Big)'\,\Big|^2\right)\mathrm{d}t. \tag{5.6}$$

Substituting estimates (5.4)–(5.5) into (5.6) we get

$$\|\,W\,\|^2\leqslant Cp^2\int\limits_{-\pi}^{\pi}\sqrt{np}\exp\Big\{-2np(1-p)\sin^2\frac{t}{2}\Big\}\,\mathrm{d}t\leqslant Cp^2.$$

Consequently,

$$\|\,((1-p)E+pE_1)^n-\exp\{np(E_1-E)\}\,\|\leqslant Cp. \tag{5.7}$$

Note, that (5.7) is trivial, for $p>1/4$. Moreover, the estimate is also true, for $np<1$, because in this case the left-hand side of (5.7) is less than $2np^2$.

The estimate (5.7) can be similarly proved for non-identically distributed Bernoulli random variables. We have not tried to get small absolute constant. For more than 50 years improvement of constant in (5.7) attracts attention of many mathematicians. Note that combined with the general property of the total variation norm (see Example 2.1) and estimate (2.8) the estimate (5.7) can be written in the following general form

$$\sup_{B\in\mathcal{F}}\|\,((1-p)E+pB)^n-\exp\{np(B-E)\}\,\|\leqslant C\min\{np^2,p\}. \tag{5.8}$$

Of course, the estimate (5.8) also holds for the uniform distance.

Franken type approximations

Let F and G be concentrated on non-negative integers, $s>1$, $\kappa=\nu_s(F)+\nu(G)_s<\infty$ and $\nu(F)_k=\nu(G)_k$, for $k=1,2,\dots,s$ and let

$$\lambda=\min\{\nu_1(F)-\nu_1^2(F)-\nu_2(F),\nu_1(G)-\nu_1^2(G)-\nu_2(G)\}>0. \tag{5.9}$$

Then, as proved in Section for local estimates

$$|\,\widehat{F}^n(t)-\widehat{G}^n(t)\,|\leqslant C(s)n\kappa|\,\sin\frac{t}{2}\,|^s. \tag{5.10}$$

Moreover, from (1.22) and (1.23) we have

$$|\,\widehat{F}(t)-\widehat{G}(t)\,|\leqslant C(s)\kappa\Big|\,\sin\frac{t}{2}\Big|^s,\quad |\,\widehat{F}'(t)-\widehat{G}'(t)\,|\leqslant C(s)\kappa\Big|\,\sin\frac{t}{2}\Big|^{s-1}. \tag{5.11}$$

Further on we assume that $n\lambda>1$. Let $a=\nu_1(F)=\nu_1(G)$. In the proof below, for the sake of brevity we shall write $\widehat{F}$ instead of $\widehat{F}(t)$. We have

$$\mathrm{e}^{-itan}(\widehat{F}^n-\widehat{G}^n)=(\mathrm{e}^{-\mathrm{i}ta}\widehat{F})^n-(\mathrm{e}^{-\mathrm{i}ta}\widehat{G})^n$$

and

$$\Big(\mathrm{e}^{-itan}(\widehat{F}-\widehat{G})\Big)'=n(\mathrm{e}^{-\mathrm{i}ta}\widehat{F})^{n-1}(\mathrm{e}^{-\mathrm{i}ta}\widehat{F})'-n(\mathrm{e}^{-\mathrm{i}ta}\widehat{G})^{n-1}(\mathrm{e}^{-\mathrm{i}ta}\widehat{G})'=$$

$$n\{(\mathrm{e}^{-\mathrm{i}ta}\widehat{F})^{n-1}-(\mathrm{e}^{-\mathrm{i}ta}\widehat{G})^{n-1}\}(\mathrm{e}^{-\mathrm{i}ta}\widehat{F})'+n(\mathrm{e}^{-\mathrm{i}ta}\widehat{G})^{n-1}\{(\mathrm{e}^{-\mathrm{i}ta}\widehat{F})'-(\mathrm{e}^{-\mathrm{i}ta}\widehat{G})'\}=$$

$$J_1+J_2+J_3.$$

Here

$$J_1=n(\mathrm{e}^{-\mathrm{i}ta}\widehat{F})^{n-1}-(\mathrm{e}^{-\mathrm{i}ta}\widehat{G})^{n-1}(\mathrm{e}^{-\mathrm{i}ta}\widehat{F})',$$

$$J_2=n(\mathrm{e}^{-\mathrm{i}ta}\widehat{G})^{n-1}\{\widehat{F}'-\widehat{G}'\}\mathrm{e}^{-\mathrm{i}ta},\quad J_3=n(\mathrm{e}^{-\mathrm{i}ta}\widehat{G})^{n-1}\mathrm{i}a\mathrm{e}^{-\mathrm{i}ta}(\widehat{G}-\widehat{F}).$$

We shall estimate each J_i separately. We have

$$|\,J_3\,|\leqslant na|\,\widehat{F}-\widehat{G}\,|\,|\,\widehat{G}\,|^{n-1}\leqslant C(s)na\kappa\Big|\,\sin\frac{t}{2}\Big|^s\exp\Big\{-2\lambda\sin^2\frac{t}{2}\Big\}\leqslant$$

$$C(s)\frac{\kappa}{n^{(s-3)/2}\lambda^{(s-1)/2}}\exp\Big\{-\lambda\sin^2\frac{t}{2}\Big\}.$$

Similarly,

$$|\,J_2\,|\leqslant n|\,\widehat{G}\,|^{n-1}|\,\widehat{F}'-\widehat{G}'\,|\leqslant C(s)\frac{\kappa}{n^{(s-3)/2}\lambda^{(s-1)/2}}\exp\Big\{-\lambda\sin^2\frac{t}{2}\Big\}.$$

For the estimate of J_1 we must note that

$$|\,(\mathrm{e}^{-\mathrm{i}ta}\widehat{F})'\,|=|\,\mathrm{e}^{-\mathrm{i}ta}\widehat{F}'-\mathrm{i}a\mathrm{e}^{-\mathrm{i}ta}\widehat{F}\,|\leqslant|\,\widehat{F}'-\mathrm{i}a\,|+a|\,\widehat{F}-1\,|\leqslant$$

$$(\nu_2(F)+\nu_1^2(F))|\,\mathrm{e}^{\mathrm{i}t}-1\,|=2(\nu_2(F)+\nu_1^2(F))\Big|\,\sin\frac{t}{2}\Big|.$$

Consequently,

$$| J_1 | \leqslant n| (\mathrm{e}^{-\mathrm{i}ta}\widehat{F})' | | \widehat{F}^{n-1} - \widehat{G}^{n-1} | \leqslant$$

$$C(s)n^2\Big| \sin\frac{t}{2}\Big|^s \exp\Big\{-2n\lambda\sin^2\frac{t}{2}\Big\}| (\mathrm{e}^{-\mathrm{i}ta}\widehat{F})' | \leqslant$$

$$C(s)\frac{\kappa}{n^{(s-3)/2}\lambda^{(s+1)/2}}(\nu_2(F)+\nu_1^2(F))\exp\Big\{-2n\lambda\sin^2\frac{t}{2}\Big\}.$$

Combining all obtained estimates with (5.6), for $b = \sqrt{n\lambda}$, we get the following final estimate

$$\| F^n - G^n \| \leqslant C(s)\frac{\kappa}{n^{(s-2)/2}\lambda^{s/2}}\bigg(1+\frac{\nu_2(F)+\nu_1^2(F)}{\lambda}\bigg). \tag{5.12}$$

For the Poisson approximation to the Binomial distribution we get the same order of accuracy as in (5.7).

Approximation of symmetric distributions

Let $F \in \mathcal{F}_Z$, for all t, have nonnegative characteristic function $\widehat{F}(t) \geqslant 0$, do not depend on n, $F\{0\} > 0$ and let F have the finite second moment. Then

$$\| F^n - \exp\{n(F-E)\} \| \leqslant Cn^{-1}. \tag{5.13}$$

For the proof of (5.13) note that

$$\widehat{F}(t) \leqslant \exp\{\widehat{F}(t)-1\},$$

and

$$| \widehat{F}^n(t) - \exp\{n(\widehat{F}(t)-1)\} | \leqslant n\exp\{(n-1)(\widehat{F}(t)-1)\}| \widehat{F}(t) - \exp\{\widehat{F}(t)-1\} |$$

$$Cn\exp\{(n-1)(\widehat{F}(t)-1)\}| \widehat{F}(t)-1 |^2 \leqslant Cn\exp\{n(\widehat{F}(t)-1)/2\}. \tag{5.14}$$

Set

$$p_k = 2F\{k\}.$$

Then $\widehat{F}(t) = \sum_j p_j\cos(tj)$ and

$$| \widehat{F}'(t) | = | \sum_{j=0}^{\infty} jp_j\sin(tj) | \leqslant \Big(\sum_{j=0}^{\infty} jp_j\Big)^{1/2}\Big(\sum_{k=0}^{\infty} p_k\sin^2(tk)\Big)^{1/2} \leqslant$$

$$2\Big(\sum_{j=0}^{\infty} jp_j\Big)^{1/2}\Big(\sum_{k=0}^{\infty} p_k\sin^2\frac{tk}{2}\Big)^{1/2} = C(1-\widehat{F}(t))^{1/2}.$$

Therefore,

$$| (\widehat{F}^n(t) - \exp\{n(\widehat{F}(t)-1)\})' | \leqslant | n\widehat{F}^{n-1}(t)\widehat{F}'(t) - n\widehat{F}'(t)\exp\{n(\widehat{F}(t)-1)\} | \leqslant$$

$$n| \widehat{F}'(t) | | \widehat{F}^{n-1}(t)(1-\widehat{F}(t)) + \widehat{F}^n(t) - \exp\{n(\widehat{F}(t)-1)\} | \leqslant$$

$$Cn(1-\widehat{F}(t))^{1/2}\exp\{(n-1)(\widehat{F}(t)-1)\}\Big((1-\widehat{F}(t))+(1-\widehat{F}(t))^2\Big)\leqslant$$
$$Cn^{-1/2}\exp\{n(\widehat{F}(t)-1)/2\}.$$

Taking $b=\sqrt{n}$ from Lemma 5.1 we get

$$\|\, F^n-\exp\{n(F-E)\}\,\|^2\leqslant Cn^{-2}\sqrt{n}\int\limits_{-\pi}^{\pi}\exp\{n(\widehat{F}(t)-1)\}\,\mathrm{d}t.$$

To end the proof of (5.13) one should note that by Lemma 3.2 we have

$$\int\limits_{-\pi}^{\pi}\exp\{n(\widehat{F}(t)-1)\}\,\mathrm{d}t\leqslant Cn^{-1/2}.$$

5.2 Total variation and the Barbour-Xia inequality

So far we encountered two situations: the general case when we must impose very stringent conditions on the behavior of distributions (obtaining estimates such as np^2), or assuming the Franken type condition. The main idea of this subsection is to show that F^n and G^n can be close even if Franken's condition fails. In doing so we utilize one result of Barbour and Xia (1999) about the difference in total variation between consequent convolutions of lattice distributions. Let $S=\xi_1+\xi_2+\cdots+\xi_n$ be independent integer valued random variables,

$$K=\sum_{i=1}^n v_i,\quad v_i=\min\{1/2;1-\|\,\mathcal{L}(\xi_i+1)-\mathcal{L}(\xi_i)\,\|/2\},\quad v^*=\max_i v_i.$$

Then

$$\|\,\mathcal{L}(S)(E_1-E)\,\|=\|\,\mathcal{L}(S+1)-\mathcal{L}(S)\,\|\leqslant 2K^{-1/2}. \tag{5.15}$$

Moreover, set $S^i=S-\xi_i$, then

$$\sup_i\|\,\mathcal{L}(S^i)(E_1-E)^2\,\|\leqslant 8(K-v^*)^{-1}. \tag{5.16}$$

We shall formulate Barbour-Xia estimate for identically distributed summands in terms of convolutions. Let $F\in\mathcal{F}_Z$, $G\in\mathcal{F}_Z$. Set

$$u_1=1-\|\,F(E_1-E)\,\|/2,\qquad u_2=1-\|\,G(E_1-E)\,\|/2. \tag{5.17}$$

The Barbour-Xia estimate can be formulated in the following way. For any natural k,

$$\|\,F^k(E_1-E)\,\|\leqslant 2(ku_1)^{-1/2}. \tag{5.18}$$

The proof of (5.18) employs special facts from the random walks and is beyond the scopes of our course. Note that similarly

$$\|\,G^k(E_1-E)\,\|\leqslant 2(ku_2)^{-1/2}. \tag{5.19}$$

We assume that the right-hand-sides in (5.18) and (5.19) are infinite, if u_1 or u_2 equals zero. The estimate (5.18) is quite sharp. However, it does not take into account possible symmetry of distributions. Combining Le Cam's approach with (5.18) we get the following general result.

Lemma 5.2 *Let, F and G be distributions concentrated on non-negative integers; for some fixed $s \geqslant 2$, $\nu_k(F) = \nu_k(G)$, $k = 1, \dots, s-1$, $\nu_s(F) + \nu_s(G) < \infty$ and $n > 6(s+1)$. Then*

$$\| F^n - G^n \| \leqslant (\nu_s(F) + \nu_s(G)) n^{-(s-2)/2} \big(u_1^{-s/2} + u_2^{-s/2}\big)(12s)^{s/2}/s!\,. \tag{5.20}$$

Proof. Let $[\cdot]$ denote the integral part of the indicated expression. Set $v(n,s) = [\,[n/2]/s]$. From the equality of factorial moments we get the following expansion in distributions

$$F - G = W(E_1 - E)^s (\nu_s(F) + \nu_s(G))/s!,$$

where W is a finite measure satisfying $\| W \| \leqslant 1$, see (2.6). Taking into account the properties of total variation norm we get:

$$\| F^n - G^n \| = \Big\| \sum_{m=0}^{n-1} F^m G^{(n-m-1)}(F - G) \Big\| \leqslant$$

$$\sum_{m=0}^{n-1} \| F^m G^{(n-m-1)}(F-G) \| \leqslant n\Big(\| F^{[n/2]}(F-G) \| + \| G^{[n/2]}(F-G) \|\Big) \leqslant$$

$$n(\nu_s(F)/s! + \nu_s(G)/s!)\Big(\| F^{[n/2]}(E_1 - E)^s \| + \| G^{[n/2]}(E_1 - E)^s \|\Big) \leqslant$$

$$n(\nu_s(F)/s! + \nu_s(F)/s!)\Big(\| F^{v(n,s)} \| (E_1 - E)^s + \| G^{v(n,s)}(E_1 - E) \|^s\Big) \leqslant$$

$$2^s n(\nu_s(F)/s! + \nu_s(G)/s!)\big((v(n,s)u_1)^{-s/2} + (v(n,s)u_2)^{-s/2}\big) \leqslant$$

$$(\nu_s(F) + \nu_s(G)) n^{-(s-2)/2}\big(u_1^{-s/2} + u_2^{-s/2}\big)(12s)^{s/2}/s!\,.$$

In the last inequality we used the following estimate

$$v(n,s) \geqslant [n/2]/s - 1 \geqslant n/(2s) - 1/s - 1 \geqslant n/(3s),$$

which follows from $n \geqslant 6(s+1)$. The proof of (5.20) is completed.

Remark 5.1 *Note that for unimodal distribution F, we can use the following estimate:*

$$u_1 \geqslant 1 - \max_k F\{k\}.$$

Indeed, let $F\{k\} \leqslant F\{k+1\}$, for $k \leqslant m$; and let $F\{k\} \geqslant F\{k+1\}$, for $k > m$. Then

$$\| F(E_1 - E) \| = \sum_{k=-\infty}^{\infty} | F\{k\} - F\{k-1\} | \leqslant 2F\{m\} \leqslant 2 \max_k F\{k\}.$$

We shall exemplify (5.20) assuming G to have the geometric (G^n to have the negative binomial) distribution, i.e. let G have the characteristic function:

$$\widehat{G}^n(t) = (p/(1 - q\mathrm{e}^{\mathrm{i}t}))^n, \quad q \leqslant 1/2, \quad p + q = 1.$$

Then

$$\widehat{G}^n(t) = \exp\Big\{n \sum_{m=1}^{\infty} \frac{q^m}{m}(\mathrm{e}^{itm} - 1)\Big\}. \tag{5.21}$$

Expanding $\widehat{G}(t)$ and $\ln \widehat{G}(t)$ in the powers of $(\mathrm{e}^{it} - 1)$ we establish that factorial moments and factorial cumulants of G equal, respectively,

$$k!\left(\frac{q}{p}\right)^k \quad \text{and} \quad (k-1)!\left(\frac{q}{p}\right)^k, \quad k = 1, 2, \dots$$

It is easy to check that Franken's condition (3.18) means that $q/p < 1/3$. Let ξ be concentrated at 4 points and have the following distribution:

$$P(\xi = 0) = 20/45, \quad P(\xi = 1) = 18/45, \quad P(\xi = 3) = 5/45, \quad P(\xi = 6) = 2/45.$$

Consequently,

$$\widehat{F}(t) = 1 + (\mathrm{e}^{it} - 1) + (\mathrm{e}^{it} - 1)^2 + (\mathrm{e}^{it} - 1)^3 + \theta |\,\mathrm{e}^{it} - 1\,|^4/3. \tag{5.22}$$

It is easy to check that $\nu_1 = 1$. Consequently, Franken's condition is not satisfied. However, it is easy to check that

$$u_1 = 1 - \big(20/45 + |20/45 - 18/45| + 18/45 + 5/45 + 5/45 + 2/45 + 2/45\big)/2 = 18/45. \tag{5.23}$$

Let η have the geometric distribution defined by (5.21) with $p = q = 1/2$. Once again, we see that Franken's conditions (5.9) is not satisfied. Moreover,

$$\widehat{G}(t) = 1 + (\mathrm{e}^{it} - 1) + (\mathrm{e}^{it} - 1)^2 + (\mathrm{e}^{it} - 1)^3 + \theta |\,\mathrm{e}^{it} - 1\,|^4. \tag{5.24}$$

The quantity u_2 can be easily computed for any geometric distribution, because $\|\, GE_1 - D \,\|$ equals to

$$p + p(1-q) + pq(1-q) + pq^2(1-q) + pq^3(1-q) + \dots = p + p^2(1 + q + q^2 + \dots) = 2p.$$

Consequently, for our example,

$$u_2 = 1 - p = 1 - 1/2 = 1/2. \tag{5.25}$$

Combining (5.22) – (5.25) with the statement of Lemma we get the following corollary.

Corollary 5.1 *Let F and G be defined as in above. Then*

$$\|\, F^n - G^n \,\| \leqslant Cn^{-1}.$$

In our example, $s = 4$. Consequently, Lemma can be applied when $n > 6(4+1) = 30$ only. However, if $n \leqslant 30$, corollary follows from the fact that the difference of two distributions is less then 1.

Exercises

1. Let $\lambda > 0$, $k \in \{1, 2 \dots\}$. To prove that

$$\|\, (E_1 - E)^k \exp\{\lambda(E_1 - E)\} \,\| \leqslant C(k)\lambda^{-k/2}.$$

2. Let $p < 1/5$. To get the estimate for

$$\|\, ((1 - 2p)E + pE_1 + pE_{-1})^n (E_1 - E) \,\|.$$

3. Let F be concentrated on non-negative integers. Let F have finite ν_1, and let F do not depend on n in any way. To get the estimate for $\|\, F^n - \exp\{n(F - E)\} \,\|$. When the obtained estimate is trivial? To give sufficient condition for the estimate to be of the order $Cn^{-1/2}$.

Bibliographical notes

Another version of (5.1) can be found in Presman (1983). Inequality (5.1) was extended to the case of l_p metrics in Šiaulys and Čekanavičius (1988). Prohorov (1953) was the first to prove the estimate (5.7). He used direct asymptotic expansion of the binomial probability. Improvement of constant in (5.7) is among the most comprehensively studied subjects in Probability theory, see Kerstan (1964), Deheuvels and Pfeifer (1986), Barbour et al. (1992), and references therein. The inequality was proved in Barbour and Xia (1999).

6 Nonuniform estimates

In many aspects nonuniform estimates for $W \in \mathcal{M}_Z$ can be obtained very similarly to the estimates in total variation.

6.1 Local estimates

Let $W \in \mathcal{M}_Z$ and let $\sum_{k\in\mathbb{Z}} k^2 |\,W\{k\}\,| < \infty$. Note that

$$\widehat{W}(\pi) = \sum_{k\in\mathbb{Z}} W\{k\}\mathrm{e}^{\mathrm{i}k\pi} = \sum_{k\in\mathbb{Z}} W\{k\}\cos(k\pi) = \sum_{k\in\mathbb{Z}} W\{k\}\cos(-k\pi) = \widehat{W}(-\pi).$$

Therefore integrating by parts the formula of inversion (3.1), for $a \neq k$, we get

$$W\{k\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \widehat{W}(t)\mathrm{e}^{-\mathrm{i}tk}\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Big(\widehat{W}(t)\mathrm{e}^{-\mathrm{i}ta}\Big)\mathrm{e}^{\mathrm{i}t(a-k)}\,\mathrm{d}t =$$

$$\frac{1}{i(k-a)}\frac{1}{2\pi}\int_{-\pi}^{\pi}\Big(\widehat{W}(t)\mathrm{e}^{-\mathrm{i}ta}\Big)'\mathrm{e}^{\mathrm{i}t(a-k)}\,\mathrm{d}t = -\frac{1}{(k-a)^2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\Big(\widehat{W}(t)\mathrm{e}^{-\mathrm{i}ta}\Big)''\mathrm{e}^{\mathrm{i}t(a-k)}\,\mathrm{d}t. \qquad (6.1)$$

Consequently, for any $a \in \mathbb{R}$,

$$|\,k-a\,|\,|\,W\{k\}\,| \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big|\Big(\widehat{W}(t)\mathrm{e}^{-\mathrm{i}ta}\Big)'\Big|\,\mathrm{d}t, \qquad (6.2)$$

$$(k-a)^2\,|\,W\{k\}\,| \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big|\Big(\widehat{W}(t)\mathrm{e}^{-\mathrm{i}ta}\Big)''\Big|\,\mathrm{d}t. \qquad (6.3)$$

In general, the process can be continued establishing estimates of the higher order. However, we must emphasize, that additional derivative usually means the reducing of the accuracy.

Nonuniform local Poisson approximation to the Binomial law

We already know that

$$\Big|\frac{\mathrm{d}}{\mathrm{d}t}\Big(\mathrm{e}^{-\mathrm{i}tnp}\Big\{((1-p)+p\mathrm{e}^{\mathrm{i}t})^n - \exp\{np(\mathrm{e}^{\mathrm{i}t}-1)\}\Big\}\Big)\Big| \leqslant$$

$$Cnp^2\Big|\sin\frac{t}{2}\Big|\exp\Big\{-2np(1-p)\sin^2\frac{t}{2}\Big\}, \qquad (6.4)$$

see (5.5). Consequently, from (6.2) we obtain the estimate

$$|\,k-np\,|\,|\,((1-p)E+p_1E_1)^n\{k\} - \exp\{np(E_1-E)\{k\}\}\,| \leqslant Cp. \qquad (6.5)$$

Usually nonuniform estimates are given in a more convenient form. Combining local estimate with (3.8), for $0 < p < 1/2$, we obtain

$$\Big(1+\frac{|\,k-np\,|}{\sqrt{np}}\Big)\Big|\,((1-p)E+pE_1)^n\{k\}-\exp\{np(E_1-E)\}\{k\}\,\Big| \leqslant C\sqrt{\frac{p}{n}}. \tag{6.6}$$

Similarly, we can get estimate for $(k-np)^2$. Note that, if the second-order nonuniform local estimates exist, then summing them it is always possible to get the estimate in total variation. However, such indirect approach means much larger absolute constant.

Nonuniform local Franken type estimates

As in previous sections, we assume that F and G are concentrated on non-negative integers, $s > 1$, $\kappa = \nu_s(F) + \nu(G)_s < \infty$ and $\nu(F)_k = \nu(G)_k$, for $k = 1, 2, \dots, s$ and let

$$\lambda = \min\{\nu_1(F) - \nu_1^2(F) - \nu_2(F), \nu_1(G) - \nu_1^2(G) - \nu_2(G)\} > 0. \tag{6.7}$$

Then, as proved in Section for estimates in total variation,

$$|\,\widehat{F}^n(t) - \widehat{G}^n(t)\,| \leqslant C(s)n\kappa|\,\sin\frac{t}{2}\,|^s \tag{6.8}$$

and

$$\Big|\Big(\mathrm{e}^{-\mathrm{i}tan}(\widehat{F}(t)-\widehat{G}(t))\Big)'\Big| \leqslant C(s)\frac{\kappa}{n^{(s-3)/2}\lambda^{(s-1)/2}}\Big(1+\frac{\nu_2(F)+\nu_1^2(F)}{\lambda}\Big). \tag{6.9}$$

Therefore, for any $k \in \mathbb{Z}$, $a \in \mathbb{R}$,

$$\Big(1+\frac{|\,k-n\nu_1(F)\,|}{\sqrt{n\lambda}}\Big)|\,F^n\{k\}-G^n\{k\}\,| \leqslant$$

$$C(s)\frac{\kappa}{n^{(s-1)/2}\lambda^{(s+1)/2}}\Big(1+\frac{\nu_2(F)+\nu_1^2(F)}{\lambda}\Big). \tag{6.10}$$

Estimate (6.10) is of the right order for the Binomial distribution. In principle, the second order estimates can be obtained as well.

6.2 Integral estimates

The idea of inversion formula for the nonuniform integral estimates is the same as for the local ones. Let $W \in \mathcal{M}_Z$,

$$U(t) = \frac{\widehat{W}(t)}{\mathrm{e}^{-\mathrm{i}t}-1}. \tag{6.11}$$

We shall assume that $U(t)$ has two continuous derivatives. Then by the formula of inversion (4.3) we get

$$W\{(-\infty, m)\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} U(t)\mathrm{e}^{-\mathrm{i}tm}\,\mathrm{d}t.$$

Note that $U(\pi) = U(-\pi)$. Therefore, applying the same reasoning as in the local case, for any $a \in \mathbb{R}$ and $k \in \mathbb{Z}$, we get

$$|\,k-a\,|\,|\,W\{(-\infty,k)\}\,| \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big|\big(U(t)\mathrm{e}^{-\mathrm{i}ta}\big)'\Big|\,\mathrm{d}t, \tag{6.12}$$

$$(k-a)^2\,|\,W\{(-\infty,k)\}\,| \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big|\Big(U(t)\mathrm{e}^{-\mathrm{i}ta}\Big)''\Big|\,\mathrm{d}t. \tag{6.13}$$

Now the estimates can be obtained in the same way as for the local case.

Nonuniform integral Poisson approximation to the Binomial law

Let $W = ((1-p)E + pE_1)^n - \exp\{np(E_1 - E)\}$, $0 < p < 1/2$. We already proved estimates

$$|\,\widehat{W}(t)\,| \leqslant Cnp^2\sin^2\frac{t}{2}\exp\Big\{-2np(1-p)\sin^2\frac{t}{2}\Big\},$$

$$|\,(\widehat{W}(t)\mathrm{e}^{-\mathrm{i}tnp})'\,| \leqslant Cnp^2\Big|\sin\frac{t}{2}\Big|\exp\Big\{-2np(1-p)\sin^2\frac{t}{2}\Big\}.$$

Taking into account that

$$|\,(U(t)\mathrm{e}^{-\mathrm{i}tnp})'\,| \leqslant \frac{|\,(\widehat{W}(t)\mathrm{e}^{-\mathrm{i}tnp})'\,|}{|\,\mathrm{e}^{-\mathrm{i}t}-1\,|} + \frac{|\,\widehat{W}(t)\,|}{|\,\mathrm{e}^{-\mathrm{i}t}-1\,|^2},$$

we easily obtain the corresponding nonuniform estimate. Once again we shall write it in a more standard form combining with the uniform result:

$$\left(1+\frac{|\,k-np\,|}{\sqrt{np}}\right)\Big|\,((1-p)E+pE_1)^n\{(-\infty,k)\} - \exp\{np(E_1-E)\}\{(-\infty,k)\}\,\Big| \leqslant$$

$$C\min\{np^2, p\}. \tag{6.14}$$

In above, we assumed that $p < 1/2$. It is not difficult to get the estimate for $1/2 < p < 1$, however, the usefulness of it will be doubtful.

Symmetric distributions

In this section we begin investigating of one special subset of all symmetric distributions. We shall consider distributions having nonnegative characteristic functions, i.e. distributions from the set $\mathcal{F}_+$. It is easy to understand that distribution of two independent random variables $\xi_1 - \xi_2$ with the same distribution F has the characteristic function $|\,\widehat{F}(t)\,|^2$ and belongs to $\mathcal{F}_+$. Let $F \in \mathcal{F}_+ \cap \mathcal{F}_Z$,

$$p_k = 2F\{k\},\quad \sigma^2 = \sum_{k=1}^{\infty}k^2p_k,\quad G = \exp\{F-E\},\quad r(t) = \sum_{k=1}^{\infty}p_k\sin^2\frac{tk}{2}.$$

Using Hölder's inequality we obtain

$$\widehat{F}(t) = 1 + \sum_{k=1}^{\infty} p_k(\cos tk - 1) \leqslant \widehat{G}(t) = \exp\{-2r(t)\},$$

$$|\,\widehat{F}'(t)\,| = \Big|\sum_{k=1}^{\infty} kp_k \sin tk\Big| \leqslant \sum_{k=1}^{\infty} k\sqrt{p_k}\sqrt{p_k|\,\sin tk\,|} \leqslant C\sigma\sqrt{r(t)},$$

$$|\,\widehat{F}''(t)\,| = \Big|\sum_{k=1}^{\infty} k^2 p_k \cos tk\Big| \leqslant \sigma^2,$$

$$|\,\widehat{G}'(t)\,| \leqslant |\,\widehat{G}(t)\,|\,|\,\widehat{F}'(t)\,| \leqslant C\sigma\sqrt{r(t)},$$

$$|\,\widehat{G}''(t)\,| \leqslant C\sigma^2, \quad |\,r(t)\,|\sigma^2 t^2, 1 \leqslant \mathrm{e}^2\mathrm{e}^{-2r(t)},$$

$$|\,\widehat{F}(t) - \widehat{G}(t)\,| \leqslant C|\,\widehat{F}(t) - 1\,|^2 \leqslant Cr^2(t),$$

$$|\,\widehat{F}'(t) - \widehat{G}'(t)\,| \leqslant |\,\widehat{F}'(t)\,|\,|\,\widehat{F}(t) - 1\,| \leqslant C\sigma r^{3/2}(t),$$

$$|\,\widehat{F}''(t) - \widehat{G}''(t)\,| \leqslant C\Big\{|\,\widehat{F}''(t)\,|\,|\,\widehat{F}(t) - 1\,| + |\,\widehat{F}'(t)\,|^2\Big\} \leqslant C\sigma^2 r(t).$$

Using these estimates we prove that

$$|\,\widehat{F}^n(t) - \widehat{G}^n(t)\,| \leqslant Cn|\,\widehat{F}(t) - \widehat{G}(t)\,|\mathrm{e}^{-2nr(t)} \leqslant Cnr^2(t)\mathrm{e}^{-2nr(t)} \leqslant C\sigma^3\sqrt{n}|\,t\,|^3\mathrm{e}^{-nr(t)},$$

$$|\,(\widehat{F}^n(t) - \widehat{G}^n(t))'\,| \leqslant C\sigma^3\sqrt{n}t^2\mathrm{e}^{-nr(t)},$$

$$|\,(\widehat{F}^n(t) - \widehat{G}^n(t))''\,| \leqslant C\sigma^3\sqrt{n}|\,t\,|\mathrm{e}^{-nr(t)}.$$

Let $W = F^n - \exp\{n(F-E)\}$. By using (1.2) (or expansion of $\widehat{F}(t) - 1$ in powers of t) we can prove that $U(t)$ and its first and second derivatives are continuous. We have

$$|\,U''(t)\,| \leqslant C\bigg(\frac{|\,\widehat{W}''(t)\,|}{|\,t\,|} + \frac{|\,\widehat{W}'(t)\,|}{t^2} + \frac{|\,\widehat{W}(t)\,|}{|\,t\,|^3}\bigg).$$

Then by (3.15)we get

$$\int_{-\pi}^{\pi} \mathrm{e}^{-nr(t)}\,\mathrm{d}t \leqslant Cn^{-1/2}(1 - F\{0\})^{-1/2}.$$

Combining all estimates with (6.13) and (4.1) we get the following nonuniform integral estimate. For any $k \in \mathbb{Z}$,

$$\bigg(1 + \frac{m^2}{\max\{1, n\sigma^2\}}\bigg)\Big|\,F^n\{(-\infty, m)\} - \exp\{n(F-E)\}\{(-\infty, m)\}\,\Big| \leqslant$$

$$C\sigma(1 - F\{0\})^{-1/2}n^{-1}. \tag{6.15}$$

Note that, for obtaining (6.15) we have not used centring. Indeed, the mean of F is already equal to zero.

6.3 Fortet - Mourier distance

Let $W \in \mathcal{M}_Z$ and let $W\{\mathbb{R}\} = 0$. We recall that then the analogue of the Fortet-Mourier distance for W is

$$|\, W\,|_{FM} = \sum_{k=-\infty}^{\infty} |\, W\{(-\infty, k)\}\,| = \sum_{k=-\infty}^{\infty} |\, W\{[k, \infty)\}\,|.$$

It is obvious, that the simplest way to obtain estimates for $|\, W\,|_{FM}$ is to sum nonuniform integral estimates of the second order. Let $F \in \mathcal{F}_+ \cap \mathcal{F}_Z$, $\sigma^2 = \sum_{k=-\infty}^{\infty} k^2 F\{k\}$ and let $n\sigma > 1$. Then from (6.15) and (1.10) we get that

$$|\, F^n - \exp\{n(F-E)\}\,|_{FM} \leqslant C\sigma^2 (1 - F\{0\})^{-1/2} n^{-1/2}.$$

We must bear in mind that to use nonuniform estimates for the Fortet-Mourier distance means quite large constants.

Exercises

1. Let $F \in \mathcal{F}_Z \cap \mathcal{F}_+$ as in (6.15). To get nonuniform local estimate of the second order for $|\, F^n - \exp\{n(F-E\}\,|$.

2. Let $F \in \mathcal{F}_Z \cap \mathcal{F}_+$ as in (6.15). To get nonuniform local estimate of the second order for $|\, F^{n+1} - F^n\,|$.

3. Let $F \in \mathcal{F}_Z$. What conditions are sufficient for the nonuniform local estimate of $F^n - \exp\{n(F-E)\}$.

4. Let $0 < p < 1/2$. To get nonuniform integral estimate for

$$\big((1-p)E + pE_1\big)^n - \exp\Big\{np(E_1 - E) - \frac{np^2}{2}(E_1 - E)^2\Big\}.$$

5. To prove that

$$|\, ((1-p)E + pE_1)^n - \exp\{np(E_1 - E)\}\,|_{FM} \leqslant Cp\sqrt{np}.$$

Bibliographical notes

The section reproduces somewhat simplified version of Čekanavičius (1993). Other applications can be found in Čekanavičius and Kruopis (2000).

7 Lower bounds

7.1 Estimating total variation through Fourier-Stieltjes transform

In general, one usually uses the following relation

$$|\,W\,| \leqslant \|\,W\,\|$$

and obtains lower bound estimates for the uniform metric. However, the following very simple relation between total variation norm and the Fourier-Stieltjes transform allows direct estimates:

$$|\,\widehat{W}(t)\,| \leqslant \|\,W\,\|. \tag{7.1}$$

Though (7.1) holds in the general situation, we shall prove it for $W \in \mathcal{M}_Z$ only. We have

$$|\,\widehat{W}(t)\,| = \Big|\sum_{k\in\mathbb{Z}} W\{k\}\mathrm{e}^{itk}\Big| \leqslant \sum_{k\in\mathbb{Z}} |\,W\{k\}\,| = \|\,W\,\|.$$

This inequality holds for all t. Therefore, we can estimate $|\,\widehat{W}(t)\,|$ suitably choosing t. We shall demonstrate the above technique on one special case. Let $F \in \mathcal{F}_+ \cap \mathcal{F}_Z$, $\sigma^2 = \sum_1^\infty k^2 2F\{k\}$. We shall assume that F does not depend on n in any way, i.e. we assume that we have the scheme of sequences. Moreover, we assume that F has the finite fourth absolute moment and $F \neq E$. Then we shall prove that

$$\|\,F^n - \exp\{n(F-E)\}\,\| \geqslant Cn^{-1}. \tag{7.2}$$

Let us consider expansion of the characteristic function in powers of t. Taking into account the symmetry of F we have

$$\widehat{F}(t) = 1 - \frac{\sigma^2 t^2}{2} + \theta C t^4.$$

Consequently, we can choose $t = t_0$ in such a way that

$$1 - \widehat{F}(t_0) = \frac{1}{an}.$$

The quantity a will be determined later. Taking into account (1.1), and (1.6) similarly to example 1.1 we get the following sequence of estimates (for the sake of brevity we omit t_0):

$$|\,\widehat{F}^n - \exp\{n(\widehat{F}-1)\} - n\exp\{(n-1)(\widehat{F}-1)\}(\widehat{F} - \exp\{\widehat{F}-1\})\,| \leqslant$$

$$\sum_{m=2}^{n} \binom{m-1}{2} |\,\widehat{F}\,|^{n-m}\,|\,\exp\{(m-s-1)(\widehat{F}-1)\}\,|\,|\,\widehat{F} - \exp\{\widehat{F}-1\}\,|^2 \leqslant$$

$$\binom{n}{2}\frac{1}{4}|\,\widehat{F}-1\,|^4 \leqslant \frac{1}{8a^4n^2},$$

$$\Big|n\exp\{(n-1)(\widehat{F}-1)\}\Big(1 + (\widehat{F}-1) - \frac{1}{2}(\widehat{F}-1)^2 - \exp\{\widehat{F}-1\}\Big)\Big| \leqslant$$

$$n\frac{|\,\widehat{F}-1\,|^3}{6}\leqslant\frac{1}{6a^3n^2},$$

$$\left|n\frac{(\widehat{F}-1)^2}{2}\big(\exp\{(n-1)(\widehat{F}-1)\}-1\big)\right|\leqslant n(n-1)\frac{|\,\widehat{F}-1\,|^3}{2}\leqslant\frac{1}{2a^3n},$$

$$\frac{n}{2}|\,n(\widehat{F}-1)\,|^2=\frac{1}{2a^2n}.$$

Collecting all estimates we see that, for $t=t_0$,

$$|\,\widehat{F}^n-\exp\{n(\widehat{F}-1)\}\,|\geqslant\frac{1}{2a^2n}\Big(1-\frac{1}{a}-\frac{1}{6a^2n}-\frac{1}{8a^2n}\Big).$$

To finish the proof of (7.2) we should choose a sufficiently large. Foe example, we can take $a=2$.

7.2 Local lower bound estimates

Let $W\in\mathcal{M}_Z$. The general idea of the lower bounds estimates are can be expressed in the following way

$$|\,WH\,|_\infty\leqslant|\,W\,|_\infty\|\,H\,\|.$$

Here $H\in\mathcal{M}$. Dividing both sides of the estimate by $\|\,H\,\|$ we get the lower bound estimate. Consequently, we need H having good properties. Usually convolution with some 'good' measure is called smoothing. There are various possible choices for H. We shall choose H to be quite close to the normal distribution.

Lemma 7.1 *Let $W\in\mathcal{M}_Z$, $b\geqslant1$, $a\in\mathbb{R}$, $s=1,2$,*

$$\widehat{\psi}_1(t)=\mathrm{e}^{-t^2/2},\quad\widehat{\psi}_2(t)=t\mathrm{e}^{-t^2/2},\tag{7.3}$$

$$V_j(b)=\int\limits_{-\infty}^{\infty}\widehat{\psi}_j(t)\widehat{W}(t/b)\exp\{-ita/b\}\,\mathrm{d}t.\tag{7.4}$$

Then

$$|\,W\,|_\infty\geqslant C_0|\,V_j(b)\,|b^{-1}.\tag{7.5}$$

Here $C_0>0.1$.

Proof. It is known that

$$\int\limits_{-\infty}^{\infty}\widehat{\psi}_j(t)\mathrm{e}^{\mathrm{i}ty}\,\mathrm{d}t=\sqrt{2\pi}\psi_j(y),\quad\psi_1(y)=\mathrm{e}^{-y^2/2},\ \psi_2(y)=\mathrm{i}y\mathrm{e}^{-y^2/2}.$$

Moreover,

$$\sum_{k\in\mathbb{Z}} \psi_1\Big(\frac{k-a}{b}\Big) \leqslant 1 + b\sqrt{2\pi}, \quad \sum_{k\in\mathbb{Z}} \psi_2\Big(\frac{k-a}{b}\Big) \leqslant 2(\mathrm{e}^{-1/2} + b). \tag{7.6}$$

Now we get the following estimate:

$$V_j(b) = \int_{-\infty}^{\infty} \exp\{-\mathrm{i}ta/b\}\widehat{\psi}_j(t)\sum_{k\in\mathbb{Z}} W\{k\}\exp\{\mathrm{i}kt/b\}\,\mathrm{d}t =$$

$$\sum_{k\in\mathbb{Z}} W\{k\}\int_{-\infty}^{\infty} \exp\{\mathrm{i}t(k-a)/b\}\widehat{\psi}_j(t)\,\mathrm{d}t = \sqrt{2\pi}\sum_{k\in\mathbb{Z}} W\{k\}\psi_j\Big(\frac{k-a}{b}\Big). \tag{7.7}$$

Consequently,

$$|\,W\,|_\infty \geqslant |\,V_s(b)\,|\bigg(\sum_{k\in\mathbb{Z}}\Big|\psi_j\Big(\frac{k-a}{b}\Big)\Big|\bigg)^{-1}$$

and we can apply (7.6). $\square$.

Why do we need two possible $V_j(b)$? The answer is determined by the method of proof. Just like in the previous section, we shall expand $\widehat{W}(t)$ in powers of t. The result will be of the form:

$$|\,\widehat{V}_j(t/b)\,| \geqslant C_1(W,b)\bigg|\int_{-\infty}^{\infty} t^s\widehat{\psi}_j(t)\,\mathrm{d}t\bigg| - C_2(W,b)\bigg|\int_{-\infty}^{\infty} t^{s+1}\widehat{\psi}_j(t)\,\mathrm{d}t\bigg| - \dots$$

Now the first integral turns zero if the function is odd. Consequently, we should choose $\widehat{\psi}_1(t)$ if s is even and $\widehat{\psi}_(t)$ if s is odd.

We demonstrate the method on the Poisson approximation to the Binomial law. Let $0 < p \leqslant 1/2$. In principle, we repeat the same scheme as in previous section. However we need to center distributions properly. Let, $W = ((1-p)E + pE_1)^n - \exp\{np(E_1 - E)\}$. Then by (1.3)

$$\widehat{W}(t)\mathrm{e}^{-\mathrm{i}tnp} =$$

$$(1-p+p\mathrm{e}^{\mathrm{i}t} - \exp\{p(\mathrm{e}^{\mathrm{i}t}-1)\})\sum_{k=1}^{n} \mathrm{e}^{-\mathrm{i}tp}\mathrm{e}^{-\mathrm{i}t(n-1)p}(1-p+p\mathrm{e}^{\mathrm{i}t})^{k-1}\exp\{(n-k)p(\mathrm{e}^{\mathrm{i}t}-1)\}. \tag{7.8}$$

We have

$$1 - p + p\mathrm{e}^{\mathrm{i}t} - \exp\{p(\mathrm{e}^{\mathrm{i}t}-1)\} = 1 + p(\mathrm{e}^{\mathrm{i}t}-1) - \sum_{j=0}^{\infty}\frac{p^j}{j!}(\mathrm{e}^{\mathrm{i}t}-1)^j =$$

$$-\frac{p^2}{2}(\mathrm{e}^{\mathrm{i}t}-1)^2 + R_1(t) = -\frac{p^2(\mathrm{i}t)^2}{2} + R_1(t) + R_2(t).$$

Here

$$|\,R_1(t)\,| \leqslant C_3p^3|\,\mathrm{e}^{\mathrm{i}t}-1\,|^3 \leqslant C_4p^3|\,t\,|^3, \quad |\,R_2(t)\,| \leqslant C_5p^2|\,t\,|^3.$$

The next step is to submit these inequalities into (7.8). However, we shall also replace $\mathrm{e}^{-\mathrm{i}tp}$ by unity and take into account that $|\,\mathrm{e}^{-\mathrm{i}tp}-1\,|\leqslant p\,|\,t\,|$. Consequently,

$$\widehat{W}(t)\mathrm{e}^{-\mathrm{i}tnp} =$$

$$\frac{p^2t^2}{2}\sum_{k=1}^{n}\mathrm{e}^{-\mathrm{i}tp}\mathrm{e}^{-\mathrm{i}t(n-1)p}(1-p+p\mathrm{e}^{\mathrm{i}t})^{k-1}\exp\{(n-k)p(\mathrm{e}^{\mathrm{i}t}-1)\}+\theta n(|\,R_1(t)+R_2(t)\,|=$$

$$\frac{p^2t^2}{2}\sum_{k=1}^{n}\mathrm{e}^{-\mathrm{i}t(n-1)p}(1-p+p\mathrm{e}^{\mathrm{i}t})^{k-1}\exp\{(n-k)p(\mathrm{e}^{\mathrm{i}t}-1)\}+R_3(t). \tag{7.9}$$

Here

$$|\,R_3(t)\,|\leqslant C_6np^2|\,t\,|^3.$$

For the centered characteristic function with two finite moments $\widehat{G}(t)$, from (1.19) we get that

$$|\,\widehat{G}(t)-1\,|\leqslant\frac{t^2}{2}\sigma^2(G),$$

where $\sigma^2(G)$ is the variance of G. Let us consider

$$G=E_{-(n-1)p}((1-p)E+pE_1)^{k-1}\exp\{(n-k)p(E_1-E)\}.$$

Note that G corresponds to the sum of independent Binomial and Poisson random variables. Hence, it has the variance equal to $(k-1)p(1-p)+(n-k)p\leqslant np$ and

$$\Big|\,\mathrm{e}^{-\mathrm{i}t(n-1)p}(1-p+p\mathrm{e}^{\mathrm{i}t})^{k-1}\exp\{(n-k)p(\mathrm{e}^{\mathrm{i}t}-1)\}-1\,\Big|\leqslant\frac{npt^2}{2}.$$

Therefore from the last estimate and (7.9) we get

$$\widehat{W}(t)\mathrm{e}^{-\mathrm{i}tnp}=\frac{np^2t^2}{2}+R_3(t)+R_4(t). \tag{7.10}$$

Here

$$|\,R_4(t)\,|\leqslant\frac{np^2t^2}{2}\cdot\frac{npt^2}{2}.$$

The second step is application of Lemma 7.1:

$$|\,V_1(b)\,|\geqslant\left|\int_{-\infty}^{\infty}\frac{np^2t^2}{2b^2}\mathrm{e}^{-t^2/2}\,\mathrm{d}t\right|-\int_{-\infty}^{\infty}\left|\,R_3\left(\frac{t}{b}\right)\,\right|\mathrm{e}^{-t^2/2}\,\mathrm{d}t-\int_{-\infty}^{\infty}\left|\,R_4\left(\frac{t}{b}\right)\,\right|\mathrm{e}^{-t^2/2}\,\mathrm{d}t\geqslant$$

$$C_7\frac{np^2}{b^2}-C_8\frac{np^2}{b^3}-C_9\frac{n^2p^3}{b^4}=C_{10}\frac{np^2}{b^2}\left(1-C_{11}\frac{1}{b}-C_{12}\frac{np}{b^2}\right). \tag{7.11}$$

Let $np<1$. Then taking sufficiently large $b\geqslant1$ from (7.11) we obtain

$$|\,V_1(b)\,|\geqslant Cnp^2. \tag{7.12}$$

Let $np \geqslant 1$. Then taking $b^2 = npC$ with sufficiently large C from (7.11) we obtain

$$|\, V_1(b)\,| \geqslant Cp. \tag{7.13}$$

Now from Lemma 7.1 we get that

$$|\, W\,|_\infty \geqslant C \min\Big\{ np^2, \sqrt{\frac{p}{n}} \Big\}.$$

Combining with the upper bound estimates, for $0 \leqslant p \leqslant 1/2$ we get the following inequality:

$$C_{13} \min\Big\{ np^2, \sqrt{\frac{p}{n}} \Big\} \leqslant \Big| \big((1-p)E + pE_1\big)^n - \exp\{np(E_1 - E)\} \Big|_\infty \leqslant$$

$$C_{14} \min\Big\{ np^2, \sqrt{\frac{p}{n}} \Big\}. \tag{7.14}$$

Thus, we see that the upper bound estimate was of the correct order.

7.3 Uniform lower bound estimates

Uniform lower bound estimates can be proved exactly in the same way as the local ones. We use the following Lemma.

Lemma 7.2 *Let $W \in \mathcal{M}_Z$, $b \geqslant 1$, $a \in \mathbb{R}$, $s = 1, 2$,*

$$\widehat{\psi}_1(t) = \mathrm{e}^{-t^2/2}, \quad \widehat{\psi}_2(t) = t\mathrm{e}^{-t^2/2}, \tag{7.15}$$

$$V_j(b) = \int\limits_{-\infty}^{\infty} \widehat{\psi}_j(t) \widehat{W}(t/b) \exp\{-ita/b\}\, \mathrm{d}t. \tag{7.16}$$

Then

$$|\, W\,| \geqslant C_0 |\, V_j(b)\,|. \tag{7.17}$$

Here $C_0 > 1/4\pi$.

Proof. Let $N \geqslant 2$ be natural number. By Abel's partial summation formula (1.7) we get

$$\sum_{k=-N}^{N} W\{k\} \psi_j\Big(\frac{k-a}{b}\Big) = \psi_j\Big(\frac{N-a}{b}\Big) \sum_{k=-N}^{N} W\{k\} +$$

$$\sum_{k \leqslant -N} W\{k\} \int\limits_{-N}^{N} \Big\{ \psi_j\Big(\frac{y-a}{b}\Big) \Big\}' \mathrm{d}y - \int\limits_{-N}^{N} \sum_{k \leqslant y} W\{k\} \Big\{ \psi_j\Big(\frac{y-a}{b}\Big) \Big\}' \mathrm{d}y.$$

The limit when $N \to \infty$ gives us the following relation

$$\sum_{k\in\mathbb{Z}} W\{k\}\psi_j\Big(\frac{k-a}{b}\Big) = -\int\limits_{-\infty}^{\infty} \sum_{k\leqslant y} W\{k\}\Big\{\psi_j\Big(\frac{y-a}{b}\Big)\Big\}' \,\mathrm{d}y. \tag{7.18}$$

The quite standard estimation shows that

$$\int\limits_{-\infty}^{\infty} \Big|\Big\{\psi_j\Big(\frac{y-a}{b}\Big)\Big\}'\Big| \,\mathrm{d}y \leqslant 2\sqrt{2\pi}.$$

By (7.7) we get

$$|\,V_j(b)\,| = \sqrt{2\pi}\Big|\sum_{k\in\mathbb{Z}} W\{k\}\psi_j\Big(\frac{k-a}{b}\Big)\,\Big| \leqslant 4\pi|\,W\,|.$$

The last inequality completes the proof. □

Let $\tilde{W}$ denotes the finite measure with the Fourier-Stieltjes transform $\widehat{W}(t/b)\exp\{-\mathrm{i}ta\}$. Then (7.18) is nothing but integration by parts in the following expression:

$$\Big|\int\limits_{-\infty}^{\infty} \psi_j(t)\,\tilde{W}\{\mathrm{d}x\}\Big| = \Big|-\int\limits_{-\infty}^{\infty} \tilde{W}\{(-\infty,x)\}\,\mathrm{d}\psi_j(x)\Big| \leqslant |\,\tilde{W}\,| \int\limits_{-\infty}^{\infty} \Big|\Big\{\psi_j\Big(\frac{x-a}{b}\Big)\Big\}'\Big|\,\mathrm{d}x,$$

It is obvious, that the difference of uniform estimates from the local ones lies in the additional multiplier b. We already established estimate for $V_1(b)$ for the Poisson approximation to the Binomial law. Therefore,without additional calculations, for $0 \leqslant p \leqslant 1/2$ we get

$$\begin{aligned} C_1\min\{np^2,p\} \leqslant \Big|\,\big((1-p)E+p(E_1-E)\big)^n - \exp\{np(E_1-E)\}\,\Big| \leqslant \\ \Big\|\,\big((1-p)E+p(E_1-E)\big)^n - \exp\{np(E_1-E)\}\,\Big\| \leqslant C_2\min\{np^2,p\}. \end{aligned} \tag{7.19}$$

Thus, our upper bound estimates are of the right order.

As the second example we will consider non-degenerate $F \in \mathcal{F}_Z \cap \mathcal{F}_+$ having four finite moments and do not depending on n in any way. We shall prove that

$$|\,F^n - \exp\{n(F-E)\}\,| \geqslant Cn^{-1}. \tag{7.20}$$

In the first subsection we already established the same order of accuracy for the total variation norm, see (7.2). Collecting the estimates of subsection 1 we get

$$|\,\widehat{F}(t) - 1\,| \leqslant \frac{\sigma^2t^2}{2}, \quad \widehat{F}(t) = 1 - \frac{\sigma^2t^2}{2} + \theta Ct^4.$$

and

$$\widehat{F}^n(t) - \exp\{n(\widehat{F}(t)-1)\} = \frac{n(\widehat{F}(t)-1)^2}{2} + \theta C|\,\widehat{F}(t)-1\,|^3 = \frac{n\sigma^4t^4}{8} + \theta n^2t^6.$$

Now it suffices to apply Lemma 7.2 with $V_1(b)$, $b = \sqrt{n}\tilde{C}$ and sufficiently large $\tilde{C}$.

7.4 Exercises

1. Let $0 \leqslant p \leqslant 1/2$. To prove that

$$\Big| ((1-p)E + pE_1)^n - \exp\Big\{np(E_1 - E) - \frac{np^2}{2}(E_1 - E)^2\Big\} \Big| \geqslant C \min\{np^3, p^{3/2}n^{-1/2}\}.$$

2. Let $F \in \mathcal{F}_Z$, $F\{0\} = p_0 \in (0,1)$ and F does not depend on n in any way. To prove that

$$|\,(F-E)\exp\{n(F-E)\}\,|_\infty \geqslant Cn^{-1}.$$

3. Let $F = (1-p)E + pE_1$, $0 < p < 1/2$, $np > 1$. To prove that

$$|\,F^n - F^{n+1}\,| \geqslant c\sqrt{\frac{p}{n}}.$$

Bibliographical notes

Relation between total variation norm and Fourier- Stieltjes transfom was noted by Studnev (196?). Lower bound estimates of (7.17) type can be found in Šiaulys and Čekanavičius (1988). Note that Kruopis (1986a) uses different from ours smoothing measure when estimating lower bounds.

8 The Stein method for Poisson approximation

8.1 The Stein equation

Nowadays Chen's adaptation of the Stein method is among the most popular techniques for estimation of the lattice approximations. H The Stein method is a very powerful technique. Summarizing the present stage of development of the method we can state that:

- The method almost without changes fits for the sum of dependent random variables as well as for the independent ones.
- The method almost without changes fits for total variation, local and Fortet-Mourier distances.
- Application of the method results in small absolute constants.
- The Stein equation can be satisfactorily solved for certain distributions only.
- So far the method is of limited use if we want to benefit from the symmetry of distributions.

Comprehensive treatment of the method can be found in Barbour et al. (1992), Chen (1998), and Barbour and Chryssaphinou (2001). We restrict ourselves to discussion of some moments only.

The Stein - Chen method does not involves characteristic functions and is based on the properties of the special difference equation. In this Section, we need additional notation. For $g : \mathbb{Z}_+ \to \mathbb{R}$, set $\| g \|_\infty = \sup_{j \geq 0} | g(j) |$, $\Delta g(j) = g(j+1) - g(j)$. Symbol f denotes function $f : \mathbb{Z}_+ \to \mathbb{R}$ in all equations, except (8.7).

We begin from the direct observation that Poisson probabilities $\pi(k)$ satisfy the following recursive relation:

$$\pi(k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda} = \frac{k+1}{\lambda}\,\frac{\lambda^{k+1}}{(k+1)!}\,\mathrm{e}^{-\lambda} = \frac{k+1}{\lambda}\,\pi(k+1),$$

or, equivalently,

$$\lambda\pi(k) - (k+1)\pi(k+1) = 0.$$

The last relation allows, for any bounded function g, to write

$$\sum_{k=0}^{\infty}(\lambda g(k+1) - kg(k))\pi(k) = 0,$$

or

$$\mathbb{E}\,(\lambda g(\eta+1) - \eta g(\eta)) = 0. \tag{8.1}$$

Here η has the Poisson distribution with the parameter λ. Now if some random variable ξ is close to η, we can expect that the difference (8.1) for ξ to be small. Let $\mathbb{I}_A(j)$ be indicator function, i.e.

$$\mathbb{I}_A(j) = \begin{cases} 0, & \text{if } j \notin A, \\ 1, & \text{if } j \in A. \end{cases}$$

Let $g(j)$ be solution to the Stein equation:

$$\lambda g(j) - jg(j+1) = \mathbb{I}_A(j) - P(\eta \in A), \quad g(0) = 0, \ j = 0, 1, \ldots \tag{8.2}$$

The essential moment is that properties of g depend on the Poisson distribution only and

$$\| g \|_\infty \leqslant \min\{1, \lambda^{-1/2}\}, \quad \| \Delta g \|_\infty \leqslant (1 - \mathrm{e}^{-\lambda})\lambda^{-1} \leqslant \min\{1, \lambda^{-1}\}. \tag{8.3}$$

The proof of (8.3) can be found in Barbour et al. (1992, p.7). Let ξ be concentrated on $\mathbb{Z}_+$. Then from (8.2) we get that

$$\mathbb{E}\big\{\lambda g(\xi + 1) - \xi g(\xi)\big\} = P(\xi \in A) - P(\eta \in A).$$

If the left-hand side of the last inequality can be estimated through g and Δg, then we automatically obtain estimate in total variation.

Almost without changes the Stein method can be rewritten for estimation of means of bounded functions and other metrics. We begin from the properties of solution to the Stein equation. Set

$$\pi\{f\} = \sum_{k=0}^{\infty} f(k)\pi(k).$$

Lemma 8.1 *Let g be solution to the Stein equation*

$$\lambda g(j+1) - jg(j) = f(j) - \pi\{f\}, \quad j = 0, 1, \ldots \tag{8.4}$$

Then:

a) for bounded $f : \mathbb{Z}_+ \to \mathbb{R}$, g has the following properties

$$\| g \|_\infty \leqslant 2\| f \|_\infty \min\{1, \lambda^{-1/2}\}, \quad \| \Delta g \|_\infty \leqslant 2\| f \|_\infty \min\{1, \lambda^{-1}\}; \tag{8.5}$$

b) for $f : \mathbb{Z}_+ \to \mathbb{R}$, satisfying

$$\| f \|_1 = \sum_{k=0}^{\infty} | f(k) | \leqslant \infty,$$

g has the following properties:

$$\| g \|_1 \leqslant 2\| f \|_1 \lambda^{-1/2}, \quad \| \Delta g \|_1 \leqslant 2\| f \|_1 \lambda^{-1}; \tag{8.6}$$

c) for $f:\mathbb{Z}_+\to\mathbb{R}$, satisfying $\|\Delta f\|_\infty \leqslant 1$, g has the following properties:

$$\|g\|_\infty \leqslant 3, \quad \|\Delta g\|_\infty \leqslant 3\lambda^{-1/2}. \tag{8.7}$$

Remark 8.1 *Note that we presented slightly more rough constants than can be obtained by solving the Stein equation, see Barbour et al. (1992, p.7, p.14). Note also that in case c) we do not assumed f to be bounded.*

All estimates are now obtained, exactly by the same scheme: we estimate the right hand side of

$$\Big|\mathbb{E}f(\xi)-\mathbb{E}f(\eta)\Big| = \Big|\mathbb{E}\big\{\lambda g(\xi+1)-\xi g(\xi)\big\}\Big|. \tag{8.8}$$

We get

a) estimate in total variation by taking, in (8.8), supremum over all $\|f(j)\|_\infty \leqslant 1$ and applying (8.5).

b) local estimate by taking, in (8.8), supremum over all $f(j)=\mathbb{I}_k(j)$ and applying (8.6).

a) estimate for Fortet-Mourier distance by taking, in (8.8), supremum over all $f(j)$ satisfying $\|\Delta f\|_\infty \leqslant 1$ and applying (8.7).

Example: Poisson approximation to the sum of Bernoulli variables

The Stein-Chen method gained its popularity from the Poisson approximation of the sum of Bernoulli random variable. Essentially, the result is an extension of Prohorov's result (5.7) with much sharper constant. Moreover, very little effort is needed to get the estimate. Let $S=\mathbb{I}_1+\mathbb{I}_2+\ldots+\mathbb{I}_n$, where $P(\mathbb{I}_j=1)=p_j=1-P(\mathbb{I}_j=0)$ and all indicators are independent. Let $S^i=S-\mathbb{I}_i$,

$$\lambda=\sum_{k=1}^n p_k.$$

Then

$$\mathbb{E}\mathbb{I}_i g(S)=p_i\mathbb{E}g(S^i+1)$$

and

$$\mathbb{E}\big\{\lambda g(S+1)-Sg(S)\big\}=\sum_{i=1}^n p_i\mathbb{E}\big\{g(S+1)-g(S^i+1)\big\}=$$

$$\sum_{i=1}^n p_i\Big\{\mathbb{E}p_i g(S^i+2)+\mathbb{E}(1-p_i)g(S^i+1)-\mathbb{E}g(S^i+1)\Big\}=\sum_{i=1}^n p_i^2\mathbb{E}\Delta g(S^i+1). \tag{8.9}$$

Let $\mathcal{L}(S)$ denotes the distribution of S and π_λ denotes Poisson distribution with parameter λ. Then, by (8.4) and (8.5) we get that,

$$\Big|\sum_{k=0}^\infty f(k)\,(P(S=k)-\pi(k))\Big| \leqslant \|\Delta g\|_\infty\sum_{i=1}^n p_i^2 \leqslant \frac{2\|f\|_\infty}{\lambda}\sum_{i=1}^n p_i^2$$

and

$$\| \mathcal{L}(S) - \pi_\lambda \| \leqslant \frac{2\sum_1^n p_i^2}{\lambda}. \tag{8.10}$$

Without much additional effort we also get the local estimate. Indeed by definition of the mean we have

$$\Big| \mathbb{E}\Delta g(S^i+1) \Big| \leqslant \sum_{k=0}^{n} | \Delta g(k) | P(S^i = k-1) \leqslant \| \Delta g \|_1 \max_{i,k} P(S^i = k).$$

Set

$$\tilde{\tau} = \lambda - \sum_{k=0}^{n} p_k^2 - \max_i p_i(1-p_i).$$

Then,

$$\max_{i,k} P(S^i = k) \leqslant 0.5\tilde{\tau}^{-1/2}.$$

The proof of the last estimate can be found in Barbour and Jensen (1989). Submitting the last estimates into (8.9) and applying (8.6) we get the local estimate

$$| \mathcal{L}(S) - \pi_\lambda |_\infty \leqslant 2\lambda^{-1}\tilde{\tau}^{-1/2} \sum_1^n p_i^2. \tag{8.11}$$

Even easier to get estimate in the Fortet-Mourier metric. Indeed, from (8.9) and (8.7) we obtain

$$| \mathcal{L}(S) - \pi_\lambda |_{FM} \leqslant 3\lambda^{-1/2} \sum_1^n p_i^2. \tag{8.12}$$

It is easy to check that, for the Binomial distribution the estimate in total variation, local estimate and estimate in the Fortet-Mourier metric are of the order $O(p)$, $O(\sqrt{p/n})$ and $O(p\sqrt{np})$, respectively.

Without much changes the Stein method can be applied to the case of dependent indicators. Let us assume that S is a sum of possibly dependent indicators. Retaining all notation as in above, we then have

$$\begin{aligned} \mathbb{E}\mathbb{I}_k g(S) &= p_k \mathbb{E}\big\{ g(S^k+1) | \mathbb{I}_k = 1 \big\}, \\ \mathbb{E}g(S) &= (1-p_k)\mathbb{E}\big\{ g(S^k+1) | \mathbb{I}_k = 0 \big\} + p_k \mathbb{E}\big\{ g(S^k+2) | \mathbb{I}_k = 1 \big\}. \end{aligned} \tag{8.13}$$

Therefore,

$$\mathbb{E}\big\{ \lambda g(S+1) - Sg(S) \big\} = \mathbb{E}\Big\{ \sum_{k=1}^{n} p_k g(S+1) - \sum_{k=1}^{n} \mathbb{I}_k g(S) \Big\} =$$

$$\sum_{k=1}^{n} p_k \Big((1-p_k)\mathbb{E}\big\{ g(S^k+1) | \mathbb{I}_k = 0 \big\} + p_k \mathbb{E}\big\{ g(S^k+2) | \mathbb{I}_k = 1 \big\} - \mathbb{E}\big\{ g(S^k+1) | \mathbb{I}_k = 1 \big\} \Big) =$$

$$\sum_{k=1}^{n} p_k \Big((1-p_k)\mathbb{E}\big\{ g(S^k+1) | \mathbb{I}_k = 0 \big\} + p_k \mathbb{E}\big\{ \Delta g(S^k+1) | \mathbb{I}_k = 1 \big\} - (1-p_k)\mathbb{E}\big\{ g(S^k+1) | \mathbb{I}_k = 1 \big\} \Big) =$$

$$\sum_{k=1}^{n} p_k^2 \mathbb{E}\{\Delta g(S^k+1)|\mathbb{I}_k=1\}+\sum_{k=1}^{n} p_k(1-p_k)\Big(\mathbb{E}\{g(S^k+1)|\mathbb{I}_k=0\}-\mathbb{E}\{g(S^k+1)|\mathbb{I}_k=1\}\Big).$$

Let η_k and η_k' be auxiliary variables defined on the same probability space and having distributions, corresponding to the conditional distributions of S^k:

$$P(\eta_k=m)=P(S^k=m|\mathbb{I}_k=1), \quad P(\eta_k'=m)=P(S^k=m|\mathbb{I}_k=0).$$

Then

$$\|\mathcal{L}(S)-\pi_\lambda\| \leqslant \|\Delta g\|_\infty \sum_{k=1}^{n} p_k^2 + \|g\|_\infty \sum_{k=1}^{n} p_k(1-p_k)\|\mathcal{L}(\eta_k)-\mathcal{L}(\eta_k')\| \leqslant$$

$$\frac{2}{\lambda}\sum_{k=1}^{n} p_k^2 + \frac{2}{\sqrt{\lambda}}\sum_{k=1}^{n} p_k(1-p_k)\|\mathcal{L}(\eta_k)-\mathcal{L}(\eta_k')\|. \tag{8.14}$$

More frequently, slightly different apperoach is used. Note that, for $0 \leqslant j \leqslant k$,

$$g(k+1)-g(j+1)=\sum_{m=j+1}^{k+1} \Delta g(m)$$

and, for any nonnegative k, j,

$$|g(k+1)-g(j+1)| \leqslant \|\Delta g\|_\infty |j-k|. \tag{8.15}$$

Now, we shall not use expansion (8.13). We have

$$\mathbb{E}\{\lambda g(S+1)-Sg(S)\}=\mathbb{E}\left\{\sum_{i=1}^{n} p_i g(S+1)-\sum_{i=1}^{n}\mathbb{I}_i g(S)\right\}=$$

$$\sum_{i=1}^{n} p_i\Big(\mathbb{E}g(S+1)-\mathbb{E}g(\eta_i+1)\Big)=\sum_{i=1}^{n} p_i\left(\sum_{k=0}^{n} g(k+1)P(S=k)-\sum_{j=0}^{n} g(j+1)P(\eta_i=j)\right)$$

$$\sum_{i=1}^{n} p_i\left(\sum_{k,j=0}^{n} P(S=k)P(\eta_i=j)(g(k+1)-g(j+1))\right).$$

Applying (8.15) we get

$$\|\mathcal{L}(S)-\pi_\lambda\| \leqslant \sum_{i=1}^{n} p_i \sum_{k,j=0}^{n} P(S=k)P(\eta_i=j)|g(k+1)-g(j+1)| \leqslant$$

$$\|\Delta g\|_\infty \sum_{i=1}^{n} p_i \sum_{k,j=0}^{n} P(S=k)P(\eta_i=j)|k-j| \leqslant \frac{2}{\lambda}\sum_{i=1}^{n} p_i \mathbb{E}|S-\eta_i|.$$

8.2 Poisson perturbations

Though the Stein equation can be directly solved for a limited set of distributions only, under certain conditions the properties of solution can be derived indirectly. This is the so called Poisson perturbation case. It was introduced in Barbour and Xia (1999). We shall give the essential idea of Poisson perturbation. Let π denote the (possibly signed) measure with generating function

$$\widehat{\pi}(z) = \sum_{j=0}^{\infty} \pi(j) z^j = \exp\Big\{\sum_{l=1}^{\infty} \lambda_l (z^l - 1)\Big\}, \quad \lambda_l \in \mathbb{R}. \tag{8.16}$$

If all $\lambda_l \geqslant 0$, π is compound Poisson distribution, otherwise π is signed measure. Let

$$\lambda = \sum_{l=1}^{\infty} l\lambda_l.$$

We define a Stein operator $\mathcal{A}$ on $g : \mathbb{Z}_+ \to \mathbb{R}$ in the following way:

$$(\mathcal{A}g)(j) := \sum_{l \geq 1} l\lambda_l g(j+l) - jg(j) = \lambda g(j+1) - jg(j) + (Ug)(j), \tag{8.17}$$

$$(Ug)(j) := \sum_{l \geq 2} l\lambda_l \sum_{k=1}^{l-1} \Delta g(j+k), \quad j = 0, 1, \dots.$$

Note that

$$\pi\{\mathcal{A}g\} = \sum_{j=0}^{\infty} (\mathcal{A}g)(j)\pi(j) = 0 \tag{8.18}$$

for all bounded g. If

$$\lambda > 0, \qquad \upsilon = \lambda^{-1} \sum_{l \geq 2} l(l-1)|\,\lambda_l\,| < 1/2, \tag{8.19}$$

then the solution of the Stein equation, for any bounded f,

$$(\mathcal{A}g)(j) = f(j) - \pi\{f\}, \quad j = 0, 1, \dots \tag{8.20}$$

satisfies inequalities

$$\|\, g \,\|_\infty \leqslant \frac{2}{1-2\upsilon}(1 \wedge \lambda^{-1/2})\|\, f \,\|_\infty, \quad \|\, \Delta g \,\|_\infty \leqslant \frac{2}{1-2\upsilon}(1 \wedge \lambda^{-1})\|\, f \,\|_\infty, \tag{8.21}$$

$$\|\, Ug \,\|_\infty \leqslant \frac{2\upsilon}{1-2\upsilon}\|\, f \,\|_\infty.$$

Let ξ be concentrated on $\mathbb{Z}_+$. Now, when approximating distribution of ξ by π, one should use (8.21) and

$$\Big|\sum_{k=0}^{\infty} f(k)\,(P(\xi = k) - \pi(k))\Big| = \Big|\,\mathbb{E}\,(\mathcal{A}g)(\xi)\,\Big|. \tag{8.22}$$

Example: Signed Poisson approximation to the sum of Bernoulli variables

We shall use perturbation technique for the improved Poisson type approximation. Just like in the first example we assume that $S = \mathbb{I}_1 + \mathbb{I}_2 + \ldots + \mathbb{I}_n$, where $P(\mathbb{I}_j = 1) = p_j = 1 - P(\mathbb{I}_j = 0)$ and all indicators are independent and $S^i = S - \mathbb{I}_i$,

$$\lambda = \sum_{k=1}^{n} p_k.$$

We shall choose approximation π_2 with the moment generating function

$$\widehat{\pi}_2(z) = \exp\Big\{\lambda(z-1) - \frac{1}{2}\sum_{i=1}^{n} p_i^2 (z-1)^2\Big\} = \exp\{\lambda_1(z-1) + \lambda_2(z^2-1)\}.$$

Here

$$\lambda_1 = \sum_{i=1}^{n} p_i(1+p_i), \quad \lambda_2 = -\frac{1}{2}\sum_{i=1}^{n} p^2. \tag{8.23}$$

Let us assume that all $p_i < 1/2$. Then

$$v = \frac{\sum_1^n p_i^2}{\lambda} < \frac{1}{2}.$$

Now, we have

$$(\mathcal{A}g)(j) = \lambda_1 g(j+1) + 2\lambda_2 g(j+2) - jg(j) = \lambda g(j+1) - jg(j) + 2\lambda\Delta g(j+1).$$

Consequently,

$$\mathbb{E}\,(\mathcal{A}g)(S) = \sum_{i=1}^{n} p_i \Big\{\mathbb{E}g(S+1) - \mathbb{E}g(S^i+1) - p_i \mathbb{E}\,\Delta g(S+1)\Big\} =$$

$$\sum_{i=1}^{n} p_i \Big\{p_i \mathbb{E}g(S^i+2) + (1-p_i)\mathbb{E}g(S^i+1) - \mathbb{E}g(S^i+1) - p_i\mathbb{E}\,\Delta g(S+1)\Big\} =$$

$$\sum_{i=1}^{n} p_i^2 \Big\{\mathbb{E}\,\Delta g(S^i+1) - \mathbb{E}\,\Delta g(S+1)\Big\} =$$

$$\sum_{i=1}^{n} p_i^2 \Big\{\mathbb{E}\,\Delta g(S^i+1) - p_i\mathbb{E}\,\Delta g(S^i+2) - (1-p_i)\mathbb{E}\,\Delta g(S^i+1)\Big\} = -\sum_{i=1}^{n} p_i^3 \mathbb{E}\,\Delta^2 g(S^i+1).$$

Thus, by (8.22)

$$\Big|\sum_{k=0}^{\infty} f(k)\,(P(\xi = k) - \pi_2(k))\Big| \leqslant \Big|\sum_{i=1}^{n} p_i^3 \mathbb{E}\,\Delta^2 g(S^i+1)\Big|. \tag{8.24}$$

However, we do not have appropriate estimate for the second difference. The problem is solved indirectly. Taking $g(j) = 0$, for $j \leqslant 0$, we obtain

$$\Big|\mathbb{E}\,\Delta^2 g(S^j+1)\Big| = \Big|\mathbb{E}\Big(g(S^j+2) - 2g(S^j+1) + g(S^j)\Big)\Big| =$$

$$\Big| \sum_{k=-2}^{\infty} P\{S^j = k\}(g(k+2) - g(k+1)) - \sum_{k=-2}^{n} P(S^j = k)(g(k+1) - g(k)) \Big| =$$

$$\Big| \sum_{k=-2}^{n} (g(k+2) - g(k+1))\Big(P(S^j = k) - P(S^j = k+1)\Big) \Big| \leqslant$$

$$\| \Delta g \|_\infty \sum_{k=-1}^{n} \Big| P(S^j = k) - P(S^j = k+1) \Big|. \tag{8.25}$$

Now we shall make use of the fact that S is unimodal. Consequently, we have

$$\sum_{k=-1}^{n} \Big| P(S^j = k) - P(S^j = k+1) \Big| \leqslant 2 \max_k P(S^j = k). \tag{8.26}$$

Collecting all estimates we obtain

$$\Big| \sum_{k=0}^{\infty} f(k)\,(P(\xi = k) - \pi_2(k)) \Big| \leqslant 2\lambda^{-1}(\tilde{\tau})^{-1/2} \| f \|_\infty \sum_{i=1}^{n} p_i^3. \tag{8.27}$$

We recall that $\tilde{\tau} = \sum_1^n p_k(1-p_k) - \max_i p_i(1-p_i)$.

Note that (8.27) is of the order $O(n^{-1/2})$ even if all $p_i = const$.

8.3 Estimating the first pseudomoment

There are various methods for estimating nonuniform estimates when applying Stein's method. Usually they heavily depend on the properties of the Poisson distribution. Here we present one quite direct approach, which allows to replace $f(j)$ by $f(j)(j-\lambda)$. Note that this new function is unbounded as $j \to \infty$. Further in this Section we assume that

a) π is the (possibly signed) measure defined by (8.16) and Poisson perturbation assumption (8.19) holds;

b) random variable ξ is concentrated on $\mathbb{Z}_+$, has at least two finite moments and $\mathbb{E}\xi = \lambda$;

For the sake of brevity, we denote $P(\xi = j) = P(j)$.

Lemma 8.2 *Let conditions a)- b) be satisfied and let, for any bounded $g : \mathbb{Z}_+ \to \mathbb{R}$,*

$$|\,\mathbf{E}(\mathcal{A}g)(\xi)\,| \leqslant \varepsilon_0 \| g \|_\infty. \tag{8.28}$$

Then, for any bounded $f : \mathbb{Z}_+ \to \mathbb{R}$,

$$\Big| \sum_{j=0}^{\infty} f(j)(j-\lambda)(P(j) - \pi(j)) \Big| \leqslant 14(1-2v)^{-2}\varepsilon_0 \| f \|_\infty. \tag{8.29}$$

Remark 8.2 *We always can choose $f(j)$ to be such that the sum in Lemma becomes*

$$\sum_{j=0}^{\infty} |\, j-\lambda \,|\,|\, P(j)-\pi(j) \,|.$$

This expression is called the first absolute pseudomoment of $\mathcal{L}(\xi)$ and π.

Proof. Set

$$h_1(j) = (j-\lambda)g(j), \quad \varphi_1(j) = (j-\lambda)(\mathcal{A}g)(j) - (\mathcal{A}h_1)(j).$$

Let us multiply (8.20)) by $(j-\lambda)P(j)$ and sum over all j. Then multiply (8.20) by $(j-\lambda)\pi(j)$ and sum over j. We get

$$\sum_{j=0}^{\infty}(\mathcal{A}g)(j)(j-\lambda)P(j) = \sum_{j=0}^{\infty} f(j)(j-\lambda)P(j),$$

$$\sum_{j=0}^{\infty}(\mathcal{A}g)(j)(j-\lambda)\pi(j) = \sum_{j=0}^{\infty} f(j)(j-\lambda)\pi(j).$$

Note that we employed the fact that $\mathbb{E}\xi = \lambda$. We have

$$\sum_{j=0}^{\infty}(\mathcal{A}h_1)(j)\pi(j) = 0.$$

Therefore,

$$\Big|\sum_{j=0}^{\infty} f(j)(j-\lambda)(P(j)-\pi(j))\Big| \leqslant |\,\mathbb{E}(\mathcal{A}h_1)(\xi)\,| + \Big|\sum_{j=0}^{\infty} \varphi_1(j)(P(j)-\pi(j))\Big|. \tag{8.30}$$

We shall prove that $h_1(j)$ is bounded. Indeed, let us take the Stein equation (8.20)

$$\lambda g(j+1) - jg(j) + (Ug)(j) = f(j) - \pi\{f\}$$

and rewrite it in the form

$$\lambda\Delta g(j) - (\lambda - j)g(j) + (Ug)(j) = f(j) - \pi\{f\}.$$

Consequently,

$$|\, h_1(j) \,| = |\, \lambda - j \,|\,|\, g(j) \,| \leqslant |\, f(j) - \pi\{f\} \,| + |\,(Ug)(j)\,| + \lambda|\,\Delta g(j)\,|.$$

Taking into account (8.21) we get

$$|\, f(j) - \pi\{f\} \,| \leqslant 2\|\, f \,\|_\infty, \quad |\,(Ug)(j)\,| \leqslant \frac{2v}{1-2v}\|\, f \,\|_\infty,$$

$$\lambda|\,\Delta g(j)\,| \leqslant \frac{2}{1-2v}\|\, f \,\|_\infty, \quad \|\, h_1(j) \,\|_\infty \leqslant \frac{2(2-v)}{1-2v}\,\|\, f \,\|_\infty.$$

Thus, by assumption (8.28)

$$|\,\mathbb{E}(\mathcal{A}h_1)(\xi)\,| \leqslant \varepsilon_0\,\frac{2(2-v)}{1-2v}\,\|\,f\,\|_\infty. \tag{8.31}$$

Now, we shall prove that $\varphi_1(j)$ is bounded. Indeed, from the Stein equation we have

$$|\,\varphi_1(j)\,| = |\,(j-\lambda)(\mathcal{A}g)(j) - (\mathcal{A}h_1)(j)\,| =$$

$$\Big|\sum_{l=1}^{\infty} l\lambda_l g(j+l)(j+l-\lambda) - \sum_{l=1}^{\infty} l\lambda_l g(j+l)(j-\lambda)\Big| = \Big|\sum_{l=1}^{\infty} l^2\lambda_l g(j+l)\Big| \leqslant$$

$$\|\,g\,\|_\infty \sum_{l=1}^{\infty} l^2|\,\lambda_l\,| \leqslant \frac{2}{(1-2v)\sqrt{\lambda}}\|\,f\,\|_\infty \sum_{l=1}^{\infty} l^2|\,\lambda_l\,|. \tag{8.32}$$

First, note that $\lambda_1 > 0$. Indeed, $l^2 \leqslant 2l(l-1)$, $l > 1$ and $v < 1/2$. Therefore,

$$\sum_{l=2} l(l-1)|\,\lambda_l\,| \leqslant \frac{\lambda}{2}$$

and

$$2\sum_{l=2} l(l-1)|\,\lambda_l\,| \leqslant \lambda = \lambda_1 + \sum_{l=2}^{\infty} l\lambda_l \leqslant \lambda_1 + \sum_{l=2}^{\infty} l(l-1)|\,\lambda_l\,|. \tag{8.33}$$

Now, we prove that $2\lambda_1 \leqslant 3\lambda$. Let

$$\sum_{l=2}^{\infty} l\lambda_l \geqslant 0.$$

Then, from the definition of λ and its non-negativeness we get $\lambda_1 \leqslant \lambda$. Let

$$\sum_{l=2}^{\infty} l\lambda_l < 0.$$

Then by elementary estimates and (8.33) we get

$$2\Big|\sum_{l=2}^{\infty} l\lambda_l\Big| \leqslant 2\sum_{l=2}^{\infty} l|\,\lambda_l\,| \leqslant 2\sum_{l=2}^{\infty} l(l-1)|\,\lambda_l\,| \leqslant \lambda = \lambda_1 - \Big|\sum_{l=2}^{\infty} l\lambda_l\Big|.$$

Consequently,

$$\Big|\sum_{l=2}^{\infty} l\lambda_l\Big| \leqslant \lambda_1/3$$

and

$$\lambda_1 = \lambda + \Big|\sum_{l=2}^{\infty} l\lambda_l\Big| \leqslant \lambda + \lambda_1/3.$$

Thus, in any case, $0 < \lambda_1 \leqslant 3\lambda/2$. Finally,

$$\sum_{l=1}^{\infty} l^2|\,\lambda_l\,| \leqslant \frac{3\lambda}{2} + 2\sum_{l=2}^{\infty} l(l-1)|\,\lambda_l\,| \leqslant \frac{3\lambda}{2} + \lambda = \frac{5\lambda}{2}.$$

Substituting these estimates into (8.32) we obtain,

$$\| \varphi_1 \| \leqslant \frac{5\lambda}{(1-2v)\sqrt{\lambda}} \| f \|_\infty. \tag{8.34}$$

Now we can directly estimate

$$\Big| \sum_{j=0}^{\infty} \varphi_1(j)(P(j) - \pi(j)) \Big| \leqslant \| \varphi_1 \|_\infty \sum_{j=0}^{\infty} | P(j) - \pi(j) | = \| \mathcal{L}(\xi) - \pi \| \| \varphi_1 \|_\infty.$$

From (8.22) we see that choosing $f(j)$ to be the sign function

$$\sum_{j=0}^{\infty} | P(j) - \pi(j) | \leqslant \Big| \mathbb{E}(\mathcal{A}g)(\xi) \Big| \leqslant \varepsilon_0 \| g \|_\infty \leqslant \varepsilon_0 \frac{2}{(1-2v)\sqrt{\lambda}}.$$

Collecting all estimates we obtain

$$| \mathbb{E}(\mathcal{A}h_1)(\xi) | + \Big| \sum_{j=0}^{\infty} \varphi_1(j)(P(j) - \pi(j)) \Big| \leqslant$$

$$\varepsilon_0 \frac{2(2-v)}{1-2v} \| f \|_\infty + \varepsilon_0 \frac{10}{(1-2v)^2} \| f \|_\infty \leqslant \varepsilon_0 \frac{14}{(1-2v)^2} \| f \|_\infty.$$

Submitting the last estimate into (8.30) we get the statement of Lemma. □

Example: Pseudomoment for the signed Poisson approximation

As in the second example we assume that $S = \mathbb{I}_1 + \mathbb{I}_2 + \ldots + \mathbb{I}_n$, where $P(\mathbb{I}_j = 1) = p_j = 1 - P(\mathbb{I}_j = 0)$ and all indicators are independent and $S^i = S - \mathbb{I}_i$,

$$\lambda = \sum_{k=1}^{n} p_k.$$

and π_2 has the moment generating function

$$\widehat{\pi}_2(z) = \exp\Big\{ \lambda(z-1) - \frac{1}{2} \sum_{i=1}^{n} p_i^2 (z-1)^2 \Big\} = \exp\{ \lambda_1(z-1) + \lambda_2(z^2-1) \}$$

with

$$\lambda_1 = \sum_{i=1}^{n} p_i(1+p_i), \quad \lambda_2 = -\frac{1}{2} \sum_{i=1}^{n} p^2.$$

We assume that all $p_i < 1/2$. Then we already proved that,

$$\Big| \mathbb{E}(\mathcal{A}g)(S) \Big| = \Big| \sum_{i=1}^{n} p_i^3 \mathbb{E} \Delta^2 g(S^i + 1) \Big|.$$

Exactly as in the previous example we establish that

$$\Big|\mathbb{E}\Delta^2 g(S^i+1)\Big| \leqslant \| g \|_\infty \sum_{k=-2}^{n} \Big| P(S^j = k+2) - 2P(S^j = k+1) + P(S^j = k) \Big| =$$

$$\| g \|_\infty \| \mathcal{L}(S^j)(E_1 - E)^2 \|.$$

For the last estimate we shall apply (5.16) obtaining

$$\| \mathcal{L}(S^j)(E_1 - E)^2 \| \leqslant 8\Big(\sum_{i=1}^{n} p_i - \max_i p_i \Big)^{-1}.$$

Therefore,

$$\Big| \mathbb{E}\,(\mathcal{A}g)(S) \Big| \leqslant 8 \sum_{k=1}^{n} p_k^3 \Big(\sum_{i=1}^{n} p_i - \max_i p_i \Big)^{-1} \| g \|_\infty.$$

Collecting all estimates, and applying Lemma 8.2we finally obtain

$$\Big| \sum_{k=0}^{\infty} f(k)(k-\lambda)(P(S=k) - \pi_2(k)) \Big| \leqslant \| f \|_\infty \frac{112}{(1-2\upsilon)^2} \sum_{i=1}^{n} p_i^3 \Big(\sum_{i=1}^{n} p_i - \max_i p_i \Big)^{-1}. \quad (8.35)$$

For the Binomial distribution (i.e. when all $p_i = p < 1/2$) the order of accuracy in (8.35) is $O(p^2)$.

Exercises

1. Let $S = \mathbb{I}_1 + \mathbb{I}_2 + \ldots + \mathbb{I}_n$, where $P(\mathbb{I}_j = 1) = 1 - P(\mathbb{I}_j = 0) = p_j \leqslant 1/4$ and all indicators are independent and let and π_3 has the moment generating function

$$\widehat{\pi}_2(z) = \exp\Big\{ \sum_{i=1}^{n} \Big(p_i(z-1) - \frac{np_i^2}{2}(z-1)^2 + \frac{np_i^3}{3}(z-1)^3 \Big) \Big\}.$$

   To estimate the accuracy of approximation in total variation, local and Fortet-Mourier distances.

2. To investigate approximation of S from previous exercise by the Binomial distribution applying perturbation technique.

Bibliographical notes

Chen's adaptation of the Stein method appeared in Chen (1975). Perturbation technique was introduced in Barbour and Xia (1999). For other applications of the Stein method see Barbour and Jensen (1989), Barbour et al. (1992), Barbour and Chryssaphinou (2001), Brown and Xia (2001), Barbour and Čekanavičius (2002), and references therein.

9 Non-lattice discrete measures

In this Section we shall discuss one method for estimation of discrete non-lattice case. What do we mean by lattice? Quite generally speaking W is lattice if, for some $h > 0$, it is concentrated on a set $K = \{a + hk,\ k \in \mathbb{Z}\}$. For convenience, it is usually assumed that h is the maximal common divisor of the points. For example, if W is concentrated at 0, 2 and 4; then $h = 2$. If W is concentrated at 0,1,3 then $h = 1$. It is obvious, that suitably shifted and normed by h, any lattice measure can be reduced to the case of measures concentrated on $\mathbb{Z}$. Then, the methods of previous sections apply. Now let us consider W concentrated at 0, 1 and $\sqrt{2}$. For such a measure we need special methods of estimation. It must be noted that discrete non-lattice distributions (measures) are rarely investigated and there are a few suitable methods only. On the other hand, they are usually covered by the general estimates obtained for compound distributions. Indeed, any discrete distribution allows representation

$$(1-p)E + pB, \qquad B \in \mathcal{F}$$

and results for the total variation from Section 2 apply.

In this Section we use Arak's inequality. Let $\mathbf{u} = (u_1, u_2, \ldots, u_N) \in \mathbb{R}^N$. Set

$$K_m(\mathbf{u}) = \left\{ \sum_{i=1}^{N} j_i u_i : j_i \in \{-m, -m+1, \ldots, m\};\ i = 1, \ldots, N \right\}. \tag{9.1}$$

We recall that $W = W^+ - W^-$ is the Jordan–Hahn decomposition of W. Set

$$\delta(W, m, \mathbf{u}) = W^+\{\mathbb{R} \setminus K_m(\mathbf{u})\} + W^-\{\mathbb{R} \setminus K_m(\mathbf{u})\}. \tag{9.2}$$

Lemma 9.1 *Let $W \in \mathcal{M}$, $W\{\mathbb{R}\} = 0$, $N, m \in \mathbb{N}$, $\mathbf{u} \in \mathbb{R}^{N+1}$, $h > 0$ and $U \in \mathcal{F}_+$. Then*

$$|\,W\,| \leqslant C_4 \int\limits_{|\,t\,| \leqslant 1/h} \left| \frac{\widehat{W}(t)}{t} \right| \mathrm{d}t +$$

$$C\big(1 + N^2 \ln(Nm+1)\big) \sup_{t \in \mathbb{R}} \frac{|\,\widehat{W}(t)\,|}{\widehat{U}(t)}\, Q(U, h) + \delta(W, m, \mathbf{u}). \tag{9.3}$$

The proof of (9.3) is beyond the scopes of this course. We can see that unlike formulas of inversion for the lattice case, (9.3) has three different parts. The first one is quite similar to the Tsaregradskii inequality and deals with the behavior of Fourier-Stieltjes transform in the neighborhood of zero. The second part directly replaces the estimate of measure by the estimate of its Fourier -Stieltjes transform. However, with some additional factor $U \in \mathcal{F}_+$. How this factor U affects the estimate? In principle, it (or rather its concentration function) should neutralize the logarithm factor.The third summand states some limits on applicability of (9.3). Indeed, the support of W, should not be much different from $K_m(\mathbf{u})$.

What can be said about the structure of $K_m(\mathbf{u})$? Probably, the idea can be best reflected by example, such as $\{1, \sqrt{2}, 2\sqrt{2}, \sqrt{3}, \sqrt{2} - \sqrt{3}, \sqrt{2} + 5\sqrt{3}\}$.

Obviously, if $supp\, W_i \subset K_m(\mathbf{u})$, $(i = 1, \dots, n)$ then

$$supp\, W_1 W_2 \cdots W_n \subset K_{nm}(\mathbf{u}). \tag{9.4}$$

For an effective application of (9.3) we must have estimates of $\delta(W, m, \mathbf{u})$. As it turns out, the support of any signed compound Poisson measure is close to some subset of $K_{my}(\mathbf{u})$ provided its compounding measure is concentrated on $K_m(\mathbf{u})$. We shall formulate this result more precisely.

Let $suppW$ denote the support of W.

Lemma 9.2 *Let $W, V \in \mathcal{M}$, $\| W \| \leqslant b_1$, $\| V \| \leqslant b_2$, $N, m \in \mathbb{N}$, $\mathbf{u} \in \mathbb{R}^{N+1}$, $supp\, W \subset K_s(\mathbf{u})$, $supp\, V \subset K_m(\mathbf{u})$. Then, for any $y \in \mathbb{N}$, the following inequalities hold*

$$\delta(W \exp\{V\}, s + my, \mathbf{u}) \leqslant b_1 \exp\{3b_2 - y\}, \tag{9.5}$$

$$\delta(\exp\{V\}, my, \mathbf{u}) \leqslant \exp\{3b_2 - y\}. \tag{9.6}$$

Proof From (9.4) we get that

$$\sum_{k=0}^{y} \frac{V^k\{\mathbb{R} \setminus K_{my}(\mathbf{u})\}}{k!} = 0. \tag{9.7}$$

Therefore,

$$\delta(\exp\{V\}, my, \mathbf{u}) \leqslant \sum_{k>y}^{\infty} \frac{\| V^k \|}{k!} \leqslant \sum_{k>y}^{\infty} \frac{\| V \|^k}{k!} \leqslant \sum_{k>y}^{\infty} \frac{b_2^k}{k!} \leqslant \mathrm{e}^{-y} \sum_{k=0}^{\infty} \frac{(\mathrm{e}b_2)^k}{k!}. \tag{9.8}$$

Similarly

$$\delta(W \exp\{V\}, s + my, \mathbf{u}) \leqslant \sum_{k>y} \frac{b_1 b_2^k}{k!} \leqslant b_1 \exp\{\mathrm{e}b_2 - y\}.$$

□

We shall demonstrate how Arak's Lemma works on a quite simple example. Let $F \in \mathcal{F}_+$ do not depend on n and be concentrated at $0, x_1, x_2, \dots, x_N$, let $F\{0\} > 0$ and all $F\{x_j\} > 0$. Then we shall prove that

$$| F^n - \exp\{n(F - E)\} | = O(n^{-1}). \tag{9.9}$$

As we'll see in subsequent sections, (9.9) can be significantly generalized by other, more elaborated, method.

The most important benefit from non-negativeness of $\widehat{F}(t)$ is the following direct estimate:

$$\widehat{F}(t) = 1 + (\widehat{F}(t) - 1) \leqslant \exp\{\widehat{F}(t) - 1\}.$$

We shall apply Arak's inequality. Let us consider all its parts. The difference between characteristic functions can be estimated directly

$$
|\,\widehat{F}(t)^n - \exp\{n(\widehat{F}(t)-1)\}\,| \leqslant n\exp\{(n-1)(\widehat{F}(t)-1)\}|\,\widehat{F}(t) - \exp\{\widehat{F}(t)-1\}\,| \leqslant
$$

$$
5n\exp\{n(\widehat{F}(t)-1)\}|\,\widehat{F}(t)-1\,|^2 \leqslant C\frac{1}{n}\exp\Big\{\frac{n}{2}(\widehat{F}(t)-1)\Big\}. \tag{9.10}
$$

Therefore, taking

$$
\widehat{U}(t) = \exp\Big\{\frac{n}{2}(\widehat{F}(t)-1)\Big\}
$$

we obtain that the second part of the estimate is majorized by

$$
C\frac{1}{n}\big(1+N^2\ln(Nm+1)\big)Q(U,h) \tag{9.11}
$$

with N, m and h still undefined. For the choice of h, let us turn to the behavior of $\widehat{F}(t)$ in the neighborhood of zero. From the expansion in moments we have

$$
|\,\widehat{F}(t)-1\,| \leqslant \frac{\sigma^2t^2}{2}, \quad \widehat{F}(t)-1 = -\frac{\sigma^2t^2}{2} + \theta a t^4. \tag{9.12}
$$

Here σ^2 is the second moment of F (i.e. the variance of corresponding random variable) and a can be expressed through fourth absolute moment. We assumed that $F \neq E$ and does not depend on n. Therefore, σ^2 and a are some absolute positive constants. Now let

$$
t^2 \leqslant \frac{\sigma^2}{4a}.
$$

Then

$$
\widehat{F}(t) - 1 \leqslant -\frac{\sigma^2t^2}{4}.
$$

Therefore, let us take

$$
h = \frac{2\sqrt{a}}{\sigma}. \tag{9.13}
$$

Then, for $|\,t\,| \leqslant 1/h$ we have $t^2 \leqslant \sigma^2/(4a)$ and

$$
n\exp\{n(\widehat{F}(t)-1)\}|\,\widehat{F}(t)-1\,|^2 \leqslant Cn\sigma^4t^4\exp\Big\{-\frac{n\sigma^2t^2}{4}\Big\}.
$$

From the last estimate we see that the first part of Arak's inequality can be estimates by

$$
\int\limits_{|\,t\,|\leqslant 1/h} \left|\frac{|\,\widehat{F}(t)^n - \exp\{n(\widehat{F}(t)-1)\}\,|}{t}\right| \mathrm{d}t \leqslant Cn\int\limits_{-\infty}^{\infty} \sigma^4|\,t\,|^3\exp\Big\{-\frac{n\sigma^2t^2}{4}\Big\}\,\mathrm{d}t \leqslant Cn^{-1}.
$$

Now let us consider the δ part. Instead of estimating the Jordan-Hahn decomposition, we shall estimate δ for F^n and for $\exp\{n(F-E)\}$, separately. Let us apply Lemma 9.2 for $V = n(F-E)$. We have

$$
\|\,V\,\| = n\|\,F-E\,\| \leqslant 2n =: b_2.
$$

Moreover, set $\mathbf{u} = (0, x_1, x_2, \dots, x_N)$, $m = 1$. Then

$$suppV \subset K_1(\mathbf{u}).$$

Therefore, by (9.6) we get that

$$\delta(\exp\{V\}, 7n, \mathbf{u}) \leqslant \mathrm{e}^{-n}.$$

Moreover,

$$\delta(F^n, 7n, \mathbf{u}) = 0.$$

Thus, we proved that the third summand in Arak's inequality is vanishing exponentially, for $m = 7n$. Once again considering the middle part of Arak's inequality we see that it remains to estimate

$$\big(1 + N^2 \ln(Nm + 1)\big) Q(U, h) \leqslant C(N) \ln n Q(U, h).$$

Recalling the fact that h does not depend on n we, can apply the properties of concentration function (1.28) and (1.30):

$$Q(U, h) \leqslant CQ(U, 0.5) \leqslant \frac{C}{\sqrt{nF\{\{x : |\,x\,| > 0\}\}}} \leqslant \frac{C}{\sqrt{n}}.$$

Collecting all estimates we finally get

$$|\,F^n - \exp\{n(F - E)\}\,| \leqslant \frac{C}{n}\left(1 + \frac{\ln n}{\sqrt{n}}\right) + C\mathrm{e}^{-n}. \tag{9.14}$$

It is obvious that (9.14) is even sharper than (9.10).

9.1 Exercises

1. Let, for some fixed x,

$$F = q_0 E + q_1 E_x + q_2 E_{-x} + q_3 E_1.$$

To estimate $|\,F^n - \exp\{n(F - E)\}\,|$ taking into account possible symmetry.

2. Let F be symmetric distribution, concentrated on the set $\{0, x_1, \dots, x_N\}$. To prove that for any natural k

$$|\,(F - E)^k \exp\{n(F - E)\}\,| \leqslant C(k) n^{-k}.$$

3. Let F be defined as before, and let $F \in \mathcal{F}_+$. To prove that

$$|\,F^n - F^{n+1}\,| \leqslant Cn^{-1}.$$

Bibliographical notes

Arak's inequality was introduced by Arak (1981). It was also applied by Čekanavičius and Wang (2003).

10 Esseen's formula of inversion

We begin investigation of the general methods. One of the most popular methods for estimating uniform distance is Esseeen's formula of inversion. Here we present one of the most general versions of this theorem.

Lemma 10.1 *Let $F(x)$ be a non-decreasing bounded function, and $G(x)$ a function of bounded variation; $F(-\infty) = G(-\infty)$. Let T be an arbitrary positive number. Then for every $b > 1/(2\pi)$ we have*

$$|\,F - G\,| \leqslant b \int\limits_{-T}^{T} \left| \frac{\widehat{F}(t) - \widehat{G}(t)}{t} \right| \mathrm{d}t + bT \sup_x \int\limits_{|y| \leqslant c(b)/T} |\,G(x+y) - G(x)\,| \, \mathrm{d}y, \tag{10.1}$$

where $c(b)$ is a positive constant depending only on b.

We recall that distribution function $F(x)$ and distribution F (measure) are related by $F(x) = F\{(-\infty, x)\}$, The last summand allows various different estimates from above. For example, it allows estimates through concentration functions, see (11.2). Sometimes the following general expression can be used: If a and b are points of continuity of the distribution function $F(x)$ then

$$F(b) - F(a) = \frac{1}{2\pi} \lim_{T\to\infty} \int\limits_{-T}^{T} \frac{\mathrm{e}^{-itb} - \mathrm{e}^{-ita}}{-it} \, \mathrm{d}t. \tag{10.2}$$

We present another version of the Esseen type inequality due to Bentkus and Götze.

Lemma 10.2 *For $F, G \in \mathcal{F}$ and $T > 0$ the follwing inequality holds:*

$$|\,F - G\,| \leqslant \frac{1}{2\pi} \int\limits_{-T}^{T} \frac{|\,\widehat{F}(t) - \widehat{G}(t)\,|}{|\,t\,|} \mathrm{d}t + \frac{1}{T} \int\limits_{-T}^{T} |\,\widehat{F}(t)\,| \, \mathrm{d}t + \frac{1}{T} \int\limits_{-T}^{T} |\,\widehat{G}(t)\,| \, \mathrm{d}t. \tag{10.3}$$

Lemma 10.1 can be simplified under additional assumptions. One of the most popular its versions was developed for the case when G has bounded density. Note that, in general, we do not assume $F(x)$ to be continuous.

Lemma 10.3 *Let $F, G \in \mathcal{F}$ and let $\sup_x G'(x) \leqslant K$. Then for every $T > 0$ and every $b > 1/(2\pi)$ we have*

$$|\,F - G\,| \leqslant b \int\limits_{-T}^{T} \left| \frac{\widehat{F}(t) - \widehat{G}(t)}{t} \right| \mathrm{d}t + C(b) \frac{K}{T}, \tag{10.4}$$

where $c(b)$ is a positive constant depending only on b.

Now we can consider some examples of application of Esseen's smoothing lemma.

10.1 The Berry-Esseen theorem

Let $F \in \mathcal{F}$ and

$$\int_{-\infty}^{\infty} x\, F\{\mathrm{d}x\} = 0, \qquad \int_{-\infty}^{\infty} x^2\, F\{\mathrm{d}x\} = \sigma^2; \qquad \int_{-\infty}^{\infty} |\, x\,|^3\, F\{\mathrm{d}x\} = \beta_3 < \infty.$$

We shall denote by Φ_σ the normal distribution with the variance σ^2, i.e. having the characteristic function

$$\widehat{\Phi}_\sigma(t) = \exp\Big\{-\frac{1}{2}\sigma^2 t^2\Big\}.$$

Theorem 10.1 *For any n the following inequality holds*

$$|\, F^n - \Phi^n \,| \leqslant C\frac{\beta_3}{\sqrt{n}\sigma^3}. \tag{10.5}$$

Proof. We shall use (10.4) with $T = \sigma^2/(4\beta_3)$. Note that $\Phi_\sigma^n = \Phi_{\sqrt{n}\sigma}$ and its density equals to

$$\phi(x) = \frac{1}{\sqrt{2\pi n}\sigma}\exp\{-x^2/(n\sigma^2)\}$$

and is bounded by $(\sqrt{2\pi n}\sigma)^{-1}$. Therefore, the second summand in (10.4) is bounded by $C\beta_3 n^{-1/2}\sigma^{-3}$.

We shall expand the characteristic function $|\,\widehat{F}(t)\,|^2$ in powers of (it). Note that $|\,\widehat{F}(t)\,|^2$ corresponds to the difference of two independent random variables $(\xi - \tilde{\xi})$, both having distribution F. Thus,

$$|\,\widehat{F}(t)\,|^2 \leqslant 1 - \sigma^2 t^2 + \frac{4}{3}\beta_3 |\,t\,|^3. \tag{10.6}$$

Consequently, for $|\,t\,| \leqslant \sigma^2/(4\beta_3)$,

$$|\,\widehat{F}(t)\,| \leqslant \exp\Big\{1 - \frac{1}{2}\sigma^2 t^2 + \frac{2}{3}\beta_3|\,t\,|^3\Big\} \leqslant \exp\Big\{-\frac{1}{3}\sigma^2 t^2\Big\}. \tag{10.7}$$

Let $|\,t\,| \leqslant \sigma^2/(4\beta_3)$ and $n\beta_3|\,t\,|^3 > 1$. Then

$$|\,\widehat{F}^n(t) - \widehat{\Phi}_\sigma^n(t)\,| \leqslant 2n\exp\Big\{-\frac{n}{3}\,\sigma^2 t^2\Big\} \leqslant 2n\beta_3|\,t\,|^3 \exp\Big\{-\frac{n}{3}\,\sigma^2 t^2\Big\}.$$

If $|\,t\,| \leqslant \sigma^2/(4\beta_3)$ and $n\beta_3|\,t\,|^3 \leqslant 1$, then we directly estimate

$$|\,\widehat{F}(t) - \widehat{\Phi}_\sigma(t)\,| \leqslant C(\beta_3|\,t\,|^3 + \sigma^4 t^4).$$

But by relation between moments

$$\sigma \leqslant \beta_3^{1/3}$$

and, for $n\beta_3|\,t\,|^3 \leqslant 1$,

$$\sigma^2 t^2 \leqslant (\beta_3|\,t\,|^3)^{2/3} \leqslant 1.$$

Consequently,

$$|\widehat{F}^n(t)-\widehat{\Phi}_\sigma^n(t)|\leqslant n\exp\Big\{-\frac{(n-1)}{3}\sigma^2t^2\Big\}|\widehat{F}(t)-\widehat{\Phi}_\sigma(t)|\leqslant$$
$$Cn\beta_3|t|^3\exp\Big\{-\frac{n}{3}\sigma^2t^2\Big\}.$$

Thus we prove that, for $|t|\leqslant\sigma^2/(4\beta_3)$,

$$|\widehat{F}^n(t)-\widehat{\Phi}_\sigma^n(t)|\leqslant Cn\beta_3|t|^3\exp\Big\{-\frac{n}{3}\sigma^2t^2\Big\}. \tag{10.8}$$

To end the proof of theorem it suffices to use the last estimate in integral:

$$\int_{-T}^{T}\left|\frac{\widehat{F}^n(t)-\widehat{\Phi}_\sigma^n(t)}{t}\right|\mathrm{d}t\leqslant C\int_0^\infty n\beta_3t^2\exp\Big\{-\frac{n}{3}\sigma^2t^2\Big\}\,\mathrm{d}t\leqslant C\frac{\beta_3}{\sqrt{n}\sigma^3}. \tag{10.9}$$

10.2 Subsequent convolutions

We shall assume that F does not depend on n, has mean zero and $1+\delta$ absolute moment $(0<\delta\leqslant 1)$ and satisfies Cramer's (C) condition:

$$\limsup_{|t|\to\infty}|\widehat{F}(t)|<1\,. \tag{C}$$

First we give some general facts about (C) condition. Any distribution with the continuous component satisfies (C) condition. Moreover, for any $\varepsilon=\varepsilon(F)>0$ and for all $|t|>\varepsilon$,

$$|\widehat{F}(t)|\leqslant e^{-C(F)}. \tag{10.10}$$

We combine this property with the fact that any nondegenerate F, for small $\varepsilon=\varepsilon(F)$ and for $|t|\leqslant\varepsilon$,

$$|\widehat{F}(t)|\leqslant\exp\{-C(F)t^2\}. \tag{10.11}$$

We shall prove that

$$|F^n-F^{n+1}|\leqslant C(F)n^{-\delta}. \tag{10.12}$$

The essence of proof lies in the fact that, because of (10.10), for any $T>\varepsilon$,

$$\frac{1}{T}\int_0^T|\widehat{F}(t)|^n\,\mathrm{d}t\leqslant\frac{1}{T}\Big(\int_0^\varepsilon|\widehat{F}(t)|^n\,\mathrm{d}t+\int_\varepsilon^T|\widehat{F}(t)|^n\,\mathrm{d}t\Big)\leqslant$$
$$\frac{1}{T}\int_0^\infty\mathrm{e}^{-C(F)nt^2}\,\mathrm{d}t+\mathrm{e}^{-C(F)n}\leqslant C(F)(T^{-1}+\mathrm{e}^{-Cn}). \tag{10.13}$$

Similarly,

$$\int_{-T}^{T}\frac{|\widehat{F}^n(t)(\widehat{F}(t)-1)|}{|t|}\,\mathrm{d}t\leqslant 2\int_0^\varepsilon\frac{|\widehat{F}^n(t)(\widehat{F}(t)-1)|}{|t|}\,\mathrm{d}t+\frac{4}{T}\int_\varepsilon^T|\widehat{F}(t)|^n\,\mathrm{d}t=$$

$$2\int_0^\varepsilon \frac{|\,\widehat{F}^n(t)(\widehat{F}(t)-1)\,|}{|\,t\,|}\,\mathrm{d}t + C(F)(T^{-1}+\mathrm{e}^{-C(F)n}). \tag{10.14}$$

But in the neighborhood of zero we can use the fact that F has finite $1+\delta$ moment and mean zero. Consequently,

$$|\,\widehat{F}(t)-1\,| = \left|\int_{-\infty}^{\infty}(\mathrm{e}^{\mathrm{i}tx}-1)\,F\{\mathrm{d}x\}\right| = \left|\int_{-\infty}^{\infty}(\mathrm{e}^{\mathrm{i}tx}-1-\mathrm{i}tx)\,F\{\mathrm{d}x\}\right| \leqslant$$

$$\int_{-\infty}^{\infty}|\,\mathrm{e}^{\mathrm{i}tx}-1-\mathrm{i}tx\,|\,F\{\mathrm{d}x\} \leqslant \frac{1}{2}\int_{|tx|\leqslant 1} t^2x^2\,F\{\mathrm{d}x\} + 2\int_{|tx|>1}|\,tx\,|\,F\{\mathrm{d}x\} \leqslant$$

$$3\int_{-\infty}^{\infty}|\,tx\,|^{1+\delta}\,F\{\mathrm{d}x\} \leqslant C(F)|\,t\,|^{1+\delta} \tag{10.15}$$

and

$$\int_0^\varepsilon \frac{|\,\widehat{F}^n(t)(\widehat{F}(t)-1)\,|}{|\,t\,|}\,\mathrm{d}t \leqslant C(F)\int_0^\infty \mathrm{e}^{-C(F)nt^2}t^\delta\,\mathrm{d}t \leqslant C(F)n^{-\delta}.$$

Now it remains to use (10.3) with $T=n$.

10.3 Exponential smoothing

We shall consider situation very similar to the previous example . Let F do not depend on n, have mean zero, finite second moment satisfy Cramer's (C) condition. We prove that, for fixed $k\in\mathbb{N}$,

$$|\,F^n(F-E)^k\,| \leqslant C(F,k)n^{-k}. \tag{10.16}$$

We do not have the difference of distributions. There are at least two possibilities. The first one, is to use (10.5). Indeed, set $W=F^n(F-E)^k$. Then, for $|\,y\,|\leqslant c(b)/T$,

$$|\,W(x+y)-W(x)\,| \leqslant \int_{-\infty}^{\infty}|\,F^n(x+y-t)-F^n(x-t)\,|\,|\,(F-E)^k\{\mathrm{d}t\}\,| \leqslant$$

$$Q(F^n,|\,y\,|)\|\,F-E\,\|^k \leqslant 2^kQ(F^n,|\,y\,|) \leqslant C(b,k)Q(F^n,c(b)/T) \leqslant C(b,k)Q(F^n,1/T).$$

Now, taking into account the properties of concentration functions, the second summand in (10.1) can be estimated by

$$bT\sup_x \int_{|y|\leqslant c(b)/T}|\,W(x+y)-W(x)\,|\,\mathrm{d}y, \leqslant C(b,k)Q(F^n,1/T) \leqslant$$

$$C(b,k)\frac{1}{T}\int_0^T|\,\widehat{F}(t)\,|^n\,\mathrm{d}t. \tag{10.17}$$

On the other hand, the same estimate one can get from (10.3). Indeed, let non-negative measures W_1 and W_2 be defined by the following relation:

$$W = \sum_{m\ even}^{n} \binom{k}{m} (-1)^m F^{k-m+n} - \sum_{m\ odd}^{n} \binom{k}{m} (-1)^{m+1} F^{k-m+n} = W_1 - W_2.$$

Due to the fact that $W\{\mathbb{R}\} = 0$, we have $\| W_1 \| = \| W_2 \|$. Therefore, we can apply (10.3) to the difference of distibutions $V_1 - V_2$, where $V_j = W_j / \| W_1 \| \in \mathcal{F}$. Thus

$$| W | = \| W_1 \| \, | V_1 - V_2 | \leqslant$$

$$\| W_1 \| \frac{1}{2\pi} \int_{-T}^{T} \frac{| \widehat{V}_1(t) - \widehat{V}_2(t) |}{| t |} \, dt + \frac{1}{T} \| W_1 \| \int_{-T}^{T} (| \widehat{V}_1(t) | + | \widehat{V}_2(t) |) \, dt \leqslant$$

$$\frac{1}{2\pi} \int_{-T}^{T} \frac{| \widehat{F}^n(t)(\widehat{F}(t) - 1)^k |}{| t |} \, dt + \frac{1}{T} \int_{-T}^{T} \sum_{m=0}^{n} \binom{k}{m} | \widehat{F}(t) |^{k-m+n} \, dt \leqslant$$

$$\frac{1}{2\pi} \int_{-T}^{T} \frac{| \widehat{F}^n(t)(\widehat{F}(t) - 1)^k |}{| t |} \, dt + \frac{2^k}{T} \int_{-T}^{T} | \widehat{F}(t) |^n \, dt. \tag{10.18}$$

The proof now is very similar to the proof of the previous example when $\delta = 1$. Indeed, we choose small $\varepsilon = \varepsilon(F)$, such that, for $| t | \leqslant \varepsilon$,

$$| \widehat{F}(t) | \leqslant \exp\{-C(F) t^2\}. \tag{10.19}$$

Now exactly as in the previous example we establish that, for $T > \varepsilon$,

$$\frac{1}{T} \int_{0}^{T} | \widehat{F}(t) |^n \, dt \leqslant C(F)(T^{-1} + e^{-C(F)n})$$

and

$$\int_{-T}^{T} \frac{| \widehat{F}^n(t)(\widehat{F}(t) - 1)^k |}{| t |} \, dt \leqslant 2 \int_{0}^{\varepsilon} \frac{| \widehat{F}^n(t)(\widehat{F}(t) - 1)^k |}{| t |} \, dt + C(F)(T^{-1} + e^{-C(F)n}).$$

From (10.15) we have

$$| \widehat{F}(t) - 1 | \leqslant C(F) | t |^2.$$

and

$$\int_{0}^{\varepsilon} \frac{| \widehat{F}^n(t)(\widehat{F}(t) - 1)^k |}{| t |} \, dt \leqslant C(F) \int_{0}^{\infty} e^{-C(F)nt^2} t^{2k-1} \, dt \leqslant C(F) n^{-k}.$$

Collecting all estimates we can apply (10.18) with $T = n^k$.

Exercises

1. To prove analogue of the Berry- Esseen theorem for the sum of independent non-identically distributed summands. That is to prove that

$$\Big|\prod_{j=1}^{n} F_j - \Phi_\sigma\Big| \leqslant C\frac{\sum_1^n \beta_3(F_i)}{(\sum_1^n \sigma^2(F_i))^3}.$$

   Here $i = 1, 2, \dots, n$ and $F_i \in \mathcal{F}$ and

$$\int_{-\infty}^{\infty} x\, F_i\{\mathrm{d}x\} = 0, \quad \int_{-\infty}^{\infty} x^2\, F_i\{\mathrm{d}x\} = \sigma^2(F_i); \quad \int_{-\infty}^{\infty} |\, x\,|^3\, F_i\{\mathrm{d}x\} = \beta_3(F_i) < \infty$$

   and Φ_σ denotes the normal distribution with the variance $\sigma^2 = \sum_1^n \sigma^2(F_i)$, i.e. having the characteristic function

$$\widehat{\Phi}_\sigma(t) = \exp\Big\{-\frac{1}{2}\sum_{i=1}^{n}\sigma^2(F_i)t^2\Big\}.$$

2. Let F do not depend on n, have two moments and satisfy Cramer's (C) condition. To prove that,

$$|\, F^n - \exp\{n(F - E)\}\,| \leqslant C(F)(n^{-1}). \tag{10.20}$$

Bibliographical comments

Lemmas 10.1 and 10.3 are due to Esseen (1945). Lemma 10.2 was proved in Bentkus and Götze (1994). The Berry-Esseen theoremwas obtained by Berry (1941) and Esseen (1942). Subsequent convolutions under (C) condition were investigated in Čekanavičius (1994). Similar problems were considered by Zaitsev (1992,1996) and Studnev (1960).

11 Centering in compound approximations

11.1 Formula of inversion

In this Section we investigate one method for approximation of suitably centered compound distributions. It involves the following formula of inversion. We recall that $W = W^+ - W^-$ is the Jordan-Hahn decomposition of W.

Lemma 11.1 *Let $W \in \mathcal{M}$, $W\{\mathbb{R}\} = 0$, $\| W \| > 0$. Then for any $h > 0$*

$$| W | \leqslant C \int_0^{1/h} \frac{| \widehat{W}(t) |}{t} \mathrm{d}t + C \| W \| \min\{Q(\tilde{W}^+, h), Q(\tilde{W}^-, h)\}. \tag{11.1}$$

Here $\tilde{W}^+ = 2W^+/\| W \|$, $\tilde{W}^- = 2W^-/\| W \|$.

The estimate (11.1) is a special version of Esseen's smoothing inequality with a different remainder.

Remark 11.1 *If $W = FV$, where $F \in \mathcal{F}$ and $V \in \mathcal{M}$, then the second summand can be replaced by $\| V \| Q(F, h)$.*

Usually (11.1) is written in the following form. Let $F \in \mathcal{F}$, $G \in \mathcal{F}$, $h > 0$. Then

$$| F - G | \leqslant C \int_0^{1/h} \left| \frac{\widehat{F}(t) - \widehat{G}(t)}{t} \right| \mathrm{d}t + C \min\{Q(F, h), Q(G, h)\}. \tag{11.2}$$

Note that both formulas are valid for arbitrary chosen distributions (measures), not only for the discrete ones. The benefit of (11.2) lies in the fact that outside the neighborhood of zero we can choose between two distributions.

11.2 Estimates for characteristic functions

For more effective application of (11.1) we need some information about the behavior of characteristic functions near the zero point. The standard approach is to assume the existence of some finite moments and to use expansion (1.19). However, F might have no finite moment. The way out it to decompose F as a sum of two distributions:

- distribution concentrated on a finite interval and
- distribution concentrated outside that interval.

  Moreover, we can

- shift distribution ensuring that the first component has the zero mean, and

– the second distribution has sufficient probabilistic mass outside the chosen interval.

We shall write this decomposition more precisely. Let $0 \leqslant p \leqslant C < 1$, and let $F \in \mathcal{F}$. Then there exist $A, B \in \mathcal{F}$, h_-, h_+ and $u \in \mathbb{R}$ such that

$$FE_u = (1-p)A + pB \quad FE_u\{(-\infty, -h_-]\} \geqslant p/2\,, \quad FE_u\{[h_+, \infty)\} \geqslant p/2\,, \tag{11.3}$$

$$A\{[-h_-, h_+]\} = 1\,, \quad B\{(-h_-, h_+)\} = 0\,, \quad \int x\, A\{\mathrm{d}x\} = 0\,. \tag{11.4}$$

Set $h = \max(h_-, h_+)$. Distribution B is concentrated outside the finite interval, which is not shorter than h. Therefore, if $h > 0$, then by definition

$$1 - Q(E_uF, h/2) \geqslant p/2.$$

Now, by the properties of concentration function (1.28) – (1.30) we get

$$Q((E_uF)^n, h) \leqslant C(np)^{-1/2}, \quad Q(\exp\{n(E_uF - E)\}, h) \leqslant C(np)^{-1/2}. \tag{11.5}$$

Distribution A is concentrated on the finite interval, consequently it has moments of all orders. Set

$$\sigma^2 = \int x^2\, A\{dx\}\,.$$

The properties of A can be summarized in the following Lemma.

Lemma 11.2 *Let A be defined as in above. Then*

a) for all $t \in \mathbb{R}$

$$|\,\widehat{A}(t) - 1\,| \leqslant \sigma^2 t^2/2\,, \tag{11.6}$$

b) if $h > 0$ then for all $|t| \leqslant 1/h$

$$\sigma^2 t^2 \leqslant 1, \quad |\,\exp\{\widehat{A}(t) - 1\}\,| \leqslant \exp\{-\sigma^2 t^2/3\}, \quad |\,\widehat{A}(t)\,| \leqslant \exp\{-\sigma^2 t^2/3\}\,. \tag{11.7}$$

Proof. Estimate (11.6) directly follows from (1.19). For the proof of (11.7) note that, for $|\,t\,| \leqslant 1/h$,

$$\left|\,\widehat{A}(t) - 1 + \frac{\sigma^2 t^2}{2}\,\right| = \left|\,\int\limits_{|x| \leqslant h} \left(\mathrm{e}^{\mathrm{i}tx} - 1 - \mathrm{i}tx + \frac{t^2 x^2}{2}\right) A\{\mathrm{d}x\}\,\right| \leqslant$$

$$\frac{1}{6} \int\limits_{|x| \leqslant h} |\,tx\,|^3\, A\{\mathrm{d}x\} \leqslant \frac{1}{6} \int\limits_{|x| \leqslant h} |\,tx\,|^2\, A\{\mathrm{d}x\} = \frac{\sigma^2 t^2}{6}.$$

The estimate (11.7) follows from :

$$|\,\widehat{A}(t)\,| \leqslant |\,\widehat{A}(t) - 1 + \sigma^2 t^2/2\,| + |\,1 - \sigma^2 t^2\,| \leqslant$$

$$\frac{\sigma^2t^2}{6}+1-\frac{\sigma^2t^2}{2}\leqslant \exp\Big\{-\frac{\sigma^2t^2}{3}\Big\}.$$

Similarly,

$$Re\,(1-\widehat{A}(t))\geqslant \frac{\sigma^2t^2}{2}-\Big|\,Re\,\Big(\widehat{A}(t)-1+\frac{\sigma^2t^2}{2}\Big)\Big|\geqslant \frac{\sigma^2t^2}{3}.$$

□

11.3 Examples

Let us observe that for any $F\in\mathcal{F}$ and $a>1$:

$$|\,(F-E)\exp\{a(F-E)\}\,|\leqslant \|\,(E_1-E)\exp\{a(E_1-E)\}\,\|\leqslant Ca^{-1/2}. \tag{11.8}$$

Estimate (11.8) can be proved by many various methods. The first inequality follows from the properties of total variation norm. The second one can be proved as in Section 5. Moreover, it can be proved directly:

$$|\,(E_1-E)\exp\{a(E_1-E)\}\,|=\Big|\sum_{k=0}^{\infty}\frac{a^k}{k!}\,\mathrm{e}^{-a}E_{k+1}-\sum_{k=0}^{\infty}\frac{a^k}{k!}\,\mathrm{e}^{-a}E_k\Big|\leqslant\frac{1}{a}\sum_{k=0}^{\infty}\frac{|\,k-a\,|}{k!}\mathrm{e}^{-a}.$$

Let η be a Poisson random variable with parameter a. Then the last sum equals

$$\mathbb{E}\,|\,\eta-a\,|\leqslant\sqrt{\mathbb{E}\,(\eta-a)^2}\leqslant\sqrt{a}.$$

Consequently, we proved (11.8) with $C=1$.

The first example will show, how appropriate centering can improve (11.8).

Example 1. Let $F\in\mathcal{F}$, $a>0$. Then

$$\inf_u|\,(FE_u-E)\exp\{a(FE_u-E)\}\,|\leqslant Ca^{-3/4}. \tag{11.9}$$

We begin the proof of (11.9) from decomposition of FE_u as shown in (11.3) and (11.4). The value of $0<p<0.5$ will be determined later. Note that

$$\exp\{a(FE_u-E)\}=\exp\{a(1-p)(A-E)\}\exp\{ap(B-E)\}. \tag{11.10}$$

By the properties of uniform distance and (11.8) we obtain

$$\begin{gathered}|\,(FE_u-E)\exp\{a(FE_u-E)\}\,|\leqslant\\ |\,(1-p)(A-E)\exp\{a(FE_u-E)\}\,|+|\,p(B-E)\exp\{a(FE_u-E)\}\,|\leqslant\\ |\,(A-E)\exp\{a(FE_u-E)\}\,|+p\,\|\,(E_1-E)\exp\{ap(E_1-E)\}\,\|\leqslant\\ |\,(A-E)\exp\{a(FE_u-E)\}\,|+C\sqrt{\frac{p}{a}}\,.\end{gathered} \tag{11.11}$$

It remains to estimate the first summand in (11.11). Let $h = max\{h_-, h_+\}$. If $h = 0$, then $A = E$ and the first summand simply vanishes. If $h > 0$, then we apply (11.1)

$$|\,(A-E)\exp\{a(FE_u-E)\}\,| \leqslant C\int\limits_0^{1/h}\left|\frac{(\widehat{A}(t)-1)\exp\{a(1-p)(\widehat{A}(t)-1)\}}{t}\right|\mathrm{d}t+$$

$$C\|\,(A-E)\exp\{a(1-p)(A-E)/2\}\,\|Q(\exp\{a(FE_u-E)/2\},h). \tag{11.12}$$

Concentration function was already estimated in (11.5). By 11.8 we get

$$\|\,(A-E)\exp\{a(1-p)(A-E)/2\}\,\| \leqslant Ca^{-1/2}.$$

Taking into account (11.6) and (11.7) we obtain

$$C\int\limits_0^{1/h}\left|\frac{(\widehat{A}(t)-1)\exp\{a(1-p)(\widehat{A}(t)-1)\}}{t}\right|\mathrm{d}t \leqslant C\int\limits_0^{\infty}\sigma^2|\,t\,|\exp\Big\{-a(1-p)\frac{\sigma^2t^2}{3}\Big\}\,\mathrm{d}t \leqslant Ca^{-1}.$$

Collecting (11.11), (11.12) and (11.5) we finally obtain:

$$|\,(FE_u-E)\exp\{a(FE_u-E)\}\,| \leqslant \frac{C}{a}+C\sqrt{\frac{p}{a}}+\frac{C}{\sqrt{a^2p}}.$$

Now it remains to choose p. However, we see that the best possible choice is

$$\sqrt{\frac{p}{a}} = \frac{1}{\sqrt{a^2p}}, \quad p = a^{-1/2}.$$

This evidently completes the proof of (11.9).

Example 2. The second example deals with the compound Poisson approximation. We shall prove that, for any $F \in \mathcal{F}$,

$$\inf_u|\,F^n - E_{-nu}\exp\{n(FE_u-E)\}\,| \leqslant Cn^{-1/3}. \tag{11.13}$$

First, we note that (11.13) can be written as

$$\inf_u|\,(FE_u)^n - \exp\{n(FE_u-E)\}\,| \leqslant Cn^{-1/3}.$$

Let $p = n^{-1/3}$ and let us choose $A \in \mathcal{F}$ and h in decomposition $E_uF = (1-p)A + pB$ as defined by (11.3) and (11.4). Note that, we can choose n sufficiently large. For example, we can take $n > 8$. From the triangle inequality we have

$$|\,(FE_u)^n - \exp\{n(FE_u-E)\}\,| \leqslant J_1 + J_2.$$

Here

$$J_1 = \Big|\,((1-p)A+pB)^n - \exp\{n(1-p)(A-E)\}((1-p)E+pB)^n\,\Big|,$$

and

$$J_2 = \Big|\ \exp\{n(1-p)(A-E)\}\big(((1-p)E+pB)^n - \exp\{np(B-E)\}\big)\ \Big|.$$

Now by the properties of metrics and (5.8) we get the following estimate

$$J_2 \leqslant \|\,((1-p)E+pB)^n - \exp\{np(B-E)\}\,\| \leqslant Cp = Cn^{-1/3}. \tag{11.14}$$

For the estimate of J_1 we shall apply (11.2). From the estimate (11.5) we obtain

$$|\,J_1\,| \leqslant C\int\limits_0^{1/h}\left|\frac{\widehat{J}_1(t)}{t}\right|\mathrm{d}t + Cn^{-1/3}. \tag{11.15}$$

Taking into account that $B\in\mathcal{F}$ and its characteristic function is less or equal to 1, we shall estimate $\widehat{J}_1(t)$ in the following way:

$$|\,\widehat{J}_1(t)\,| \leqslant \sum_{k=0}^{n}\binom{n}{k}(1-p)^k p^{n-k}|\,\widehat{B}(t)^{n-k}\,|\,\Big|\ \widehat{A}(t)^k - \exp\{n(1-p)(\widehat{A}(t)-1)\}\ \Big| \leqslant$$

$$\sum_{k=1}^{n}\binom{n}{k}(1-p)^k p^{n-k}\Big|\ \widehat{A}(t)^k - \exp\{k(\widehat{A}(t)-1)\}\ \Big| +$$

$$\sum_{k=0}^{n}\binom{n}{k}(1-p)^k p^{n-k}\Big|\ \exp\{k(\widehat{A}(t)-1)\} - \exp\{n(1-p)(\widehat{A}(t)-1)\}\ \Big| \tag{11.16}$$

Applying (11.6) and (11.7) we obtain

$$\Big|\ \widehat{A}(t)^k - \exp\{k(\widehat{A}(t)-1)\}\ \Big| \leqslant Ck(\sigma^2t^2)\exp\Big\{-\frac{k\sigma^2t^2}{3}\Big\}.$$

Consequently, for $k>0$,

$$\int\limits_0^{1/h}\left|\frac{\widehat{A}(t)^k - \exp\{k(\widehat{A}(t)-1)\}}{t}\right|\mathrm{d}t \leqslant \frac{C}{k} \leqslant \frac{C}{k+1}.$$

Now applying (1.11) we get the estimate for the first sum:

$$\sum_{k=1}^{n}\binom{n}{k}(1-p)^k p^{n-k}\Big|\ \widehat{A}(t)^k - \exp\{k(\widehat{A}(t)-1)\}\ \Big| \leqslant$$

$$C\sum_{k=1}^{n}\binom{n}{k}(1-p)^k p^{n-k}(k+1)^{-1} \leqslant \frac{C}{np} \leqslant Cn^{-1/3}. \tag{11.17}$$

Similarly,

$$\Big|\ \exp\{k(\widehat{A}(t)-1)\} - \exp\{n(1-p)(\widehat{A}(t)-1)\}\ \Big| \leqslant$$

$$C|\,k-n(1-p)\,|\sigma^2t^2\exp\Big\{-\min\{k,n(1-p)\}\frac{\sigma^2t^2}{3}\Big\}.$$

Let $k > 0$. Then

$$\int_0^{1/h} \left| \frac{\exp\{k(\widehat{A}(t)-1)\} - \exp\{n(1-p)(\widehat{A}(t)-1)\}}{t} \right| \mathrm{d}t \leqslant$$

$$C\frac{|\,k-n(1-p)\,|}{k+1} + C\frac{|\,k-n(1-p)\,|}{n(1-p)}. \tag{11.18}$$

For $k = 0$, we can prove (11.18) directly. It suffices to note that

$$\sigma^2 = \int_{-\infty}^{\infty} x^2\, A\{\mathrm{d}x\} = \int_{-h}^{h} x^2\, A\{\mathrm{d}x\} \leqslant h^2.$$

Therefore,

$$\int_0^{1/h} n(1-p)\sigma^2 t\,\mathrm{d}t \leqslant Cn(1-p)\frac{\sigma^2}{h^2} \leqslant Cn(1-p).$$

Thus, we established that (11.18) holds for $k = 0$ too. Let ξ be the Binomial random variable with parameters n and $1-p$. Then taking into account (1.11) we obtain

$$|\,J_1\,| \leqslant C \int_0^{1/h} \left| \frac{\widehat{J}_1(t)}{t} \right| \mathrm{d}t \leqslant$$

$$C\sum_{k=0}^{n} \binom{n}{k} (1-p)^k p^{n-k} |\,k-n(1-p)\,| \left(\frac{1}{n(1-p)} + \frac{1}{k+1)} \right) \leqslant$$

$$\frac{C}{n(1-p)}\mathbb{E}\,|\,\xi - \mathbb{E}\xi\,| + C\sqrt{\mathbb{E}(\xi-\mathbb{E}\xi)^2} \left(\sum_{k=0}^{n} \binom{n}{k} (1-p)^k p^{n-k} \frac{1}{(k+1)^2} \right)^{1/2} \leqslant$$

$$C\sqrt{\frac{p}{n}} + C\sqrt{\frac{np(1-p)}{n^2p^2}} \leqslant Cn^{-1/3}. \tag{11.19}$$

Collecting (11.14) and (11.19) we complete the proof of (11.13).

Exercises

1. To prove that

$$\inf_u |\,(FE_u - E)^k \exp\{a(FE_u - E)\}\,| \leqslant C(k)a^{-k/2-k/(2k+2)}.$$

2. To prove that

$$\inf_u |\,(FE_u)^{n+1} - (FE_u)^n\,| \leqslant Cn^{-1/3}.$$

3. To get the estimate for

$$\inf_u |\,(FE_u - E)^k (FE_u)^n\,|.$$

Bibliographical notes

The inversion formula (11.1) was introduced by Le Cam (1965). Decomposition of FE_u and properties of characteristic functions can be traced back to Prohorov (1955). The estimate (11.13) was obtained by Le Cam (1965) and is a slightly different version of Kolmogorov's (1963) result.

12 The triangle function method

In this section, we give the necessary facts and statements, needed for application of the triangle function method, but usually omit their proofs (which, as a rule, are very sophisticated). All omitted proofs can be found in Arak and Zaĭtsev (1988).

12.1 General idea

The triangle function method was primarily designed for the estimation of compound distributions. First of all we discuss the general idea. In this section we assume that $W \in \mathcal{M}$ and $W\{\mathbb{R}\} = 0$. For $h > 0$ let us introduce the following pseudo-metric:

$$|\, W \,|_h = \sup_z |\, W\{[z, z+h]\} \,|. \tag{12.1}$$

It is easy to note that

$$|\, W \,| \leqslant \sup_{h \geqslant 0} |\, W \,|_h \leqslant 2|\, W \,|.$$

If $F \in \mathcal{F}$, then $|\, F \,|_h = Q(F, h)$. Let $\mathbb{I}\,\{[z, z+h]\}$ be indicator function, i.e.

$$\mathbb{I}\,\{[z, z+h]\}(x) = \begin{cases} 1, & \text{if } x \in [z, z+h], \\ 0, & \text{if } x \notin [z, z+h]. \end{cases}$$

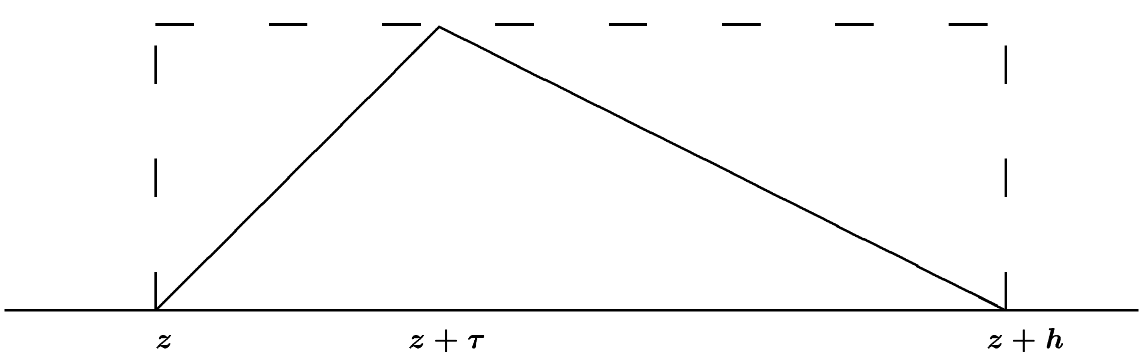
Then

$$|\, W \,|_h = \sup_z \left| \int_{-\infty}^{\infty} \mathbb{I}\,\{[z, z+h]\}(x)\, W\{\mathrm{d}x\} \right|.$$

The triangle function method means replacement of the indicator function by some other function with better properties. Let us introduce the set of triangle functions

$$\{f_{z,h,\tau}(x) \;:\; z \in \mathbb{R},\; 0 < \tau \leqslant h\}\,.$$

Here $f_{z,h,\tau}(x)$ increases linearly from 0 to 1 on $[z, z+\tau]$, decreases linearly to 0 on $[z+\tau, z+h]$ and is equal to zero for $x \notin [z, z+h]$.



Now set

$$|W|_{h,\tau} = \sup_z \left| \int_{-\infty}^{\infty} f_{z,h,\tau}(x)\, W\{\mathrm{d}x\} \right|. \tag{12.2}$$

Pseudo-metric $|\cdot|_{h,\tau}$ has the following useful properties. Let $W, V \in \mathcal{M}$, $W\{\mathbb{R}\} = 0$, $a > 0$. Then

$$|WV|_{h,\tau} \leqslant \|V\|\, |W|_{h,\tau}, \quad |aW|_{h,\tau} = a|W|_{h,\tau}, \tag{12.3}$$

$$|W|_{h,\tau} \leqslant |W|_{h,\omega} + |W|_{\omega,\tau}. \tag{12.4}$$

It is obvious, that $f_{z,h,\tau}(x)$ is not a very good replacement for $\mathbb{I}\{[z, z+\tau]\}$. On the other hand, we can replace the indicator function by the sum of triangle functions with weights and apply (12.4). The general idea of the triangle function method is the following:

> If we can estimate $|W|_{h,\tau}$ for a sufficiently large set of h and τ, then we can get estimates for $|W|_h$ and for $|W|$.

This fact is formulated in Lemma 12.1 below. However, firstly we need an additional information. As a rule, the triangle function method is used when we can get the estimate

$$|\widehat{W}(t)| \leqslant C\varepsilon \widehat{D}(t), \tag{12.5}$$

where D is symmetric infinitely divisible distribution. Usually $D = \exp\{\lambda(F - E)\}$, for some symmetric $F \in \mathcal{F}$ and $\lambda > 0$. Set

$$\gamma_h = Q(D, h).$$

Typically, the triangle function method allows estimates of $|W|_{h,\tau}$ in terms of γ_h. Then, as follows from Lemma 12.1 below, it is possible to get the estimate for $|\cdot|_h$ and, consequently, for $|\cdot|$.

We recall that, for $F \in \mathcal{F}$, $F^{(-)}$ denotes distribution, for any Borel set X, satisfying $F^{(-)}\{X\} = F\{-X\}$. Note that $\widehat{F^{(-)}}(t) = \widehat{F}(-t)$.

Lemma 12.1 *Let $a > 0$, $b > 0$ be some absolute constants, $W \in \mathcal{M}$, $W\{\mathbb{R}\} = 0$, $\gamma_0 > 0$, and let, for any h and τ such that $0 < \tau \leqslant h/2$ and $\gamma_\tau \geqslant \gamma_h/8$,*

$$\max\{|W|_{h,\tau}, |W^{(-)}|_{h,\tau}\} \leqslant C_1 \varepsilon \gamma_h^a (|\ln \gamma_h| + 1)^b. \tag{12.6}$$

Then, for any $h > 0$,

$$|W|_h \leqslant C_2 \leqslant C_1 \varepsilon \gamma_h^a (|\ln \gamma_h| + 1)^b. \tag{12.7}$$

Remark 12.1 *Taking into account that $\gamma_h \leqslant 1$, we can see that, from (12.7) it is not difficult to obtain estimate $|W| \leqslant C\varepsilon$.*

Passing from the estimates valid for all $0 \leqslant \tau \leqslant h/2$ to the estimates valid for all $0 \leqslant \tau \leqslant h/2$ *and* $\gamma_\tau \geqslant \gamma_h/8$ is far from trivial. The following Lemma, though somewhat cumbersome, gives a quite good idea about the structure off estimates allowing this step.

Lemma 12.2 *Let $W \in \mathcal{M}$, $W\{\mathbb{R}\} = 0$ and let H_1 be finite non-negative measure, for any $\tau > 0$, allowing decomposition*

$$H_1 = H_{\tau 2} + H_{\tau 3} + H_{\tau 4}.$$

Here non-negative measures $H_{\tau 2}$, $H_{\tau 3}$ and $H_{\tau 4}$ satisfy relations

$$H_{\tau 3}\{\mathbb{R}\} \leqslant C_3(|\ln \gamma_\tau| + 1)^3, \tag{12.8}$$

$$H_{\tau 2}\{[-\tau, \tau]\} = H_{\tau 3}\{[-\tau, \tau]\} = H_{\tau 4}\{\mathbb{R} \setminus [-\tau, \tau]\} = 0. \tag{12.9}$$

Let, for any $0 \leqslant \tau \leqslant h/2$,

$$\max\{|W|_{h,\tau}, |W^{(-)}|_{h,\tau}\} \leqslant C_4 \varepsilon \Big\{ \gamma_h^\beta (|\ln \gamma_h| + 1)^\delta (H_1\{\{x : |x| > \tau\}\})^m + \gamma_h^\alpha \Big\}, \tag{12.10}$$

$$\max\{|W|_{h,\tau}, |W^{(-)}|_{h,\tau}\} \leqslant C_5 \varepsilon \left\{ \frac{(|\ln \gamma_\tau| + 1)^s}{\sqrt{H_{\tau 2}\{\{x : |x| > h/4\}\}}} + \gamma_h^\alpha \right\}, \tag{12.11}$$

where $0 < \beta, m, \alpha \leqslant C_6$ and $0 \leqslant \delta, s \leqslant C_7$ and $\varepsilon > 0$. Then, for any τ and h such that $0 < \tau \leqslant h/2$ and $\gamma_\tau \geqslant \gamma_h/8$,

$$|W|_{h,\tau} \leqslant C_8 \varepsilon \gamma_h^a (|\ln \gamma_h| + 1)^b. \tag{12.12}$$

Here

$$a = \min\left\{\alpha, \frac{\beta}{2m+1}\right\}, \quad b = \frac{2ms + \delta}{2m+1}.$$

Moreover, for all $h > 0$,

$$|W|_h \leqslant C_9 \varepsilon \gamma_h^a (|\ln \gamma_h| + 1)^b. \tag{12.13}$$

Proof. For $u > 0$ let

$$\nu_u(y) = H_{u2}\{\{x : |x| > y\}\}; \quad \kappa_u = \gamma_u^{-2\beta/(2m+1)} (|\ln \gamma_u| + 1)^{2(s-\delta)/(2m+1)}.$$

Let $0 < \tau \leqslant h/2$ and $\gamma_h \leqslant 8\gamma_\tau$. Then

$$|\ln \gamma_\tau| \leqslant |\ln \gamma_h| + \ln 8.$$

We consider three cases:

1. $\nu_\tau(h/4) \geqslant \kappa_h$;
2. $\nu_\tau(\tau) \leqslant \kappa_h$;
3. $\nu_\tau(h/4) < \kappa_h < \nu_\tau(\tau)$.

In the case 1, we can apply (12.11). In the case 2, we have

$$H_1\{\{x : |\,x\,| > \tau\}\} \leqslant H_{\tau 2}\{\{x : |\,x\,| > \tau\}\} + H_{\tau 3}\{\{x : |\,x\,| > \tau\}\} \leqslant$$

$$\nu_\tau(\tau) + C(|\,\ln\gamma_h\,| + 1)^3 \leqslant C\Big(\gamma_h^{-2\beta/(2m+1)}(|\,\ln\gamma_h\,| + 1)^{2(s-\delta)/(2m+1)} + (|\,\ln\gamma_h\,| + 1)^3\Big).$$

Consequently,

$$(H_1\{\{x : |\,x\,| > \tau\}\})^m \leqslant C\Big\{\gamma_h^{-2\beta m/(2m+1)}(|\,\ln\gamma_h\,| + 1)^{2(s-\delta)m/(2m+1)} + (|\,\ln\gamma_h\,| + 1)^{3m}\Big\}.$$

Now the estimate (12.12) follows from (12.10). In the case 3, we note that $\nu_\tau(y)$ is non-increasing and we can find y such that

$$\tau < y < h/4, \quad \nu_\tau(2y) \leqslant \kappa_h \leqslant \nu_\tau(y/2).$$

Setting $\omega = 2y$, we can rewrite the last inequalities as

$$2\tau < \omega < h/2, \quad \nu_\tau(\omega) \leqslant \kappa_h \leqslant \nu_\tau(\omega/4).$$

By the property of pseudometric (12.4):

$$|\,W\,|_{h,\tau} \leqslant |\,W\,|_{h,\omega} + |\,W\,|_{\omega,\tau}.$$

For the estimate of $|\,W\,|_{\omega,\tau}$ it suffices to apply (12.11) replacing h by ω and taking into account that $\nu_\tau(\omega/4) \geqslant \kappa_h$. For the estimate of $|\,W\,|_{h,\omega}$ we apply (12.10). Note that

$$H_1\{\{x : |\,x\,| > \omega\}\} \leqslant \nu_\tau(\omega) + H_{\tau 3}\{\{x : |\,x\,| > \tau\}\} \leqslant \kappa_h + C(|\,\ln\gamma_h\,| + 1)^3.$$

Collecting all estimates we get (12.12). The estimate (12.13) follows from Lemma 12.1. $\square$.

Note that, idea of the proof can be applied in many similar situations which involve two estimates outside the finite interval $[-\tau, \tau]$ through some measure $H_{\tau 2}$.

Considering (12.10) and (12.11) we see that the main requirement is two different estimates for distributions which are concentrated outside finite interval $[-\tau, \tau]$. It is not a coincidence. The triangle function method is close to the one used in previous Section. The main idea is to decompose measure under estimation into two components: the measure concentrated on some finite interval and the other one concentrated outside this interval. The second measure must be estimated as in Lemma 12.2, because the standard estimate through concentration function as in (11.1) or (11.2) is too rough and, in general, can not ensure the accuracy better than $n^{-1/2}$. On the other hand, for the measure concentrated on the finite interval we can use the standard approach estimating it through Fourier transform in the neighborhood of zero. For this we need some properties of the Fourier transforms for $f_{z,h,\tau}(x)$.

12.2 Fourier transform of the triangle function

As we see from Lemma 12.2, usually it suffices to get estimates of the special form for $|\,W\,|_{h,\tau}$ and $0<\tau\leqslant h/2$. We summarize the technical aspects of the triangle function method in the following statements:

1. For measures concentrated on $[-\tau,\tau]$ the Parseval's identity and properties of $\widehat{f}_{z,h,\tau}(t)$ are used.
2. For estimation of the measure concentrated outside $[-\tau,\tau]$ the triangle function is replaced by some special function which has very similar properties for $x\in K_m(\mathbf{u})$.
3. As a rule, the measure concentrated outside $[-\tau,\tau]$ can be decomposed as a sum of two components: the one concentrated on the neighborhood of $K_m(\mathbf{u})$ with $\mathbf{u}\in\mathbb{R}^\ell$ and the one concentrated outside this neighborhood. Moreover, the second component has small total variation norm.

We discuss every statement separately.

1. The triangle function $f_{z,h,\tau}(x)$ has Fourier transform with quite good properties in the neighborhood of zero.

Lemma 12.3 *Let $z\in\mathbb{R}$, $0<\tau\leqslant h/2$. Then*

a) for all $t\in\mathbb{R}$

$$|\,\widehat{f}_{z,h,\tau}(t)\,|\leqslant \min\left\{\frac{2}{|\,t\,|},\frac{h}{2}\right\}; \tag{12.14}$$

b) for $|\,t\,|>1/\tau$

$$|\,\widehat{f}_{z,h,\tau}(t)\,|\leqslant \frac{8\tau}{1+(t\tau)^2}. \tag{12.15}$$

Lemma 12.3 and Parseval's identity (1.26) allows to get the estimate for W concentrated on $[-\tau,\tau]$:

$$\left|\int\limits_{-\infty}^{\infty} f_{z,h,\tau}(x)\,W\{\mathrm{d}x\}\right|\leqslant\frac{1}{2\pi}\int\limits_{-\infty}^{\infty}|\,\widehat{f}_{z,h,\tau}(t)\,|\,|\,\widehat{W}(t)\,|\,\mathrm{d}t\leqslant C\int\limits_{-\infty}^{\infty}\frac{|\,\widehat{W}(t)\,|}{|\,t\,|}\,\mathrm{d}t. \tag{12.16}$$

However, if W is concentrated outside the interval $[-\tau,\tau]$, the estimate (12.16) is too rough.

2. We recall that

$$K_m(\mathbf{u})=\left\{\sum_{i=1}^{N} j_iu_i : j_i\in\{-m,-m+1,\dots,m\};\ i=1,\dots,N\right\}, \tag{12.17}$$

see (9.1). For W concentrated on $K_m(\mathbf{u})$ we used Arak's lemma 9.1. The following general Lemma shows that $f_{z,h,\tau}$ can be replaced by special function ω which has good properties in the neighborhood of $K_m(\mathbf{u})$.

Lemma 12.4 *Let ℓ and m be positive integers, $\mathbf{u} = (u_1, u_2, \dots, u_\ell) \in \mathbb{R}^\ell$, $z \in \mathbb{R}$ and $0 < \tau \leqslant h/2$. Then there exists a continuous function $\omega(x)$, $x \in \mathbb{R}$, with the following properties:*

$$0 \leqslant \omega(x) \leqslant f_{z,h,\tau}(x) \quad \text{for all } x \in \mathbb{R}, \tag{12.18}$$

$$\omega(x) = f_{z,h,\tau}(x) \quad \text{for } x \in [K_m(\mathbf{u})]_{m\tau}, \tag{12.19}$$

$$\sup_{t\in\mathbb{R}} |\,\widehat{\omega}(t)\,| \leqslant h/2, \tag{12.20}$$

$$\int\limits_{-\infty}^{\infty} \sup_{|s|\geqslant|t|} |\,\widehat{\omega}(t)\,| \,\mathrm{d}t \leqslant C\ell^2 \ln(\ell m + 1); \tag{12.21}$$

and, for any $F \in \mathcal{F}_+$,

$$\int\limits_{-\infty}^{\infty} |\,\widehat{\omega}(t)\,|\,\widehat{F}(t)\,\mathrm{d}t \leqslant CQ(F,h)\ell^2 \ln(\ell m + 1)\,. \tag{12.22}$$

We recall that by $[X]_\tau$ we denote a closed τ - neighborhood of the set X. The proof of (12.18)-(12.21) is quite sophisticated and can be found in Arak and Zaĭtsev (1988). The estimate (12.22) follows from Lemma 1.3.

We have

$$\int\limits_{-\infty}^{\infty} f_{z,h,\tau}(x)\,W\{\mathrm{d}x\} = \int\limits_{-\infty}^{\infty} \omega(x)\,W\{\mathrm{d}x\} + \int\limits_{\mathbb{R}^*} (f_{z,h,\tau}(x) - \omega(x))\,W\{\mathrm{d}x\}. \tag{12.23}$$

Here $\mathbb{R}^* = \mathbb{R} \setminus [K_m(\mathbf{u})]_{m\tau}$. It is obvious, that the first integral can be estimated by the Parseval's identity and Lemma 12.4. Meanwhile, the estimate of the second integral essentially depends on the choice of $\mathbf{u}$ and m, which we discuss below. Essentially $\omega(x)$ is used for obtaining estimates of the type (12.11).

3. In principle, for any $F \in \mathcal{F}$, it is always possible to choose $K_1(\mathbf{u})$ in such a way that almost all probabilistic mass of F is concentrated in the neighborhood of $K_1(\mathbf{u})$. We formulate this statement as Lemma.

Lemma 12.5 *Let $F_i \in \mathcal{F}$, $(i = 1, 2, \dots, n)$, $\tau > 0$, $a \in (0, 1]$. Then there exists $\mathbf{u} \in \mathbb{R}^\ell$ such that*

$$\ell \leqslant C(|\,\ln \gamma_{\tau,a}\,| + 1), \tag{12.24}$$

$$\sum_{i=1}^{n} F_i\{\mathbb{R} \setminus [K_1(\mathbf{u})]_\tau\} \leqslant Ca^{-1}\left(|\,\ln \gamma_{\tau,a}\,| + 1\right)^3. \tag{12.25}$$

Here

$$\gamma_{\tau,a} = Q\Big(\exp\Big\{a\sum_{i=1}^{n}(\tilde{F}_i - E)\Big\}, \tau\Big), \tag{12.26}$$

and $\tilde{F}_i \in \mathcal{F}$ *has the characteristic function* $Re\,\widehat{F}_i(t)$.

We shall formulate the variant of Lemma 12.5 for symmetric F.

Lemma 12.6 *Let* $F \in \mathcal{F}$ *be symmetric distribution,* $\gamma_\tau = Q(\exp\{\lambda(F - E)\}, \tau)$, $\tau > 0$, $\lambda > 0$. *Then there exists* $\mathbf{u} \in \mathbb{R}^\ell$ *such that*

$$\ell \leqslant C(|\ln \gamma_\tau| + 1), \tag{12.27}$$

$$F\{\mathbb{R} \setminus [K_1(\mathbf{u})]_\tau\} \leqslant \frac{C}{\lambda}\big(|\ln \gamma_\tau| + 1\big)^3. \tag{12.28}$$

Similar estimates can be obtained from Lemma 12.5 for other distributions or their mixtures. We need one additional Lemma for estimation of the characteristic function.

Lemma 12.7 *Let* $H = (1-p)F + pG$, $F, G \in \mathcal{F}$, $0 \leqslant p \leqslant 1$, F *is symmetric distribution and, for all* t, $\widehat{F}(t) \geqslant -\alpha > -1$. *Then, for any* $t \in \mathbb{R}$,

$$|\widehat{H}(t)| \leqslant \exp\Big\{\frac{(1-\alpha)(1-p)}{1 + p + \alpha(1-p)}\,(Re\,\widehat{H}(t) - 1)\Big\}. \tag{12.29}$$

Corollary 12.1 *Let* $\alpha \in [0, 1)$ *and let* F *be symmetric distribution, for any* $t \in \mathbb{R}$, *satisfying* $\widehat{F}(t) \geqslant -\alpha$. *Then, for any* $t \in \mathbb{R}$,

$$|\widehat{F}(t)| \leqslant \exp\Big\{\frac{1-\alpha}{1+\alpha}\,(\widehat{F}(t) - 1)\Big\}. \tag{12.30}$$

Now we can study how all Lemmas are applied in practice.

12.3 First example

Let F be symmetric distribution, $\lambda > 1$. Then

$$|(F - E)\exp\{\lambda(F - E)\}| \leqslant \frac{C}{\lambda}. \tag{12.31}$$

We recall the fact that, for any $G \in \mathcal{F}$,

$$\|\,(G-E)\exp\{\lambda(G-E)\}\,\| \leqslant \frac{C}{\sqrt{\lambda}}.$$

Thus, we see that the symmetry of distribution radically improves the accuracy. Note also that suitable centering gives an intermediate result, see (11.13). We prove (12.33) step by step.

Step 0. *Preliminary investigation.* First we check what estimate can be obtained for the Fourier-Stieltjes transform, i.e. (12.5). It can be easily established that

$$|\,(\widehat{F}(t)-1)\exp\{\lambda(\widehat{F}(t)-1)\}\,| \leqslant \frac{C}{\lambda}\exp\Big\{\frac{\lambda}{2}\,(\widehat{F}(t)-1)\Big\}.$$

Thus, the Fourier-Stieltjes transform is of the type (12.5) with

$$\widehat{D}(t) = \exp\Big\{\frac{\lambda}{2}(\widehat{F}(t)-1)\Big\}. \tag{12.32}$$

We apply the triangle function method. Therefore, we shall prove more general result than (12.31), i.e. we prove that if $h > 0$, then

$$|\,(F-E)\exp\{\lambda(F-E)\}\,|_h \leqslant \frac{C}{\lambda}\,\gamma_h^{1/3}(|\,\ln\gamma_h\,|+1)^{7/3}. \tag{12.33}$$

Here

$$\gamma_h = Q(D,h) = Q\bigg(\exp\Big\{\frac{\lambda}{2}\,(F-E)\Big\},h\bigg).$$

As was noted above, taking supremum over all $h > 0$ we get (12.31).

Step 1. *Decomposition of F.* Similarly to the the previous section we decompose F into the sum of distributions concentrated on finite interval and outside that interval. According to Lemma 12.6, for any $\tau > 0$, there exists $\mathbf{u} \in \mathbb{R}^{\ell}$ such that

$$\ell \leqslant C(|\,\ln\gamma_\tau\,|+1), \tag{12.34}$$

$$F\{\mathbb{R}\setminus[K_1(\mathbf{u})]_\tau\} \leqslant \frac{C}{\lambda}\,\big(|\,\ln\gamma_\tau\,|+1\big)^3. \tag{12.35}$$

Setting

$$q = F\{[-\tau,\tau]\}, \quad s = F\{[K_1(\mathbf{u})]_\tau\setminus[-\tau,\tau]\}, \quad d = F\{\mathbb{R}\setminus[K_1(U)]_\tau\},$$

we decompose F as a mixture of distributions $A, W, \Psi \in \mathcal{F}$ concentrated on the sets

$$[-\tau,\tau], \qquad [K_1(\mathbf{u})]_\tau\setminus[-\tau,\tau], \qquad \mathbb{R}\setminus[K_1(\mathbf{u})]_\tau\,,$$

respectively, i.e.,

$$F = qA + sW + d\Psi = qA + rV,$$

$$F - E = q(A-E) + s(W-E) + d(\Psi-E) = q(A-E) + r(V-E),$$

$$A\{[-\tau,\tau]\} = W\{[K_1(\mathbf{u})]_\tau \setminus [-\tau,\tau]\} = \Psi\{\mathbb{R} \setminus [K_1(\mathbf{u})]_\tau\} = 1. \tag{12.36}$$

From (12.35) it follows that

$$d \leqslant \frac{C}{\lambda}\,(|\ln \gamma_\tau| + 1)^3. \tag{12.37}$$

Step 2. *Estimating measures containing* $(A-E)$. Our goal is to get the estimate for $|\cdot|_{h,\tau}$ and to use Lemma 12.2. We begin from the part of measure containing the difference $(A-E)$. It is easy to see that

$$|\,(F-E)\exp\{\lambda(F-E)\}\,|_{h,\tau} \leqslant$$

$$|\,q(A-E)\exp\{\lambda(F-E)\}\,|_{h,\tau} + |\,r(V-E)\exp\{\lambda(F-E)\}\,|_{h,\tau}. \tag{12.38}$$

Let us estimate the first component of (12.38).

Lemma 12.8 *For all* $0<\tau\leqslant h/2$

$$|\,q(A-E)\exp\{\lambda(F-E)\}\,|_{h,\tau} \leqslant \frac{C}{\lambda}\gamma_h.$$

Proof. For the sake of brevity, set

$$g(t) = q\,|\,\widehat{f}_{z,h,\tau}(t)\,|\,|\,\widehat{A}(t)-1\,|\exp\Big\{\frac{\lambda}{2}(\widehat{F}(t)-1)\Big\}. \tag{12.39}$$

Applying Parseval's identity, we get

$$|\,q(A-E)\exp\{\lambda(F-E)\}\,|_{h,\tau} \leqslant$$

$$\frac{q}{2\pi}\int\limits_{-\infty}^{\infty}|\,\widehat{f}_{z,h,\tau}(t)\,|\,|\,\widehat{A}(t)-1\,|\exp\{\lambda(\widehat{F}(t)-1)\}\,\mathrm{d}t = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} g(t)\widehat{D}(t)\,\mathrm{d}t, \tag{12.40}$$

see (12.16). Here $\widehat{D}(t)$ is defined by (12.32). Estimating $g(t)$ we can distinguish between two cases: a) when t is near zero; b) when t is far from zero. We consider both cases separately.

a) Let $|\,t\,|\leqslant 1/\tau$, then $\widehat{f}_{z,h,\tau}(t)$ is not small (see Lemma 12.3):

$$|\,\widehat{f}_{z,h,\tau}(t)\,| \leqslant \frac{2}{|\,t\,|}.$$

On the other hand, A is concentrated on $[-\tau,\tau]$ and, for small t,

$$\frac{\sigma^2t^2}{3} \leqslant 1 - Re\,\widehat{A}(t) = 1-\widehat{A}(t) \leqslant \frac{\sigma^2t^2}{2}. \tag{12.41}$$

Here

$$\sigma^2 = \int x^2\,A\{dx\}\,,$$

see (11.6) and (11.7). Therefore, for all $|\,t\,|\leqslant 1/\tau$,

$$g(t) \leqslant \frac{2q}{|\,t\,|}\,|\,\widehat{A}(t)-1\,|\exp\Big\{\frac{q\lambda}{2}\,(\widehat{A}(t)-1)\Big\}\;\exp\Big\{\frac{p\lambda}{2}(\widehat{V}(t)-1)\Big\} \leqslant$$

$$Cq\sigma^2|\,t\,|\exp\Big\{-\frac{q\lambda\sigma^2t^2}{6}\Big\} \leqslant \frac{C}{\lambda}\sqrt{\lambda q\sigma^2}\exp\Big\{-\frac{q\lambda\sigma^2t^2}{12}\Big\} = \frac{C}{\lambda}\,g_1(t). \tag{12.42}$$

b) Now let $|t| > 1/\tau$. Then by (12.15)

$$|\widehat{f}_{z,h,\tau}(t)| \leqslant \frac{8\tau}{1+(t\tau)^2}$$

and

$$g(t) \leqslant C\frac{\tau}{1+(t\tau)^2}\,q(1-\widehat{A}(t))\exp\Big\{\frac{\lambda}{2}\,q(\widehat{A}(t)-1)\Big\} \leqslant \frac{C}{\lambda}\,\frac{\tau}{1+(t\tau)^2} = \frac{C}{\lambda}\,g_2(t). \qquad (12.43)$$

Collecting (12.42) and (12.43) we see that

$$g(t) \leqslant \frac{C}{\lambda}\begin{cases} g_1(t), & \text{if } |t| \leqslant 1/\tau, \\ g_2(t), & \text{if } |t| > 1/\tau. \end{cases}$$

Of course, we can roughly estimate integral in (12.40) by $C\lambda^{-1}$. However, we want to preserve γ_τ. Therefore, we shall apply Lemma 1.3. Note that $g_1(t)$ and $g_2(t)$ are even and vanishing as $|t| \to \infty$. Therefore

$$\sup_{s:|s|\geqslant|t|} |g(s)| \leqslant \frac{C}{\lambda}\big(g_1(t)+g_2(t)\big)$$

and

$$\int_{-\infty}^{\infty} \sup_{s:|s|\geqslant|t|} |g(s)|\,\mathrm{d}t \leqslant \frac{C}{\lambda}\int_{-\infty}^{\infty}\big(g_1(t)+g_2(t)\big)\,\mathrm{d}t \leqslant \frac{C}{\lambda}.$$

Moreover, by (12.14)

$$\sup_t g(t) \leqslant \sup_t |\widehat{f}_{z,h,\tau}(t)|\,q(1-\widehat{A}(t))\exp\Big\{\frac{q\lambda}{2}(\widehat{A}(t)-1)\Big\} \leqslant$$
$$\frac{C}{\lambda}\sup_t |\widehat{f}_{z,h,\tau}(t)| \leqslant \frac{Ch}{\lambda}.$$

Now applying Lemma 1.3 we get

$$\int_{-\infty}^{\infty} g(t)\widehat{D}(t)\,\mathrm{d}t \leqslant \frac{C}{\lambda}\,Q(D,h) = \frac{C}{\lambda}\,\gamma_h.$$

□.

Remark 12.2 *There are other choices of γ_h. For example. it is possible to prove Lemma 12.8 with $\tilde{\gamma}_h = Q(\exp\{\lambda(F-E)\},h)$. However then the estimate is $C\lambda^{-1}\tilde{\gamma}_h^{1/2}$.*

Step 3. *Directly estimating measure concentrated outside* $[-\tau,\tau]$. Now we begin estimation of the second part in (12.38). Taking into account Lemma 12.2, two estimates for it are needed. We begin from the easier part.

Lemma 12.9 *For all* $0 < \tau \leqslant h/2$

$$|\, r(V-E)\exp\{\lambda(F-E)\}\,|_{h,\tau} \leqslant Cr\gamma_h. \tag{12.44}$$

Proof. By (12.3) we have

$$|\, r(V-E)\exp\{\lambda(F-E)\}\,|_{h,\tau} \leqslant Cr\|\, V-E\,\|\,|\exp\{\lambda(F-E)\}\,|_{h,\tau} \leqslant$$
$$Cr|\exp\{\lambda(F-E)\}\,|_{h,\tau} \leqslant Cr\|\exp\{\lambda(F-E)/2\}\,\|\,|\, D\,|_{h,\tau} \leqslant Cr|\, D\,|_{h,\tau}.$$

We recall, that $0 \leqslant f_{z,h,\tau}(x) \leqslant 1$. Therefore

$$|\, D\,|_{h,\tau} \leqslant \sup_z \int_z^{z+h} f_{z,h,\tau}(x)\, D\{\mathrm{d}x\} \leqslant \sup_z \int_z^{z+h} D\{\mathrm{d}x\} = \sup_z D\{[z, z+h]\} = Q(D,h) = \gamma_h.$$

The last estimate completes the proof of (12.44). □

Step 4. *Second estimate for measure concentrated outside* $[-\tau,\tau]$. Now comes the tricky part. Set

$$\nu_\tau(y) = sW\big\{\{x : |\, x\,| > y\}\big\}. \tag{12.45}$$

Lemma 12.10 *For all* $0 < \tau \leqslant h/2$

$$|\, r(V-E)\exp\{\lambda(F-E)\}\,|_{h,\tau} \leqslant \frac{C}{\lambda}(|\ln\gamma_\tau\,| + 1)^3(\lambda\nu_\tau(h/4))^{-1/2}. \tag{12.46}$$

Proof. By (12.3) we have

$$|\, r(V-E)\exp\{\lambda(F-E)\}\,|_{h,\tau} =$$
$$|\, r(V-E)\exp\{s\lambda(W-E) + \lambda d(\Psi - E) + \lambda q(A-E)\}\,|_{h,\tau} \leqslant$$
$$|\, r(V-E)\exp\{s\lambda(W-E)\}\,|_{h,\tau}\,\|\exp\{\lambda d(\Psi-E) + \lambda q(A-E)\}\,\| =$$
$$|\, r(V-E)\exp\{s\lambda(W-E)\}\,|_{h,\tau} = |\,(s(W-E) + d(\Psi-E))\exp\{s\lambda(W-E)\}\,|_{h,\tau} \leqslant$$
$$|\, s(W-E)\exp\{s\lambda(W-E)\}\,|_{h,\tau} + Cd|\exp\{s\lambda(W-E)\}\,|_{h,\tau}. \tag{12.47}$$

We begin from the estimate of the second summand. Because $0 \leqslant f_{z,h,\tau}(x) \leqslant 1$ we get

$$|\exp\{s\lambda(W-E)\}\,|_{h,\tau} \leqslant Q(\exp\{s\lambda(W-E)\}, h) \leqslant 4Q(\exp\{s\lambda(W-E)\}, h/4).$$

By the property of concentration functions (1.30) we get

$$Q(\exp\{s\lambda(W-E)\}, h/4) \leqslant \frac{C}{\sqrt{s\lambda W\{\{x : |\, x\,| > h/4\}\}}} = \frac{C}{\sqrt{\lambda\nu_\tau(h/4)}}. \tag{12.48}$$

We decomposed F in such a way that d was small. Combining the last estimate with (12.37) we get

$$d|\exp\{s\lambda(W-E)\}\,|_{h,\tau} \leqslant \frac{C}{\lambda}(|\ln\gamma_\tau\,| + 1)^3(\lambda\nu_\tau(h/4))^{-1/2}. \tag{12.49}$$

Thus, it remains to estimate the first summand in (12.47). First, let us assume that $s\lambda < 2$. Then we can take into account that $0 \leqslant f_{z,h,\tau}(x) \leqslant 1$ and (12.3) and obtain

$$| \, s(W-E)\exp\{s\lambda(W-E)\} \, |_{h,\tau} \leqslant$$

$$\| \, W-E \, \| sQ(\exp\{s\lambda(W-E)\}, h) \leqslant CsQ(\exp\{s\lambda(W-E)\}, h/4) \leqslant$$

$$Cs(s\lambda)^{-1}Q(\exp\{s\lambda(W-E)\}, h/4) \leqslant \frac{C}{\lambda}\, Q(\exp\{s\lambda(W-E)\}, h/4).$$

By the Le Cam inequality (1.29):

$$Q(\exp\{s\lambda(W-E)\}, h/4) \leqslant \frac{C}{\sqrt{s\lambda W\{\{x : |\, x \,| > h/4\}\}}} = \frac{C}{\sqrt{\lambda\nu_\tau(h/4)}}. \tag{12.50}$$

Therefore, for $\lambda s < 2$,

$$| \, s(W-E)\exp\{s\lambda(W-E)\} \, |_{h,\tau} \leqslant \frac{C}{\lambda}\, (\lambda\nu_\tau(h/4))^{-1/2}. \tag{12.51}$$

Now let us assume that $\lambda s \geqslant 2$. We shall employ the fact that W is concentrated on $[K_1(\mathbf{u})]_\tau \setminus [-\tau, \tau]$ and, consequently, we can apply (12.23) with suitable $\omega(x)$. We choose ω to be defined as in Lemma 12.4 with $\mathbf{u}$ and ℓ as in above and with $m = 4[s\lambda] + 1$. Here $[\cdot]$ means the integral part. Then, just like in (12.23), we get

$$| \, s(W-E)\exp\{s\lambda(W-E)\} \, |_{h,\tau} \leqslant \int\limits_{-\infty}^{\infty} \omega(x)\, s(W-E)\exp\{s\lambda(W-E)\}\{\mathrm{d}x\} +$$

$$\int\limits_{\mathbb{R}^*} (f_{z,h,\tau}(x) - \omega(x))\, s(W-E)\exp\{s\lambda(W-E)\}\{\mathrm{d}x\} = J_1 + J_2. \tag{12.52}$$

Here $\mathbb{R}^* = \mathbb{R} \setminus [K_m(\mathbf{u})]_{m\tau}$. Let us estimate J_1. We shall apply Parseval's identity and the properties of $\omega(x)$ (12.18)–(12.22). We have

$$| \, J_1 \, | \leqslant C \int\limits_{-\infty}^{\infty} | \, \widehat{\omega}(t) \, | s(1 - \widehat{W}(t))\exp\{s\lambda(\widehat{W}(t) - 1)\}\, \mathrm{d}t \leqslant$$

$$\frac{C}{\lambda} \int\limits_{-\infty}^{\infty} | \, \widehat{\omega}(t) \, | \exp\Big\{\frac{s\lambda}{2}(\widehat{W}(t) - 1)\Big\}\, \mathrm{d}t \leqslant \frac{C}{\lambda}\, Q\Big(\exp\Big\{\frac{s\lambda}{2}(W-E)\Big\}, h\Big)\ell^2 \ln(\ell m + 1).$$

Now, $Q(\cdot, h)$ is already estimated in (12.48). Moreover, ℓ satisfies estimates (12.27). Therefore,

$$| \, J_1 \, | \leqslant \frac{C}{\lambda}\, (\lambda\nu_\tau(h/4))^{-1/2}(| \ln\gamma_\tau \,| + 1)^2(| \ln\gamma_\tau \,| + 1 + \ln m).$$

For estimate of $\ln m$, we note that by the Le Cam inequality (1.29) and because of the support of W:

$$\gamma_\tau = Q(\exp\{(\lambda q(A-E) + \lambda s(W-E) + d(\Psi - E))/2\}, \tau) \leqslant$$

$$Q(\exp\{\lambda s(W-E)/2\},\tau) \leqslant \frac{C}{\sqrt{\lambda s W\{\{x : |\,x\,| > \tau\}\}}} = \frac{C}{\sqrt{s\lambda}}.$$

Consequently, noting that $\gamma_\tau \leqslant 1$ (because it is concentration function), we get

$$\gamma_\tau \leqslant C(\lambda s)^{-1/2}, \quad \lambda s \leqslant C\gamma_\tau^{-1/2}, \quad m \leqslant 4\lambda s + 1 \leqslant C\gamma_\tau^{-1/2}.$$

Therefore,

$$\ln m \leqslant C(|\,\ln \gamma_\tau\,| + 1)$$

and

$$|\,J_1\,| \leqslant \frac{C}{\lambda}\,(\lambda \nu_\tau(h/4))^{-1/2}(|\,\ln \gamma_\tau\,| + 1)^3. \tag{12.53}$$

Now, let us return to (12.52) and estimate J_2. We shall use the following facts

1. W is concentrated on the set $[K_1(\mathbf{u})]_\tau$.
2. W^k is concentrated on the set $[K_k(\mathbf{u})]_{k\tau}$.
3. $f_{z,h,\tau}(x) - \omega(x) = 0$, for $x \in [K_m(\mathbf{u})]_{m\tau}$.
4. $0 \leqslant f_{z,h,\tau}(x) - \omega(x) \leqslant 1$.

Now we can employ the exponential structure of measure

$$\exp\{s\lambda(W-E)\} = \sum_{k \leqslant 4\lambda s} \frac{(s\lambda)^k W^k}{k!}\mathrm{e}^{-s\lambda} + \sum_{k > 4\lambda s} \frac{(s\lambda)^k W^k}{k!}\mathrm{e}^{-s\lambda}.$$

Consequently,

$$J_2 = \sum_{k \geqslant 0} \frac{(s\lambda)^k}{k!}\mathrm{e}^{-s\lambda} s \int\limits_{\mathbb{R}^*} (f_{z,h,\tau}(x) - \omega(x))\,(W-E)W^k\{\mathrm{d}x\} =$$

$$\sum_{k \geqslant m-1} \frac{(s\lambda)^k}{k!}\mathrm{e}^{-s\lambda} s \int\limits_{\mathbb{R}^*} (f_{z,h,\tau}(x) - \omega(x))\,(W-E)W^k\{\mathrm{d}x\}.$$

Therefore

$$|\,J_2\,| \leqslant \sum_{k > 3\lambda s} \frac{(s\lambda)^k}{k!}\mathrm{e}^{-s\lambda} s \|\,(W-E)W^k\,\| \leqslant 2 \sum_{k > 3\lambda s} \frac{(s\lambda)^k}{k!}\mathrm{e}^{-s\lambda} s \leqslant$$

$$2s\mathrm{e}^{-3\lambda s} \sum_{k > 3\lambda s} \frac{(\mathrm{e}s\lambda)^k}{k!}\mathrm{e}^{-s\lambda} \leqslant 2s\exp\{(\mathrm{e}-4)\lambda s\} \leqslant 2s\exp\{-\lambda s\} \leqslant$$

$$\frac{C}{\lambda}\exp\{-\lambda s/2\} \leqslant \frac{C}{\lambda}\frac{1}{\sqrt{s\lambda}} \leqslant \frac{C}{\lambda}\,(\lambda \nu_\tau(h/4))^{-1/2}. \tag{12.54}$$

Collecting estimates (12.49), (12.51), (12.53) and (12.54) we complete the proof of Lemma 12.10. □

Step 5. *Collecting all in one place.* Now we can make the final step. Let us collect the estimates of Lemmas 12.8 –12.10.

Lemma 12.11 *For all* $0 < \tau \leqslant h/2$

$$|\,(F-E)\exp\{\lambda(F-E)\}\,|_{h,\tau} \leqslant \frac{C}{\lambda}\big(\gamma_h + \lambda r \gamma_h\big)$$

and

$$|\,(F-E)\exp\{\lambda(F-E)\}\,|_{h,\tau} \leqslant \frac{C}{\lambda}\big(\gamma_h + (|\,\ln\gamma_\tau\,|+1)^3(\lambda\nu_\tau(h/4))^{-1/2}\big).$$

Now it remains to check that Lemma 12.11 has the same structure as Lemma 12.2. Indeed, it suffices to take

$$H_1 = \lambda F, \quad H_{\tau 2} = \lambda s W.$$

Indeed, then

$$H_{\tau 2}\{\{x : |\,x\,| > h/4\}\} = \lambda\nu_\tau(h/4)$$

and

$$\lambda r = \lambda F\big\{\{x : |\,x\,| > \tau\}\big\} = H_1\big\{\{x : |\,x\,| > \tau\}\big\}.$$

Moreover, F is symmetric and $F = F^{(-)}$. Therefore, to end the proof of (12.31) it suffices to collect the powers of γ_h and $(|\,\ln\gamma_\tau\,|+1)$.

12.4 Second example

Let $F \in \mathcal{F}_+$, $N \leqslant C$ and n be natural numbers. Then

$$|\,F^n(F-E)^N\,| \leqslant Cn^{-N}. \tag{12.55}$$

In comparison to the previous example, we introduce two new moments: first, we have N'th convolution of $(F-E)$; second, the main smoothing distribution is not compound Poisson one. The triangle function method is adapted to fit these changes. Again we present the proof in blocks.

Step 0. *Preliminary investigation.* First we check what estimate can be obtained for the Fourier-Stieltjes transform. Taking into account that $\widehat{F}(t) \leqslant \exp\{\widehat{F}(t)-1\}$ we get

$$|\,\widehat{F}(t)^n(\widehat{F}(t)-1)^N\,| \leqslant |\,(\widehat{F}(t)-1)^N\exp\{n(\widehat{F}(t)-1)\}\,| \leqslant Cn^{-N}\exp\Big\{\frac{n}{2}\,(\widehat{F}(t)-1)\Big\}.$$

Thus, the Fourier-Stieltjes transform is of the type (12.5) with

$$\widehat{D}(t) = \exp\Big\{\frac{n}{2}(\widehat{F}(t)-1)\Big\}. \tag{12.56}$$

As usual, when the triangle function method is applied, we prove more general result than (12.55), i.e. we prove that for any $h > 0$

$$|\,F^n(F-E)^N\,|_h \leqslant Cn^{-N}\gamma_h^{1/(2N+1)}(|\,\ln\gamma_h\,|+1)^{6N(N+1)/(2N+1)}. \tag{12.57}$$

Here

$$\gamma_h = Q(D, h) = Q\biggl(\exp\Bigl\{\frac{n}{2}\,(F-E)\Bigr\}, h\biggr).$$

Taking supremum over all $h > 0$ we get (12.55).

Step 1. *Decomposition of F.* We decompose F exactly as in previous example. For convenience we repeat that decomposition. For any $\tau > 0$, there exists $\mathbf{u} \in \mathbb{R}^\ell$ such that

$$\ell \leqslant C(|\ln \gamma_\tau| + 1), \tag{12.58}$$

$$F\{\mathbb{R} \setminus [K_1(\mathbf{u})]_\tau\} \leqslant \frac{C}{n}\bigl(|\ln \gamma_\tau| + 1\bigr)^3. \tag{12.59}$$

Moreover,

$$F = qA + sW + d\Psi = qA + rV,$$
$$A\{[-\tau,\tau]\} = W\{[K_1(\mathbf{u})]_\tau \setminus [-\tau,\tau]\} = \Psi\{\mathbb{R} \setminus [K_1(\mathbf{u})]_\tau\} = 1. \tag{12.60}$$

Here

$$q = F\{[-\tau,\tau]\}, \quad s = F\{[K_1(\mathbf{u})]_\tau \setminus [-\tau,\tau]\}, \quad d = F\{\mathbb{R} \setminus [K_1(U)]_\tau\}.$$

Though we do not know what properties has Ψ its weight is small:

$$d \leqslant \frac{C}{n}(|\ln \gamma_\tau| + 1)^3. \tag{12.61}$$

Step 2. *Estimating measures containing* $(A-E)$. The general idea is to consider measures containing factor $(A-E)$ applying Parseval's identity. For the present example, this will require more elaborated approach. In the previous case, we could use the property of compound Poisson distribution allowing easily separate two components

$$\exp\{F-E\} = \exp\{q(A-E)\}\exp\{r(V-E)\}.$$

However, for F^n we need the different approach. We shall replace F^n by $(qA+rE)^n(qE+rV)^n$. Note that

$$(qA+rE)(qE+rV) = (E+q(A-E))(E+r(V-E)) = E+q(A-E)+r(V-E),$$

$$F = qA+rV = E+q(A-E)+r(V-E), \quad F-(qA+rE)(qE+rV) = -q(A-e)r(V-E).$$

Set

$$\Delta_1 = (F-E)^N\bigl\{F^n - (qA+rE)^n(qE+rV)^n\bigr\}.$$

Note, that we investigate symmetric distributions, i.e. having real characteristic functions. Therefore

$$q\widehat{A}(t) + r \geqslant q\widehat{A}(t) + r\widehat{V}(t) = \widehat{F}(t) \geqslant 0.$$

Consequently,

$$q\widehat{A}(t) + r = 1 + q(\widehat{A}(t) - 1) \leqslant \exp\{q(\widehat{A}(t)-1)\}$$

and, similarly,

$$q + r\widehat{V}(t) \leqslant \exp\{r(\widehat{V}(t)-1)\}.$$

Lemma 12.12 *For all* $0 < \tau \leqslant h/2$

$$|\,\Delta_1\,|_{h,\tau} \leqslant Cn^{-N}\,\gamma_h.$$

Proof. We begin from the estimate of $\widehat{\Delta}_1(t)$. Note that

$$|\,(q\widehat{A}(t)+r)(q+r\widehat{V}(t))\,| \leqslant \exp\{q(\widehat{A}(t)-1)+r(\widehat{V}(t)-1)\} = \exp\{\widehat{F}(t)-1\}.$$

Therefore, by (1.9)

$$|\,\widehat{\Delta}_1(t)\,| \leqslant |\,\widehat{F}(t)-1\,|^N n\exp\{(n-1)(\widehat{F}(t)-1)\}q|\,\widehat{A}(t)-1\,|r|\,\widehat{V}(t)-1\,| \leqslant$$
$$Cn|\,\widehat{F}(t)-1\,|^N\exp\{n(\widehat{F}(t)-1)/4\}r|\,\widehat{V}(t)-1\,|\exp\{nr(\widehat{V}(t)-1)/4\}q|\,\widehat{A}(t)-1\,|\widehat{D}(t) \leqslant$$
$$Cn^{-N}q|\,\widehat{A}(t)-1\,|\widehat{D}(t) = Cg(t)\widehat{D}(t).$$

Applying Parseval's identity, we get

$$|\,\Delta_1\,|_{h,\tau} \leqslant C\int_{-\infty}^{\infty} g(t)\widehat{D}(t)\,\mathrm{d}t, \tag{12.62}$$

The following steps are identical to those of Lemma 12.8. Estimating $g(t)$ we distinguish between two cases: a) when t is near zero; b) when t is far from zero, and obtain

$$g(t) \leqslant Cn^{-N}\begin{cases} g_1(t), & \text{if } |\,t\,| \leqslant 1/\tau,\\ g_2(t), & \text{if } |\,t\,| > 1/\tau.\end{cases}$$

Here

$$g_1(t) = \sqrt{nq\sigma^2}\exp\Big\{-\frac{qn\sigma^2t^2}{12}\Big\},\quad g_2(t) = C\,\frac{\tau}{1+(t\tau)^2} \tag{12.63}$$

and

$$\sigma^2 = \int x^2\,A\{dx\}\,.$$

Note that $g_1(t)$ and $g_2(t)$ are even and vanishing as $|\,t\,| \to \infty$. Therefore

$$\sup_{s:|s|\geqslant|t|} |\,g(s)\,| \leqslant Cn^{-N}\,\big(g_1(t)+g_2(t)\big)$$

and

$$\int_{-\infty}^{\infty}\sup_{s:|s|\geqslant|t|}|\,g(s)\,|\,\mathrm{d}t \leqslant Cn^{-N}\int_{-\infty}^{\infty}\big(g_1(t)+g_2(t)\big)\,\mathrm{d}t \leqslant Cn^{-N}.$$

Moreover, by (12.14)

$$\sup_t g(t) \leqslant Cn^{-N}\sup_t|\,\widehat{f}_{z,h,\tau}(t)\,| \leqslant Cn^{-N}h.$$

Now applying Lemma 1.3 we get

$$\int_{-\infty}^{\infty} g(t)\widehat{D}(t)\,\mathrm{d}t \leqslant Cn^{-N}\gamma_h.$$

$\square$.

Remark 12.3 *Note that, in principle, we could obtain the estimate of the order* $O(n^{-N-1})$.

Now we decompose

$$(F-E)^N = r^N (V-E)^N + \sum_{j=1}^{N} \binom{N}{j} q^j (A-E)^j r^{N-j} (V-E)^{N-j}.$$

Set

$$\Delta_2 = (qA+rE)^n (qE+rV)^n \{(F-E)^N - r^N (V-E)^N\} =$$

$$(qA+rE)^n (qE+rV)^n \sum_{j=1}^{N} \binom{N}{j} q^j (A-E)^j r^{N-j} (V-E)^{N-j}.$$

Lemma 12.13 *For all* $0 < \tau \leqslant h/2$

$$|\, \Delta_2 \,|_{h,\tau} \leqslant C n^{-N} \gamma_h.$$

Proof. The proof of Lemma 12.13 is similar to the proof of Lemma 12.12. We note that

$$|\, \widehat{\Delta}_2(t) \,| \leqslant q |\, \widehat{A}(t) - 1 \,| \sum_{j=1}^{N} \binom{N}{j} q^{j-1} |\, \widehat{A}(t) - 1 \,|^{j-1} r^{N-j} |\, \widehat{V}(t) - 1 \,|^{N-j} \widehat{D}(t)^2 \leqslant$$

$$C n^{-N+1} q |\, \widehat{A}(t) - 1 \,| \widehat{D}(t)^{3/2} = C g(t) \widehat{D}(t).$$

The following proof is identical to the previous one and is omitted. $\square$

Step 3. *Directly estimating measure concentrated outside* $[-\tau, \tau]$. The first step in estimating measures concentrated outside finite interval is straightforward. Set

$$\Delta_3 = (qA+rE)^n (qE+rV)^n r^N (V-E)^N.$$

Lemma 12.14 *For all* $0 < \tau \leqslant h/2$

$$|\, \Delta_3 \,|_{h,\tau} \leqslant C r^N \gamma_h. \tag{12.64}$$

Proof. By (12.3) and $0 \leqslant f_{z,h,\tau}(x) \leqslant 1$we obtain

$$|\, \Delta_3 \,|_{h,\tau} r \leqslant C r^N |\, (qA+rE)^n (qE+rV)^n \,|_{h,\tau} \leqslant C r^N Q((qA+rE)^n (qE+rV)^n, h).$$

For the proof that Q can be majorized by γ_h we apply the properties of concentration functions (1.31) - (1.32):

$$Q((qA+rE)^n (qE+rV)^n, h) \leqslant Ch \int\limits_{|t|<1/h} (q\widehat{A}(t) + r)^n (q + r\widehat{V}(t))^n \,\mathrm{d}t \leqslant$$

$$Ch \int\limits_{|t|<1/h} \exp\{n(q(\widehat{A}(t)-1)+r(\widehat{V}(t)-1))\}\,\mathrm{d}t = Ch \int\limits_{|t|<1/h} \exp\{n(\widehat{F}(t)-1)\}\,\mathrm{d}t \leqslant$$

$$Ch \int\limits_{|t|<1/h} \exp\{n(\widehat{F}(t)-1)/2\}\,\mathrm{d}t \leqslant C\gamma_h.$$

The last estimate completes the proof of (12.64). □

Step 4. *Second estimate for measure concentrated outside* $[-\tau,\tau]$. In principle, all ideas are the same as in the first example. Essentially we should

1. Reduce compound measures containing A. We already separated such measures as convolution factors and, usually it suffices to replace them by their norms.

2. Reduce Ψ. We expand compound measures in powers of Ψ and replace them by $\|\Psi\|=1$.

Just like in the proof of the first example set

$$\nu_\tau(y) = sW\big\{\{x : |x| > y\}\big\}. \tag{12.65}$$

Lemma 12.15 *For all* $0<\tau\leqslant h/2$

$$|\Delta_3|_{h,\tau} \leqslant Cn^{-N}(|\ln\gamma_\tau|+1)^{3+3N}(n\nu_\tau(h/4))^{-1/2}. \tag{12.66}$$

Proof. The first essential step is to check that the proof is needed for small d only. Indeed, let $d\geqslant 1/3$. Then, exactly as in the proof of lemma 12.14

$$|\Delta_3|_{h,\tau} \leqslant C\gamma_h \leqslant Cd\gamma_h.$$

But, by (12.61) $d\leqslant C(|\ln\gamma_\tau|+1)^3$ and, as proved in Lemma 12.10

$$\gamma_h \leqslant CQ(\exp\{ns(W-E)/2\}, h/4) \leqslant C(n\nu_\tau(h/4)^{-1/2}.$$

Thus, it remains to prove Lemma for $d<1/3$. We recall that $\|G\|=1$, for any $G\in\mathcal{F}$. Therefore, $\|(qA+rE)^n\|=1$, $\|\Psi-E\|\leqslant 2$ and by (12.3) we have

$$|\Delta_3|_{h,\tau} \leqslant C|r^N(V-E)^N(qE+rV)^n|_{h,\tau} = C|(s(W-E)+d(\Psi-E))^N(qE+rV)^n|_{h,\tau} \leqslant$$

$$C\sum_{k=0}^{N}\binom{N}{k}d^{N-k}|s^k(W-E)^k(qE+rV)^n|_{h,\tau}.$$

We can define $U\in\mathcal{F}$ by

$$qE+rV = qE+sW+d\Psi = (1-d)U+d\Psi, \quad U = \frac{q}{1-d}E+\frac{s}{1-d}W. \tag{12.67}$$

Now once again applying (12.3) we get

$$| s^k(W-E)^k(qE+rV)^n |_{h,\tau} \leqslant \sum_{j=0}^{n} \binom{n}{j} d^{n-j}(1-d)^j | s^k(W-E)^k U^j |_{h,\tau}.$$

Collecting the last estimates we see, that

$$| \Delta_3 |_{h,\tau} \leqslant \sum_{k=0}^{N} \binom{N}{k} d^{N-k} \sum_{j=0}^{n} \binom{n}{j} d^{n-j}(1-d)^j | s^k(W-E)^k U^j |_{h,\tau}. \tag{12.68}$$

For the sake of brevity set

$$\Delta_4 = s^k(W-E)^k U^j.$$

Let $j > 0$. We choose ω to be defined as in Lemma 12.4 with $\mathbf{u}$ and ℓ as in above and with $m = [6sn] + N + 1$. Here $[\cdot]$ means the integral part. Then, just like in (12.23), we get

$$| \Delta_4 |_{h,\tau} \leqslant \int_{-\infty}^{\infty} \omega(x)\, \Delta_4\{\mathrm{d}x\} + \int_{\mathbb{R}^*} (f_{z,h,\tau}(x) - \omega(x))\, \Delta_4\{\mathrm{d}x\} = I_1 + I_2. \tag{12.69}$$

Here $\mathbb{R}^* = \mathbb{R} \setminus [K_m(\mathbf{u})]_{m\tau}$. Let us estimate I_1. Applying Parseval's identity and the properties of $\omega(x)$ (12.18)–(12.22) we obtain

$$| I_1 | \leqslant C \int_{-\infty}^{\infty} | \widehat{\omega}(t) | s^k(1-\widehat{W}(t))^k | \widehat{U}(t) |^j \,\mathrm{d}t.$$

Now we make use of the fact that $d < 1/3$. Then

$$q + s\widehat{W}(t) = a + s\widehat{W}(t) + d - d \geqslant q\widehat{A}(t) + s\widehat{W}(t) + d\widehat{\Psi}(t) - d \geqslant \widehat{F}(t) - 1/3 \geqslant -1/3.$$

Therefore,

$$\widehat{U}(t) \geqslant -\frac{1}{3(1-d)} \geqslant -\frac{1}{2}$$

and, by (12.30),

$$| \widehat{U}(t) | \leqslant \exp\Big\{\frac{s}{9}\,(\widehat{W}(t) - 1)\Big\}.$$

Consequently,

$$s^k(1-\widehat{W}(t)) | \widehat{U}(t) |^j \leqslant C(k) j^{-k} \exp\Big\{\frac{js}{12}\,(\widehat{W}(t) - 1)\Big\}.$$

Taking into account the last estimates and (12.22) we get

$$| I_1 | \leqslant C(k) j^{-k} \int_{-\infty}^{\infty} | \widehat{\omega}(t) | \exp\Big\{\frac{s}{12}\,(\widehat{W}(t) - 1)\Big\}\,\mathrm{d}t \leqslant$$

$$C(k) j^{-k} Q(\exp\{sj(W-E)/12\}, h)\ell^2 \ln(\ell m + 1).$$

All factors are already estimated in the first example. Note also that $k \leqslant N \leqslant C$. Collecting those estimates we finally get

$$|\, I_1 \,| \leqslant Cj^{-k}(|\, \ln \gamma_\tau \,| + 1)^3 Q(\exp\{sj(W-E)/12\}, h). \tag{12.70}$$

Now we estimate I_2. Let $ns \geqslant 2$. We have

$$|\, I_2 \,| \leqslant \sum_{l=0}^{j} \binom{j}{l} \left(\frac{s}{1-d}\right)^l \left(\frac{q}{1-d}\right)^{j-l} \int_{\mathbb{R}^*} (f_{z,h,\tau}(x) - \omega(x)) \,|\, s^k (W-E)^k W^l \,|\{\mathrm{d}x\}$$

Due to the choice of m, we get

$$|\, I_2 \,| \leqslant s^k \sum_{l \geqslant 3ns} \binom{j}{l} \left(\frac{s}{1-d}\right)^l \left(\frac{q}{1-d}\right)^{j-l}. \tag{12.71}$$

For the estimate of the last sum we employ probabilistic interpretation and Chebyshev's inequality. Let $S_j = \xi_1 + \cdots + \xi_j$, where ξ_j are independent Bernoulli variables taking 0 with $q/(1-d)$. It is not difficult to note that the right-hand side of (12.71) is equal to

$$s^k P(S_j \geqslant 6ns) \leqslant s^k \mathrm{e}^{-6ns} \mathbb{E} \exp\{S_j\} =$$

$$s^k \mathrm{e}^{-6ns} \left(\frac{q}{1-d} + \frac{s\mathrm{e}}{1-d}\right)^j \leqslant Cs^k \exp\{-6ns + 3sj\} \leqslant Cn^{-k}(ns)^{-1/2} \leqslant (n\nu_\tau(h/4))^{-1/2}.$$

Therefore,

$$|\, I_2 \,| \leqslant Cn^{-k}(ns)^{-1/2} \leqslant (n\nu_\tau(h/4))^{-1/2}.$$

Just like in the previous example we prove that the estimate also holds for $ns < 2$. Collecting estimates of I_1 and I_2 we get

$$|\, \Delta_4 \,|_{h,\tau} \leqslant Cj^{-k}(|\, \ln \gamma_\tau \,| + 1)^3 (j\nu_\tau(h/4))^{-1/2} + Cn^{-k}(ns)^{-1/2} \leqslant (n\nu_\tau(h/4))^{-1/2}. \tag{12.72}$$

The next step is to put (12.72) into (12.68). We get

$$|\, \Delta_3 \,|_{h,\tau} \leqslant \sum_{k=0}^{N} \binom{N}{k} d^{N-k} \sum_{j=0}^{n} \binom{n}{j} d^{n-j}(1-d)^j n^{-k} (n\nu_\tau(h/4))^{-1/2} +$$

$$\sum_{k=0}^{N} \binom{N}{k} d^{N-k} \left\{ \sum_{j=1}^{n} \binom{n}{j} d^{n-j}(1-d)^j j^{-k} (|\, \ln \gamma_\tau \,| + 1)^3 Q(\exp\{sj(W-E)/12\}, h) + d^n \right\} =$$

$$I_3 + I_4. \tag{12.73}$$

It is not difficult to estimate I_3. Indeed by (12.61)

$$I_3 \leqslant (n\nu_\tau(h/4))^{-1/2}(|\, \ln \gamma_\tau \,| + 1)^{3N} n^{-N} \sum_{k=0}^{N} \binom{N}{k} \leqslant$$

$$Cn^{-N}(|\, \ln \gamma_\tau \,| + 1)^{3N} (n\nu_\tau(h/4))^{-1/2}. \tag{12.74}$$

The estimate of I_4 is much more complicated. We know that d is small. Therefore, d^n also is small. However, we need $n\nu_\tau(h/4)$ and, consequently, more elaborated approach. We have

$$I_4 \leqslant C(N)(|\ln\gamma_\tau|+1)^3 \sum_{k=0}^{N} d^{N-k}\Big\{\sum_{j=1}^{n}\binom{n}{j} d^{n-j}(1-d)^j j^{-k} Q(\exp\{sj(W-E)/12\},h)+d^n\Big\}.$$

For the expression in brackets, we apply (1.31):

$$\sum_{j=1}^{n}\binom{n}{j} d^{n-j}(1-d)^j j^{-k} Q(\exp\{sj(W-E)/12\},h)+d^n \leqslant$$

$$C\sum_{j=1}^{n}\binom{n}{j} d^{n-j}(1-d)^j j^{-k} h\int_{-1/h}^{1/h} \exp\{sj(\widehat{W}(t)-1)/12\}\,\mathrm{d}t + d^n \leqslant$$

$$Ch\int_{-1/h}^{1/h}\Big\{\sum_{j=1}^{n}\binom{n}{j} d^{n-j}(1-d)^j j^{-k}\exp\{sj(\widehat{W}(t)-1)/12\}+d^n\Big\}\,\mathrm{d}t.$$

Applying Hölder's inequality we get

$$\Big\{\sum_{j=1}^{n}\binom{n}{j} d^{n-j}(1-d)^j j^{-k}\exp\{sj(\widehat{W}(t)-1)/12\}+d^n\Big\}^2 \leqslant \Big\{\sum_{j=1}^{n}\binom{n}{j} d^{n-j}(1-d)^j j^{-2k}+d^n\Big\}\times$$

$$\Big\{\sum_{j=1}^{n}\binom{n}{j} d^{n-j}(1-d)^j \exp\{sj(\widehat{W}(t)-1)/6\}+d^n\Big\} = \tilde{S}S.$$

Note that

$$\binom{n}{j}\frac{1}{(j+1)(j+2)\dots(j+2k)} = \binom{n+2k}{j+2k}\frac{1}{(n+1)(n+2)\dots(n+2k)}.$$

We recall that $d<1/3$. Therefore, we directly prove that

$$\tilde{S} \leqslant C(n(1-d))^{-2k} \leqslant Cn^{-2k}.$$

On the other hand,

$$S = \big((1-d)\exp\{s(\widehat{W}(t)-1)/6\}+d\big)^n \leqslant \exp\{(1-d)(\exp\{sn(\widehat{W}(t)-1)/6\}-1)\}.$$

Applying inequality

$$\mathrm{e}^{-x}-1 \leqslant -x+\frac{x^2}{2} = -x\Big(1-\frac{x}{2}\Big), \quad x>0,$$

we get

$$\exp\{s(\widehat{W}(t)-1)/6\}-1 \leqslant \frac{s(\widehat{W}(t)-1)}{6}\Big(1-\frac{s(1-\widehat{W}(t))}{12}\Big) \leqslant$$

$$\frac{5s}{36}(\widehat{W}(t)-1).$$

Therefore,

$$S \leqslant \exp\{10sn(\widehat{W}(t)-1)/118\}.$$

Collecting all estimates and consequently applying (1.32), (12.61) and (12.50) we get

$$I_4 \leqslant C(|\ln\gamma_\tau|+1)^3\sum_{k=0}^{N} d^{N-k}n^{-k}Ch\int\limits_{-1/h}^{1/h}\exp\{5sn(\widehat{W}(t)-1)/118\}\,dt \leqslant$$

$$C(|\ln\gamma_\tau|+1)^3 Q(\exp\{5sn(W-E)/118\},h)\sum_{k=0}^{N} d^{N-k}n^{-k} \leqslant$$
$$Cn^{-N}(|\ln\gamma_\tau|+1)^{3+3N}Q(\exp\{5sn(W-E)/118\},h) \leqslant$$
$$Cn^{-N}(|\ln\gamma_\tau|+1)^{3+3N}(n\nu_\tau(h/4))^{-1/2}.$$

From the last estimate, (12.74) and (12.73) we get the statement of Lemma. □

Step 5. *Collecting all in one place.* Now we can make the final step. Let us collect the estimates of Lemmas 12.12 –12.15.

Lemma 12.16 *For all* $0<\tau\leqslant h/2$

$$|\,F^n(F-E)^N\,|_{h,\tau} \leqslant Cn^{-N}\big(\gamma_h+(nr)^N\gamma_h\big)$$

and

$$|\,F^n(F-E)^N\,|_{h,\tau} \leqslant Cn^{-N}\big(\gamma_h+(|\ln\gamma_\tau|+1)^{3+3N}(n\nu_\tau(h/4))^{-1/2}\big).$$

The last part of the proof coincides with the one from the first example. Indeed, it suffices to take

$$H_1 = nF, \quad H_{\tau 2} = nsW.$$

□

Exercises

1. To prove Lemma 12.13.

2. Let F be symmetric,$\lambda>0$ and let $N\leqslant C$ be natural number. To prove that

$$|\,(F-E)^N\exp\{\lambda(F-E)\}\,| \leqslant C\lambda^{-N}.$$

Bibliographical notes

The triangle function method was introduced by Arak (1980). Zaĭtsev (1989) extended method to the multivariate case. Almost all lemmas can be found in Arak and Zaĭtsev (1988). Lemmas 12.1 and 12.2 are slightly different versions of Lemmas 3.1 (p.67) and 3.2 (p. 68) from the mentioned book. Example 1 was proved by Presman (1981). Example 2 is a partial case of the result proved in Čekanavičius (1989). Other examples of application of the triangle function method can be found in Čekanavičius (1995, 1997).

13 Combining various techniques

In Section 2 we already demonstrated, how to use combinatorics and estimates for total variation. Here we give more examples.

Example 1

In the previous Section, we proved that, for $F \in \mathcal{F}_+$, $s \leqslant C$ and n natural

$$|\,F^n(F-E)^s\,| \leqslant C(s)n^{-s}. \tag{13.1}$$

As it turns out, from (13.1) we can obtain many other estimates. For example, let $\lambda > 0$, and let $s \leqslant C$ and n be natural numbers, $F \in \mathcal{F}_+$. Then

$$|\,(F-E)^s \exp\{\lambda(F-E)\}\,| \leqslant C(s)\lambda^{-s}. \tag{13.2}$$

Proof. From the definition of the exponential measure we get

$$|\,(F-E)\exp\{\lambda(F-E)\}\,| \leqslant \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}|\,F^k(F-E)^s\,| \leqslant \mathrm{e}^{-\lambda}+$$

$$C(s)\sum_{k=1}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!\,k^s} \leqslant C(s)\sum_{k=0}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^k}{(k+s)!} = C(s)\lambda^{-s}\sum_{k=0}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^{k+s}}{(k+s)!} \leqslant C(s)\lambda^{-s}.$$

Example 2

Let $F \in \mathcal{F}_+$, n be natural number. Then

$$|\,F^n - \exp\{n(F-E)\}\,| \leqslant Cn^{-1}. \tag{13.3}$$

Proof. The norm of any distribution equals 1. Therefore, if $n \leqslant 10$, the proof follows from the fact that

$$|\,F^n - \exp\{n(F-E)\}\,| \leqslant \|\,F^n - \exp\{n(F-E)\}\,\| \leqslant \|\,F^n\,\| + \|\,\exp\{n(F-E)\}\,\| = 2.$$

Let us assume that $n > 10$, and let m be the integral part of $n/3$, i.e. the largest integer less than $n/3$. Then by (1.3), (2.7), (13.2) and (13.1)

$$|\,F^n - \exp\{n(F-E)\}\,| \leqslant \sum_{k=1}^{n}|\,(F-\exp\{F-E\})F^{k-1}\exp\{(n-k)(F-E)\}\,| \leqslant$$

$$C\sum_{k=1}^{n}|\,(F-E)^2F^{k-1}\exp\{(n-k)(F-E)\}\,| \leqslant$$

$$Cn\Big\{|\,(F-E)^2F^m\,| + |\,(F-E)^2\exp\{m(F-E)\}\,|\Big\} \leqslant Cnm^{-2} \leqslant Cn^{-1}.$$

Example 3

Let $F \in \mathcal{F}_+$, $a > 0$, $b > 0$. Then

$$|\exp\{a(F-E)\} - \exp\{b(F-E)\}| \leqslant C\,\frac{|a-b|}{b}. \tag{13.4}$$

Proof. Let $b < a$. Then by (13.2) and (2.7)

$$|\exp\{a(F-E)\} - \exp\{b(F-E)\}| = |\exp\{b(F-E)\}(\exp\{(a-b)(F-E)\} - E)| \leqslant$$

$$C|\exp\{b(F-E)\}(a-b)(F-E)| \leqslant C\,\frac{a-b}{b}.$$

Let $a < b \leqslant 2a$. Then similarly

$$|\exp\{a(F-E)\} - \exp\{b(F-E)\}| \leqslant C\,\frac{b-a}{a} \leqslant 2C\,\frac{b-a}{b}.$$

Let $2a \leqslant b$. Then

$$|\exp\{a(F-E)\} - \exp\{b(F-E)\}| \leqslant$$

$$\|\exp\{a(F-E)\}\| + \|\exp\{b(F-E)\}\| = 2 = 4\cdot\frac{1}{2} \leqslant 4\,\frac{b-a}{b}.$$

Example 4

Let $F \in \mathcal{F}_+$, n be natural number. Then

$$\left|\left(\frac{1}{2}E + \frac{1}{2}F^2\right)^n - \exp\{n(F-E)\}\right| \leqslant Cn^{-2}. \tag{13.5}$$

Proof. Set $G = 0.5E + 0.5F^2$. Just like in example 2 we prove that it suffices to consider $n > 10$. Moreover, setting m to be the integral part of $n/3$ we get

$$|G^n - \exp\{n(F-E)\}| \leqslant$$

$$Cn|G^m(G - \exp\{F-E\})| + Cn|\exp\{m(F-E)(G-\exp\{F-E\})\}|. \tag{13.6}$$

Now by (2.7)

$$G - \exp\{F-E\} = E + (F-E) + \frac{1}{2}(F-E)^2 - \exp\{F-E\} = \Theta(F-E)^3.$$

Here $\|\Theta\| \leqslant 1/3$. Therefore, estimating the second summand in (13.6) and applying (13.2) we get

$$Cn|\exp\{m(F-E)(G-\exp\{F-E\})\}| \leqslant Cn|\exp\{m(F-E)\}(F-E)^3| \leqslant Cnm^{-3} \leqslant Cn^{-2}.$$

The first summand in (13.6) is estimated by (13.1)

$$Cn|G^m(G-\exp\{F-E\})| \leqslant Cn|G^m(F-E)^3| \leqslant Cn\,\frac{1}{2^m}\sum_{k=0}^{m}\binom{m}{k}|F^{2k}(F-E)^3| \leqslant$$

$$Cn\,\frac{1}{2^m}\Big\{2+\sum_{k=0}^{m}\binom{m}{k}\frac{C}{(2k)^3}\Big\}\leqslant Cn\,\frac{1}{2^m}\sum_{k=0}^{m}\binom{m}{k}\frac{C}{(k+1)(k+2)(k+3)}=$$

$$C\frac{n}{(m+1)(m+2)(m+3)}\,\frac{1}{2^m}\sum_{k=0}^{m}\binom{m+3}{k+3}\leqslant Cn^{-2}.$$

The last estimate completes the proof of example.

Example 5

Let F be symmetric, n be natural number. Then

$$|\,F^n-\exp\{n(F-E)\}\,|\leqslant Cn^{-1/2}. \tag{13.7}$$

Note that the accuracy in (13.7) is much weaker than for $F\in\mathcal{F}_+$; see (13.1). Of course, any F from $\mathcal{F}_+$ is symmetric distribution. The crucial difference between two kinds lies in the fact that the characteristic functions of some symmetric distributions can take value -1. It is known that, in general, (13.7) is of the right order.
Proof. We shall combine the approach used in the proof (11.13) and (13.3). Let us decompose F by the as in (11.3) and (11.4) with $p=1/2$. Note that, due to the symmetry of F, we can take $u=0$. Now, as it follows from (11.13) and (11.19),

$$\Big|\,F^n-\exp\{n(1-p)(A-E)\}((1-p)E+pB)^n\,\Big|\leqslant C\sqrt{\frac{p}{n}}+C\sqrt{\frac{np(1-p)}{n^2p^2}}\leqslant Cn^{-1/2}.$$

On the other hand, $\widehat{B}(t)\geqslant -1$ and $(1-p)+p\widehat{B}(t)=0.5+0.5\widehat{B}(t)\geqslant 0$. Thus, $((1-p)E+pB)\in\mathcal{F}_+$, and by (13.3)

$$|\,((1-p)E+pB)^n-\exp\{np(B-E)\}\,|\leqslant Cn^{-1}.$$

It remains to apply the triangle inequality

$$|\,F^n-\exp\{n(F-E)\}\,|\leqslant|\,F^n-\exp\{n(1-p)(A-E)\}((1-p)E+pB)^n\,|+$$
$$\|\,\exp\{n(1-p)A(\,-E)\}\|\,|\,((1-p)E+pB)^n-\exp\{np(B-E)\}\,|.$$

Example 6

We shall assume that F does not depend on n, has two moments and satisfies Cramer's (C) condition:

$$\limsup_{|t|\to\infty}\,|\,\widehat{F}(t)\,|\,<1\,. \tag{C}$$

Then, for fixed $k\in\mathbb{N}$,

$$|\,F^n(F-E)^k\,|\leqslant C(F)n^{-k}. \tag{13.8}$$

and

$$|\,F^n-\exp\{n(F-E)\}\,|\leqslant C(F)n^{-1}; \tag{13.9}$$

see (10.16) and (10.20).

We shall extend (13.9) and prove that, for any $p \to 0$ and any $B \in \mathcal{F}$ (which might depend on n, have no finite moments etc.,

$$|\,((1-p)F + pB)^n - \exp\{n(1-p)(F-E) + np(B-E)\}\,| \leqslant C(F)(n^{-1} + p); \qquad (13.10)$$

Proof. We shall use the fact that $\|\, B^{n-j} \,\| = 1$, because $B^{n-j} \in \mathcal{F}$. Therefore applying (1.36) we get

$$\begin{aligned}
&|\,((1-p)F + pB)^n - \exp\{n(1-p)(F-E) + np(B-E)\}\,| \leqslant \\
&\sum_{j=1}^{n} \binom{n}{j} (1-p)^j p^{n-j} |\, F^j - \exp\{j(F-E)\}\,| + \\
&\sum_{j=1}^{n} \binom{n}{j} (1-p)^j p^{n-j} |\, \exp\{j(F-E)\} - \exp\{n(1-p)(F-E)\}\,| + \\
&|\, \exp\{n(1-p)(F-E)\}((1-p)E + pB)^n - \exp\{np(B-E)\})\,| = J_1 + J_2 + J_3.
\end{aligned}$$

Taking into account (13.9) the first sum can be estimated just like in (11.16) geting the estimate $J_1 \leqslant C(F)n^{-1}$. For the second sum we first just like in example 3 obtain

$$|\, \exp\{j(F-E)\} - \exp\{n(1-p)(F-E)\}\,| \leqslant C(F)\left(\frac{|\, n - j(1-p)\,|}{n}\right).$$

and then, just like in the proof of (11.19)

$$C \sum_{j=0}^{n} \binom{n}{j} (1-p)^j p^{n-j} \frac{|\, j - n(1-p)\,|}{n} \leqslant C\sqrt{\frac{p}{n}};$$

and, therefore,

$$J_2 \leqslant C(F)\left(\sqrt{\frac{p}{n}}\right).$$

Finally, by the Poisson approximation to the Binomial law (see (8.10), (5.8)

$$J_3 \leqslant \|\, ((1-p)E + pB)^n - \exp\{np(B-E)\})\,\| \leqslant Cp.$$

This evidently complete the proof of (13.10).

Exercises

1. Let $F \in \mathcal{F}_+$. To prove that

$$|\, F^n - \exp\{n(F-E)\}(E - n(F-E)^2/2)\,| \leqslant Cn^{-1}.$$

2. Let $F \in \mathcal{F}_+$, a, b natural numbers. To prove that

$$|\, F^a - F^b \,| \leqslant C\,\frac{|\, a - b\,|}{a}.$$

3. Let $F \in \mathcal{F}_+$, a, h natural numbers. To prove that

$$\Big| F^a - \exp\Big\{ \frac{a}{h}\,(F^h - E) \Big\} \Big| \leqslant C\,\frac{h}{a}\,.$$

4. Let ξ and η be two independent integer-valued random variables $P(\xi = k) = p_k$, $P(\eta = k) = q_k$, (k=0,1,...) and let

$$\varphi(F) = \sum_{k=0}^{\infty} p_k F^k, \quad \psi(F) = \sum_{k=0}^{\infty} q_k F^k.$$

Let $F \in \mathcal{F}_+$. to prove that

$$|\,\varphi(F) - \psi(F)\,| \leqslant \mathbb{E}\,\frac{|\,\xi - \eta\,|}{\max\{\xi, \eta\} + 1}\,.$$

Bibliographical notes

Estimates (13.2) and (13.5) are partial cases of more general results from Čekanavičius (1995). Estimate (13.3) was obtained in Arak (1980) by the triangle functions method. Estimate (13.7) was proved by Zaĭtsev (1983a). Other examples, (13.4) and different proof of (13.7) can be found in Arak and Zaĭtsev (1988). Moreover, there many other generalizations of the results of this Section; see, for example, Čekanavičius (1989), (1997). Exercises also can be found in the mentioned papers.

14 Estimating abolutely continuous distributions

For continuous random variables it is more convenient to use distribution functions. The distribution function and $F(x)$ is defined as $F(x) = F\{(-\infty, x\}$. The distribution F is absolutely continuous if and only if, for every x,

$$F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y. \tag{14.1}$$

Here $f(y)$ is non-negative function integrable on the real line, the so-called density function. Integral in (14.1) is Lebesgue integral.

If F is absolutely continuous distribution with the density $f(x)$ and $G \in \mathcal{F}$, then FG is also absolutely continuous distribution having density:

$$p(y) = \int_{-\infty}^{\infty} f(y - x)\,G\{\mathrm{d}x\}.$$

If, in addition, G has the density $g(x)$ then FG has the density

$$p(y) = \int_{-\infty}^{\infty} f(y - x) g(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} g(y - x) f(x)\,\mathrm{d}x.$$

Let distribution F have density $f(x)$. Then its characteristic function is

$$\widehat{F}(t) = \int\limits_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx} f(x)\,\mathrm{d}x.$$

If the characteristic function $\widehat{F}(t)$ is absoulutely integrable on $\mathbb{R}$ then F has a continuous density $f(x)$ and, for all $x \in \mathbb{R}$, the following formula of inversion holds:

$$f(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}tx} \widehat{F}(t)\,\mathrm{d}t. \tag{14.2}$$

In this Section we shall consider the scheme of sequences, that is we assume that distributions do not depend on n.

14.1 Local estimates for bounded densities

If F has a bounded density then

$$\int_{-\infty}^{\infty} |\,\widehat{F}(t)\,|^2 dt \leqslant C(F). \tag{14.3}$$

Moreover, there exists $\varepsilon = \varepsilon(F)$, such that

$$|\,\widehat{F}(t)\,| \leqslant \begin{cases} \exp\{-C_1(F)t^2\}, & \text{if } |\,t\,| \leqslant \varepsilon, \\ \exp\{-C_2(F)\}, & \text{if } |\,t\,| > \varepsilon. \end{cases}$$

Combining, these properties and the inversion formula (14.2) it is not difficult to get estimates.

Example 1. Let F and G be continuous distributions, having bounded densities $f(x)$ and $g(x)$ respectively, and let $s-1$ $(s \geqslant 1)$ pseudomoment equal to zero and s'th absolute pseudomoment be finite. That is, for $k = 1, 2, \ldots, s-1$

$$\int\limits_{-\infty}^{\infty} x^k (f(x) - g(x))\,\mathrm{d}x = 0, \qquad \int\limits_{-\infty}^{\infty} |\,x\,|^s |\,f(x) - g(x)\,|\,\mathrm{d}x < C(F, G) < \infty. \tag{14.4}$$

Let $n > 1$ and let denote the densities of F^n and G^n by $f_n(x)$ and $g_n(x)$, respectively. Then

$$\sup_x |\,f_n(x) - g_n(x)\,| \leqslant C(F, G) n^{-(s-1)/2}. \tag{14.5}$$

Proof. From expansion in moments we have that, for all t,

$$|\,\widehat{F}(t) - \widehat{G}(t)\,| \leqslant C(F, G, s) |\,t\,|^s.$$

Moreover, we can find $\varepsilon = \varepsilon(F, G)$ such that estimates for the characteristic functions in above apply. Then by formula of inversion (14.2).

$$\sup_x |\,f_n(x) - g_n(x)\,| \leqslant C \int\limits_{-\infty}^{\infty} |\,\widehat{F}^n(t) - \widehat{G}^n(t)\,|\,\mathrm{d}t \leqslant$$

$$
\int_{-\varepsilon}^{\varepsilon} |\,\widehat{F}^n(t) - \widehat{G}^n(t)\,|\,\mathrm{d}t + \int_{|t|>\varepsilon} (|\,\widehat{F}(t)\,|^n + |\,\widehat{G}(t)\,|^n)\ \mathrm{d}t \leqslant
$$

$$
C(F,G) \int_{-\varepsilon}^{\varepsilon} n|\,\widehat{F}(t) - \widehat{G}(t)\,|\exp\{-C_1(F,G)(n-1)t^2\}\,\mathrm{d}t +
$$

$$
C(F,G)\exp\{-C_2(F,G)(n-2)\} \int_{-\infty}^{\infty} (|\,\widehat{F}(t)\,|^2 + |\,\widehat{G}(t)\,|^2)\,\mathrm{d}t \leqslant
$$

$$
C(F,G,s) \int_{-\varepsilon}^{\varepsilon} n|\,t\,|^s \exp\{-C_1(F,G)nt^2\}\,\mathrm{d}t + C(F,G)\exp\{-C_2(F,G)n\} \leqslant
$$

$$
C(F,G,s)n^{-(s-1)/2}.
$$

14.2 Integral estimates

14.3 Estimates in total variation

For the estimates in total variation of absolutely continuous distributions F and G having densities $f(x)$ and $g(x)$ we recall that

$$
\|\,F - G\,\| = \int_{-\infty}^{\infty} |\,f(x) - g(x)\,|\,\mathrm{d}x. \tag{14.6}
$$

One of the most convenient formula of inversion for continuous distributions is the following one.

Lemma 14.1 *Let $F, G \in \mathcal{F}$ be absolutely continuous distributions having densities $f(x)$ and $g(x)$, respectively. Let*

$$
\int_{-\infty}^{\infty} |\,x\,|\,|\,f(x) - g(x)\,|\,\mathrm{d}x < \infty.
$$

Then

$$
\|\,F - G\,\| \leqslant \frac{1}{2}\Big(\int_{-\infty}^{\infty} |\,\widehat{F}(t) - \widehat{G}(t)\,|^2\,\mathrm{d}t \int_{-\infty}^{\infty} |\,\widehat{F}'(t) - \widehat{G}'(t)\,|^2\,\mathrm{d}t\Big)^{1/4} \tag{14.7}
$$

Sometimes it is more convenient to use inversion formula, which is mor similar to the discrete case.

Lemma 14.2 *Let $F, G \in \mathcal{F}$ be absolutely continuous distributions having densities $f(x)$ and $g(x)$, respectively. Let*

$$\int_{-\infty}^{\infty} |\, x \,|\, |\, f(x) - g(x) \,| \, \mathrm{d}x < \infty.$$

Then, for any $b > 0$,

$$\| \, F - G \, \| \leqslant \frac{1}{2\sqrt{2}} \Big(\int_{-\infty}^{\infty} \Big(b |\, \widehat{F}(t) - \widehat{G}(t) \,|^2 + b^{-1} |\, \widehat{F}'(t) - \widehat{G}'(t) \,|^2 \Big) \, \mathrm{d}t \Big)^{1/2} \tag{14.8}$$

Example 1. Let F and G be continuous distributions, having bounded densities $f(x)$ and $g(x)$ respectively, means equal to zero, and let $s-1$ ($s \geqslant 1$) pseudomoment equal to zero and s'th absolute pseudomoment be finite. That is, for $k = 1, 2, \ldots, s-1$

$$\int_{-\infty}^{\infty} x^k (f(x) - g(x)) \, \mathrm{d}x = 0, \qquad \int_{-\infty}^{\infty} |\, x \,|^s |\, f(x) - g(x) \,| \, \mathrm{d}x < C(F, G) < \infty. \tag{14.9}$$

Then

$$\| \, F^n - G^n \, \| \leqslant C(F, G) n^{-(s-2)/2}. \tag{14.10}$$

We shall apply (14.9) with $b = \sqrt{n}$. There exist $\varepsilon = \varepsilon(f, G)$ such that, for $|\, t \,| \leqslant \varepsilon$,

$$\max\{|\, \widehat{F}(t) \,|, |\, \widehat{G}(t) \,|\} \leqslant \exp\{-C_1(F, G) t^2\} \tag{14.11}$$

and, for $|\, t \,| > \varepsilon$,

$$\max\{|\, \widehat{F}(t) \,|, |\, \widehat{G}(t) \,|\} \leqslant \exp\{-C_2(F, G)\}. \tag{14.12}$$

We assumed that the mean of F is zero. Thus, the derivative of $\widehat{F}(t)$ is bounded from above by

$$|\, \widehat{F}'(t) \,| \leqslant C(F), \quad |\, \widehat{F}'(t) \,| \leqslant C(F) |\, t \,|,$$

and similar estimates hold for $|\, \widehat{G}'(t) \,|$.

Due to the bounded densities $|\, \widehat{F}(t) \,|^2$ and $|\, \widehat{G}(t) \,|^2$ are integrable. Therefore

$$\int_{|t| > \varepsilon} (|\, \widehat{F}(t)^n - \widehat{G}(t) \,|^n) \, \mathrm{d}t \leqslant C(F, G) \exp\{-C_2(F, G)(n-2)\} \int_{|t| > \varepsilon} (|\, \widehat{F}(t) \,|^2 + |\, \widehat{G}(t) \,|^2) \, \mathrm{d}t \leqslant$$

$$C(F, G) \exp\{-C(F, G) n\}$$

and

$$\int_{|t| > \varepsilon} (|\, (\widehat{F}(t)^n - \widehat{G}(t)^n)' \,|) \, \mathrm{d}t \leqslant n \int_{|t| > \varepsilon} (|\, \widehat{F}'(t) \,|\, |\, \widehat{F}(t) \,|^{n-1} + |\, \widehat{G}'(t) \,|\, |\, \widehat{G}(t) \,|^{n-1}) \, \mathrm{d}t \leqslant$$

$$nC(F, G) \exp\{-C_2(F, G)(n-3)\} \int_{|t| > \varepsilon} (|\, \widehat{F}(t) \,|^2 + |\, \widehat{G}(t) \,|^2) \, \mathrm{d}t \leqslant C(F, G) \exp\{-C(F, G) n\}.$$

Therefore, the estimate reduces to the case

$$\| F^n - G^n \| \leqslant \frac{1}{2\sqrt{2}}\left(\int\limits_{-\varepsilon}^{\varepsilon} \left(\sqrt{n}\,|\,\widehat{F}^n(t) - \widehat{G}^n(t)\,|^2 + \frac{1}{\sqrt{n}}|\,(\widehat{F}^n(t) - \widehat{G}^n(t))'\,|^2\right) \mathrm{d}t\right)^{1/2} + C(F,G)n^{-(s-2)/2}. \tag{14.13}$$

From expansion in moments we have that, for all t,

$$|\,\widehat{F}(t) - \widehat{G}(t)\,| \leqslant C(F,G,s)|\,t\,|^s, \quad |\,\widehat{F}'(t) - \widehat{G}'(t)\,| \leqslant C(F,G,s)|\,t\,|^{s-1}.$$

Consequently, for $|\,t\,| \leqslant \varepsilon$,

$$|\,\widehat{F}^n(t) - \widehat{G}^n(t)\,| \leqslant n\exp\{-C_1(F,G)nt^2\}|\,\widehat{F}(t) - \widehat{G}(t)\,| \leqslant C(F,G)n\exp\{-C_1(F,G)nt^2\}|\,t\,|^s.$$

Similarly, for $|\,t\,| \leqslant \varepsilon$,

$$|\,(\widehat{F}^n(t) - \widehat{G}^n(t))'\,| \leqslant n|\,\widehat{F}'(t)\,|\,|\,\widehat{F}^{n-1}(t) - \widehat{G}^{n-1}(t)\,| + n|\,\widehat{G}(t)\,|^{n-1}|\,\widehat{F}'(t) - \widehat{G}'(t)\,| \leqslant$$
$$C(F,G)n^2\exp\{-C_1(F,G)nt^2\}|\,t\,|^{s+1} + C(F,G)n\exp\{-C_1(F,G)nt^2\}|\,t\,|^{s-1}.$$

Now it suffices to substitute the last estimates into (14.13).