Estimation errors for functionals on measure spaces

N. Golyandina and V. Nekrutkin
St. Petersburg University, St. Petersburg, Russia

Abstract: The article is devoted to investigation of bias and mean-square deviation while estimating smooth functionals on measure spaces. Such problems appear when one uses a Monte-Carlo procedure to solve certain linear or nonlinear equation on measures (f.e., nonlinear Boltzmann-like equations), and wants to estimate and diminish the part of an error corresponding to simulation of an initial distribution. It is shown that by means of a special variant of stratification technique both bias and variance of this part of error may be reduced to $o(1/n)$ instead of usual $O(n^{-1})$. Several simple examples of stratification for linear functionals (i.e., integrals) are represented.

Keywords and phrases: Non-linear functionals, measure-valued random variables, stratification technique.

---

1 Introduction

We start with an example elucidating the problem under consideration. Let (D,ρ) be some metric space with Borel σ-algebra $\mathcal{B}$. Consider linear evolution equation on measures

$$\mu_t = \int\limits_0^t e^{-(t-s)} \int\limits_D T(\,\cdot\,;u)\mu_s(du)ds + e^{-t}\mu\,, \tag{1}$$

where $\mu \in H$, H is the set of all probabilistic measures on $(D,\mathcal{B})$, $T(\,\cdot\,;u) \in H$ for any u and $T(A;\,\cdot\,)$ is $\mathcal{B}$-measurable for any $A \in \mathcal{B}$. Surely, the solution $\mu_t = \mu_t(\mu)$ is a probabilistic measure too.

Usual Monte-Carlo procedure for estimating the integral $\psi(\mu_t)=\int_D f d\mu_t$ may be described as following. One simulates n independent copies $\zeta_i(s)$ of jump Markov process with the initial distribution $\mathcal{L}(\zeta_i(0))=\mu$, jump law T and independent exponentially distributed (with the unit scale parameter) time intervals between jumps, and takes

$$\eta_n=\frac{1}{n}\sum_{i=1}^{n} f(\zeta_i(t))$$

as the estimator of $\psi(\mu_t)$.

Let us consider this procedure from the other viewpoint. If we collect processes $\zeta_i(s)$ onto one *n-particle process* $\eta^{(n)}(s)=\Big(\zeta_1(s),\ldots,\zeta_n(s)\Big)\in D^n$, then $\eta^{(n)}(s)$ is also a jump Markov process. Its exponential scale parameter is equal to n, and the jump law may be expressed as following. If $\mathbf{u}=(u_1,\ldots,u_n)$ is the phase coordinate of the n-particle process before a jump, we randomly choose one of the particles, and supposing this particle has number i, it jumps to a new position due to the distribution $T(\,\cdot\,;u_i)$. The initial distribution of the process is equal to $\mu^{\otimes n}$ (here $\mu_1\otimes\mu_2$ stands for the outer product of measures μ_1 and μ_2, $\mu^{\otimes n}=\mu\otimes\ldots\otimes\mu$ (n times)).

From the other hand, any n-particle process $\eta^{(n)}(s)$ produces the corresponding measure-valued process

$$\xi_s^{(n)}=\frac{1}{n}\sum_{i=1}^{n}\delta_{\zeta_i(s)},$$

where δ_x is the Dirac measure, concentrated at the point x. Such processes may be called as *empirical*. Surely, $\xi_s^{(n)}$ is also a jump Markov process.

In terms of the empirical process we have $\eta_n=\psi(\xi_t^{(n)})$, and the estimation error $\psi(\xi_t^{(n)})-\psi(\mu_t)$ may be represented in the form

$$\begin{aligned}\psi(\xi_t^{(n)})&-\psi(\mu_t(\mu))\\ &=\Big(\psi(\xi_t^{(n)})-\psi(\mu_t(\xi_0^{(n)}))\Big)+\Big(\psi(\mu_t(\xi_0^{(n)}))-\psi(\mu_t(\mu))\Big)\,,\end{aligned}\tag{2}$$

where $\mu_t(\xi_0^{(n)})$ stands for the solution of (1) with the initial value $\xi_0^{(n)}$.

As the start point of the process $\xi_s^{(n)}$ is also $\xi_0^{(n)}$, the first addend in the right-hand side of (2) describes the difference between transition mechanisms of the evolution equation (1) and Markov process $\xi_t^{(n)}$, while the second may be called as the part of the error, produced by the initial value $\xi_0^{(n)}$. Our aim is to investigate statistical properties of this second addend.

If we fix time moment t and denote $F : H \mapsto \mathbb{R}$ by $F = \psi \circ \mu_t$ (note that $\mu_t : H \mapsto H$), then it occurs that we are interesting in the difference $F(\xi_0^{(n)}) - F(\mu)$. When we deal with the linear functional ψ and with linear equation (1), F may be considered as linear. But if we want to estimate some nonlinear functional (such as $\psi(\mu_t) = (\int f d\mu_t)^2$) or if we consider nonlinear evolution equation (f.e., Boltzmann-like equation, where similar n-particle processes may be applied to), F is nonlinear.

Thus the first problem under consideration is to investigate the bias and variance for the estimator $F(\xi_n)$ of $F(\mu)$, where F is some smooth functional (the exact definitions will be given further) and ξ_n is the random empirical measure

$$\xi_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\zeta_i}$$

with independent ζ_i, $\mathcal{L}(\zeta_i) = \mu$.

The second problem is to understand the role of independence for ζ_i. Surely, if ζ_i are 'almost independent', the result must be the same as if they are independent. Can one choose dependent $\zeta_i = \zeta_i^{(n)}$ in such a way, that for any smooth F both bias and variance would be less, than in the simplest case of independence? It occurs that generally the answer is positive — if one manages to apply the proper stratification technique, both error characteristics would have the form $o(1/n)$, while in the case of independence they are $O(1/n)$.

To demonstrate such propositions we must deal with functions defined on the appropriate measure spaces, their derivative mappings etc. Definitions and simplest properties of such objects are collected in the Section 2.

Section 3 contains main theoretical results of the paper. Theorem 3.1 tells us that for independent $\zeta_i^{(n)}$ and under some smoothness conditions on F, the bias $m_0(F) = \mathbf{E}F(\xi_n) - F(\mu)$ has the form $A(F)/n + o(1/n)$. As for $A(F)$, it is a function of the second derivative $d^2F(\mu)/d\mu^2$.

The same result takes place if $\zeta_i^{(n)}$ are 'almost independent'. More precisely, these random variables must be equidistributed with the distribution μ, their mutual distribution must be invariant under their permutation, and the condition $\mathcal{P}_2^{(n)} = \mathcal{L}(\zeta_1^{(n)}, \zeta_2^{(n)}) = \mu^{\otimes 2} + o(1/n)$ must hold.

The Corollary 3.1 deals with $s_0^2(F) = \mathbf{E}(F(\xi_n) - F(\mu))^2$. It is proved that under the same conditions, $s_0^2(F) = B(F)/n + o(1/n)$ with $B(F)$ depending on the first derivative $dF(\mu)/d\mu$.

The second proposition of the Theorem 3.1 claims that if

$$\mathcal{P}_2^{(n)} = \frac{n+1}{n}\,\mu^{\otimes 2} - \frac{1}{n}\,\nu_n + o(1/n), \tag{3}$$

where ν_n is a probabilistic measure on Borel subsets of D^2, ν_n weakly converges to ν, and ν is the 'diagonal measure' $\mathcal{L}(\zeta,\zeta)$, $\mathcal{L}(\zeta) = \mu$, then $m_0(F) = o(1/n)$ for any sufficiently smooth F. Moreover, $s_0^2(F)$ have the same form here. The proof of these propositions is based on the technical Lemma 3.1, which gives the appropriate variant of the Taylor decomposition formula applied to the spaces under consideration.

The last Section 4 is devoted to investigation of the condition (3) and to several examples. It occurs that the proper variant of the stratification technique leads to (3). Roughly speaking (see Corollary 4.1 for precise results), this stratification may be described as following.

For each n let us consider such n subdomains $D_{i,n} \subset D$, that (up to some small number of subdomains) $\mathrm{diam}(D_{i,n}) \to 0$ uniformly on i. Let $\mu_i^{(n)} \in H$, $\mu = \sum_{i=1}^n \mu_i^{(n)}/n$ and $\mu_i^{(n)}(D_{i,n}) = 1$. Suppose $\widehat{\zeta}_i^{(n)}$ are independent and $\mathcal{L}(\widehat{\zeta}_i^{(n)}) = \mu_i^{(n)}$. Denote $(\zeta_1^{(n)}, \ldots, \zeta_n^{(n)})$ the result of random permutation for random variables $(\widehat{\zeta}_1^{(n)}, \ldots, \widehat{\zeta}_n^{(n)})$.

Then, by the assertion of the Corollary 4.1, the conditions of the second proposition of the Theorem 3.1 are fulfilled.

Linear functional $F(\mu) = \int f d\mu$, $f \in \mathbf{C}(D)$, is the simplest example of smooth functional on measure space. Surely, the case of independent ζ_i with $\mathcal{L}(\zeta_i) = \mu$ is trivial here, but integrals are good area for stratification, as one can try to get more precise results for them. Several examples of such a kind may be found in the Subsection 4.2.

If $D = [0,1]$ and μ is the Lebesgue measure on D, the simplest stratification with $D_{i,n} = [(i-1)/n, i/n]$ and $\mu_i^{(n)}$ taken as the uniform distribution on $D_{i,n}$ leads to error $o(n^{-1})$ for any continuous f (this is just the proposition of the Corollary 4.1) and to $O(n^{-3})$ if $f \in \mathbf{C}^1(D)$ (or even f is Lipschitz continuous). This corresponds to the old result due to Bahvalov (1964). Moreover, in the case $f \in \mathbf{C}^1(D)$ the version of Central Limit Theorem may be proved. Other similar examples deal with discrete measures, general one-dimensional integrals and multidimensional case.

The last example is based on the Theorem 3.1 itself but not on the Corollary 4.1. The equality (12) shows that to diminish the variance, random variables $\zeta_i^{(n)}$ must 'avoid each other' with the probability equivalent to $1/n$. Due to the Corollary 4.1 this effect is provided with the help of small supports $D_{i,n}$ of distributions $\mu_{i,n}$. Here we produce such a property in the more direct way.

Denote $D = (0,1]$ and $\rho(x,y) = \min\{|x-y|, 1-|x-y|\}$. This means that we consider D not as the interval, but as the circle. Let B_n be such subset of D^n, that $\rho(x_i,x_j) > 1/2n$ if $i \neq j$ and $(x_1,\dots,x_n) \in B_n$.

For fixed $n > 2$ consider random vector $\eta_n = (\zeta_1^{(n)},\dots,\zeta_n^{(n)})$ which is uniformly distributed on B_n. It may be easily seen that all conditions of the Theorem 3.1 (second propositions) are fulfilled, in particular (3) takes place. Moreover, exact calculations show that such 'stratification without stratification' may give better results than the usual division of $[0,1]$ onto n equal intervals.

---

2 Strong weakly-continuous derivatives

Let us start with some general terms and notation.

For Banach spaces E_1 and E_2 with norms $\|\cdot\|_1$ and $\|\cdot\|_2$ we denote by $\mathbf{L}(E_1,E_2)$ the Banach space of bounded linear operators acting from E_1 to E_2. Let U be a certain open subset of E_1. The mapping $g' : U \mapsto \mathbf{L}(E_1,E_2)$ is called as the *derivative mapping* (see Cartan (1967) for details) or briefly as the *derivative* of the mapping $g : U \mapsto E_2$, if for any $\mu, \widetilde{\mu} \in U$ with $\|\widetilde{\mu}-\mu\|_1 \to 0$

$$\|g(\widetilde{\mu}) - g(\mu) - g'(\mu)(\widetilde{\mu}-\mu)\|_2 = o(\|\widetilde{\mu}-\mu\|_1). \tag{4}$$

Note, that the derivative g' may be considered as the mapping $U \times E_1 \mapsto E_2$, which is linear on the second argument.

Let us introduce notation for derivatives of the k-th order, using the second derivative as an example. By definition, the second derivative mapping is a mapping

$$g'' : U \mapsto \mathbf{L}(E_1 \times E_1, E_2),$$

which satisfies the relation analogous to (4). Therefore we use the notation $g''(\mu)$ or $d^2g(\mu)/d\mu^2$ for the value of this mapping at the point $\mu \in U$.

On the other hand one may deal with the second derivative as a mapping $U \times E_1 \times E_1 \mapsto E_2$, which is linear on the second and third arguments and is invariant under their permutation. Then we denote the second derivative as

$$\partial^2 g : U \times E_1 \times E_1 \mapsto E_2.$$

Thus $g''(\mu)(\nu_1, \nu_2)$ and $\partial^2 g(\mu; \nu_1, \nu_2)$ mean the same for $\mu \in U$ and $\nu_1, \nu_2 \in E_1$.

Now we suppose that (D, ρ) is some complete metrical separable compact space, $\mathcal{F}^\rho$ is the topology, induced by ρ, $\mathcal{B}$ is the corresponding Borel σ-algebra and E is the linear space of Radon measures on $(D, \mathcal{B})$. It is well known that the space E with the norm $\|\cdot\| = \mathrm{Var}(\cdot)$ is a Banach space (not separable in general).

Let $\mathcal{F}_w$ be the topology induced by week convergence of measures on E, (namely, μ_n weakly converges to μ, $\mu_n \xrightarrow{w} \mu$, if $\int f d\mu_n \to \int f d\mu$ for any bounded continuous function f on D).

Definition 2.1 The function $\psi : E \mapsto \mathbb{R}$ is called *w-continuously differentiable*, if there exists a derivative mapping ψ' which is continuous as a mapping $(E, \mathcal{F}_w) \times (E, \mathcal{F}_w) \mapsto \mathbf{R}$.

The following functional spaces will be used further:

1. $\mathbf{C}(E)$ is the space of continuous on $(E, \mathcal{F}_w)$ functions,
2. $\widetilde{\mathbf{C}}^1(E)$ is a set of all w-continuously differentiable functions;
3. $\mathbf{C}^1(E) = \mathbf{C}(E) \cap \widetilde{\mathbf{C}}^1(E)$.

The analogous definition may be done for $\mathbf{C}^k(E)$, $k \geq 2$. The following lemma gives useful representation for derivative mappings, induced by functions from $\mathbf{C}^1(E)$.

Lemma 2.1 *Let $\psi \in \mathbf{C}^1(E)$. Then there exists such $\mathcal{F}_w \times \mathcal{F}^\rho$-continuous function $\psi' : E \times D \mapsto \mathbf{R}$, that for any $\nu \in E$*

$$\partial^1 \psi(\mu; \nu) = \int_D \psi'(\mu, v) \nu(dv).$$

Proof. Let $u \in D$. Denote $\psi'(\mu, u) = \partial^1 \psi(\mu; \delta_u) = \int_D \psi'(\mu, v) \delta_u(dv)$. As $\psi \in \mathbf{C}^1(E)$, the function ψ' is continuous on its arguments. Besides, as D is a compact set, $\psi'(\mu, \cdot)$ is bounded for fixed μ. Evidently,

$$\partial^1 \psi(\mu; \nu_n) = \int_D \psi'(\mu, v) \nu_n(dv) \qquad (5)$$

for any discrete measure $\nu_n = \sum\limits_{i=1}^{n} \alpha_i^{(n)} \delta_{u_i^{(n)}}$. Let $\nu \in E$. As the linear span of the set of Dirac measures is dense in $(E, \mathcal{F}_w)$, there exists such sequence $\{\nu_n\}$, that $\nu_n \xrightarrow{w} \nu$. Taking $n \to \infty$ in (5) we get the result. ∎

Note that these definitions and features may be extended to higher derivatives. For example, the analogue of Lemma 2.1 for second derivatives may be stated as following.

Lemma 2.2 *Let $\psi \in \mathbf{C}^2(E)$. Then there exists such $\mathcal{F}_w \times \mathcal{F}^\rho \times \mathcal{F}^\rho$-continuous function $\psi'' : E \times D \times D \mapsto \mathbf{R}$, that for any $\nu_1, \nu_2 \in E$*

$$\partial^2 \psi(\mu; \nu_1, \nu_2) = \int\limits_{D^2} \psi''(\mu, v_1, v_2) \nu_1(dv_1) \nu_2(dv_2).$$

---

3 General results

Let $\mu \in H$. For $n \geq 1$ and $1 \leq i \leq n$ consider random variables $\zeta_i^{(n)}$, $\tilde{\zeta}_i^{(n)} \in D$, such that $\mathcal{L}(\zeta_1^{(n)}, \ldots, \zeta_n^{(n)}) = \mathcal{L}(\tilde{\zeta}_1^{(n)}, \ldots, \tilde{\zeta}_n^{(n)})$ and $\zeta_i^{(n)}$, $\tilde{\zeta}_j^{(n)}$ are independent for any i and j. Let

$$\xi_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\zeta_i^{(n)}}, \qquad \tilde{\xi}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\tilde{\zeta}_i^{(n)}}$$

and define $\mu_n = \mathcal{E}(\xi_n)$ as such distribution that

$$\int\limits_D f d\mu_n = \mathbf{E} \int\limits_D f d\xi_n,$$

$f \in \mathbf{C}(D)$. The following technical proposition is the base of further estimates.

Lemma 3.1 *If uniformly on $1 \leq i, j \leq n,\ i \neq j$,*

$$\|\mathcal{L}(\zeta_i^{(n)}, \zeta_j^{(n)}) - \mathcal{L}(\zeta_i^{(n)}) \otimes \mathcal{L}(\zeta_j^{(n)})\| \underset{n\to\infty}{\longrightarrow} 0 \tag{6}$$

and $\|\mu_n - \mu\| \underset{n\to\infty}{\longrightarrow} 0$, *then for any* $F \in \mathbf{C}^3(E)$

$$
\begin{aligned}
m_0(F) &\overset{def}{=} \mathbf{E}F(\xi_n) - F(\mu) = \mathbf{E}\partial^1 F(\mu; \xi_n - \mu) \\
&+\mathbf{E}\int_0^1 (1-\tau)\,\partial^2 F(\alpha_n; \widetilde{\xi}_n - \mu, \widetilde{\xi}_n - \mu)\,d\tau + o(1/n) \qquad (7)
\end{aligned}
$$

with $\alpha_n = \mu + \tau(\xi_n - \mu)$.

Proof. On the strength of the Taylor formula with the integral remainder term it is sufficient to prove that

$$
\begin{aligned}
\Lambda_n &= \mathbf{E}\,\partial^2 F(\alpha_n; \xi_n - \mu, \xi_n - \mu) \\
&= \mathbf{E}\,\partial^2 F(\alpha_n; \widetilde{\xi}_n - \mu, \widetilde{\xi}_n - \mu) + o(1/n) = \widetilde{\Lambda}_n + o(1/n).
\end{aligned}
$$

Let $f_2(\nu; v_1, v_2) = \partial^2 F(\nu; \delta_{v_1}, \delta_{v_2})$. Further we omit the upper index while writing down $\zeta_i^{(n)}$ or $\widetilde{\zeta}_i^{(n)}$. As $\partial^2 F$ is linear on second and third arguments,

$$
\begin{aligned}
\Lambda_n &= \mathbf{E}\frac{1}{n^2}\sum_{i\neq j} f_2(\alpha_n; \zeta_i, \zeta_j) + \mathbf{E}\frac{1}{n^2}\sum_i f_2(\alpha_n; \zeta_i, \zeta_i) \\
&-\frac{2}{n}\mathbf{E}\sum_i f_2(\alpha_n; \zeta_i, \mu) + \mathbf{E}f_2(\alpha_n; \mu, \mu) \\
&= I_1 + I_2 + I_3 + I_4.
\end{aligned}
$$

Let us rewrite I_1 in another form (I_2 and I_3 may be rewritten in the same manner). Denote $\Delta^{(i)} = \delta_{\widetilde{\zeta}_i} - \delta_{\zeta_i}$, $\Delta^{(ij)} = \Delta^{(i)} + \Delta^{(j)}$, $\xi_n^{(ij)} = \xi_n + \Delta^{(ij)}/n$ and

$$
\Phi(\nu; \nu_1, x, y) = \frac{df_2(\nu; x, y)}{d\nu}(\nu_1).
$$

Using Taylor formula for the expansion of the function $f_2(\alpha_n; \zeta_i, \zeta_j)$ on the first argument at the point $\alpha_n^{(ij)} = \mu + \tau(\xi_n^{(ij)} - \mu)$, we obtain

$$
\begin{aligned}
&f_2(\alpha_n; \zeta_i, \zeta_j) \\
&\quad = f_2(\alpha_n^{(ij)}; \zeta_i, \zeta_j) - \frac{\tau}{n}\,\Phi(\alpha_n^{(ij)}; \Delta^{(ij)}, \zeta_i, \zeta_j) + \frac{\tau}{n}\,o\left(\|\Delta^{(ij)}\|\right)
\end{aligned}
$$

with $\|\Delta^{(ij)}\| \le 4$. Now, applying the properties of ζ_i and $\widetilde{\zeta}_i$, we can change our notation and write ζ_i instead of $\widetilde{\zeta}_i$, ξ_n instead of $\xi_n^{(ij)}$, etc. Therefore

$$I_1 = \frac{1}{n^2}\sum_{i\neq j}\mathbf{E}f_2(\alpha_n;\widetilde{\zeta}_i,\widetilde{\zeta}_j)$$
$$+\frac{\tau}{n^3}\sum_{i\neq j}\mathbf{E}\Phi(\alpha_n;\Delta^{(ij)},\widetilde{\zeta}_i,\widetilde{\zeta}_j)+o(n^{-1}) = I_1^{(1)}+I_1^{(2)}+o(n^{-1}).$$

Analogously $I_l = I_l^{(1)}+I_l^{(2)}+o(n^{-1})$ for $l=2,3$, and $I_1^{(1)}+I_2^{(1)}+I_3^{(1)}+I_4 = \widetilde{\Lambda}_n$. Thus $\Lambda_n = \widetilde{\Lambda}_n+\tau J+o(n^{-1})$, where

$$J = \frac{1}{n^3}\sum_{i\neq j}\mathbf{E}\Phi(\alpha_n;\Delta^{(ij)},\widetilde{\zeta}_i,\widetilde{\zeta}_j)+\frac{1}{n^3}\sum_{i}\mathbf{E}\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,\widetilde{\zeta}_i)$$
$$-\frac{2}{n^2}\sum_i \mathbf{E}\int\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,v)\mu(dv) = J_1+J_2+J_3.$$

Now we must prove that $J=o(n^{-1})$. Note that $J_2 = o(n^{-1})$. Using the invariant features of Φ under permutations of two last arguments, we have

$$J_1 = \frac{1}{n^3}\sum_{i,j}\mathbf{E}\Phi(\alpha_n;\Delta^{(ij)},\widetilde{\zeta}_i,\widetilde{\zeta}_j)+o(n^{-1})$$
$$= \frac{2}{n^2}\sum_i\mathbf{E}\int\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,v)\widetilde{\xi}_n(dv)+o(n^{-1}). \tag{8}$$

Let $(\zeta_1^*,\ldots,\zeta_n^*)$ be another independent copy of $(\zeta_1,\ldots,\zeta_n)$. Denoting

$$\xi_n^* = \frac{1}{n}\sum_{i=1}^n \delta_{\zeta_i^*},$$

we see that $\mathcal{E}(\xi_n^*) = \mu_n$. Let us prove that

$$J_1 = \frac{2}{n^2}\sum_i\mathbf{E}\int\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,v)\mu_n(dv)+o(n^{-1}). \tag{9}$$

By (8) one can obtain that up to $o(n^{-1})$

$$J_1 = \frac{2}{n^3}\sum_{i,j}\mathbf{E}\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,\widetilde{\zeta}_j)$$
$$= \frac{2}{n^3}\sum_{i\neq j}\mathbf{E}\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,\widetilde{\zeta}_j)+\frac{2}{n^3}\sum_i\mathbf{E}\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,\widetilde{\zeta}_i)$$

$$= \frac{2}{n^2}\sum_i \mathbf{E}\int \Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,v)\xi_n^*(dv)$$

$$+\frac{2}{n^3}\sum_{i\neq j}\mathbf{E}\left(\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,\widetilde{\zeta}_j)-\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,\zeta_j^*)\right)$$

$$+\frac{2}{n^3}\sum_i \mathbf{E}\left(\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,\widetilde{\zeta}_i)-\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,\zeta_i^*)\right)$$

$$= \frac{2}{n^2}\sum_i \mathbf{E}\int \Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,v)\mu_n(dv)+\frac{2}{n^3}\sum_{i\neq j}\Gamma_{ij}^{(n)},$$

where $\Gamma_{ij}^{(n)} = \mathbf{E}\left(\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,\widetilde{\zeta}_j)-\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,\zeta_j^*)\right)$.

All we need now is to prove that $\Gamma_{ij}^{(n)} = o(1)$ uniformly on i and j. Denote

$$\phi(\widetilde{\zeta}_i,v) = \mathbf{E}\left(\Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,v)\mid \widetilde{\zeta}_i\right).$$

Then by (6)

$$\Gamma_{ij}^{(n)} = \mathbf{E}\left(\phi(\widetilde{\zeta}_i,\widetilde{\zeta}_j)-\phi(\widetilde{\zeta}_i,\zeta_j^*)\right)\underset{n\to\infty}{\longrightarrow} 0$$

and (9) takes place.

As $\|\mu_n-\mu\|\to 0$ and Φ is bounded on any compact set, then for any i

$$\mathbf{E}\int \Phi(\alpha_n;\Delta^{(i)},\widetilde{\zeta}_i,v)(\mu_n(dv)-\mu(dv)) = o(1)$$

and $J_1 = o(n^{-1})$. By the same way one may see that $J_3 = o(n^{-1})$. ∎

Let now $\mu\in H$, $\mathcal{L}(\zeta_i^{(n)}) = \mu$ and suppose that for any n the mutual distribution of random variables $\zeta_1^{(n)},\dots,\zeta_n^{(n)}$ is invariant under their permutation. Denote $\mathcal{L}(\zeta_i^{(n)},\zeta_j^{(n)}) = \mathcal{P}_2^{(n)}$, $i\neq j$, $\xi_n = \sum\limits_{i=1}^n \delta_{\zeta_i^{(n)}}/n$ and $\mathcal{W}_n = \mathcal{P}_2^{(n)}-\mu\otimes\mu$.

Theorem 3.1 1. *If $F\in \mathbf{C}^3(E)$ and $\|\mathcal{W}_n\| = o(1/n)$, then*

$$m_0(F) = \mathbf{E}F(\xi_n)-F(\mu) = \frac{1}{2n}A(F)+o(1/n), \tag{10}$$

where

$$A(F) = \int_D f_2(\mu;v,v)\mu(dv)-\int_{D^2} f_2(\mu;v_1,v_2)\mu(dv_1)\mu(dv_2). \tag{11}$$

2. *Let* $F \in \mathbf{C}^5(E)$. *Suppose that*

$$\mathcal{P}_2^{(n)} = \frac{n+1}{n}\,\mu \otimes \mu - \frac{1}{n}\,\nu_n + r_n, \tag{12}$$

where $\nu_n \xrightarrow{w} \mathcal{L}(\zeta, \zeta)$, $\mathcal{L}(\zeta) = \mu$ *and* $\|r_n\| = o(1/n)$. *Then* $m_0(F) = o(1/n)$.

Proof. 1. Denote $g_\tau(v_1, v_2) = f_2(\alpha_n; v_1, v_2)$. If we prove that

$$m_0(F) = \frac{1}{2n} A(F) + \int_0^1 (1-\tau) \int_{D^2} \mathbf{E} g_\tau \, d\mathcal{W}_n \, d\tau + o(1/n) \tag{13}$$

for $V_n = \|\mathcal{W}_n\| \to 0$, then the first part of the theorem would be demonstrated.

As $\mathcal{E}(\xi_n) = \mu$, the conditions of Lemma 3.1 are fulfilled and (7) takes place. Moreover $\mathbf{E}\partial^1 F(\mu; \xi_n - \mu) = 0$ and therefore

$$\begin{aligned}
m_0(F) &= \mathbf{E} \int_0^1 (1-\tau) \int_{D^2} g_\tau \, d\Big((\widetilde{\xi}_n - \mu) \otimes (\widetilde{\xi}_n - \mu) \Big) \, d\tau + o(1/n) \\
&= \mathbf{E} \int_0^1 (1-\tau) \Big(\frac{1}{n^2} \sum_{i,j} g_\tau(\widetilde{\zeta}_i, \widetilde{\zeta}_j) - \frac{2}{n} \sum_i \int_D g_\tau(\widetilde{\zeta}_i, v) \mu(dv) \\
&+ \int_{D^2} g_\tau(v_1, v_2) \mu(dv_1) \mu(dv_2) \Big) d\tau + o(1/n) \\
&= \mathbf{E} \int_0^1 (1-\tau) \Big(\frac{n-1}{n} g_\tau(\widetilde{\zeta}_1, \widetilde{\zeta}_2) + \frac{1}{n} g_\tau(\widetilde{\zeta}_1, \widetilde{\zeta}_1) \\
&- 2 \int_D g_\tau(\widetilde{\zeta}_1, v) \mu(dv) + \int_{D^2} g_\tau(v_1, v_2) \mu(dv_1) \mu(dv_2) \Big) d\tau + o(1/n).
\end{aligned}$$

Hence up to $o(n^{-1})$

$$\begin{aligned}
m_0(F) = \mathbf{E} \int_0^1 (1-\tau) \bigg(\int_{D^2} g_\tau(v_1, v_2) \mathcal{W}_n(dv_1 dv_2) + \\
+ \frac{1}{n} \Big(\int_D g_\tau(v, v) \mu(dv) - \int_{D^2} g_\tau(v_1, v_2) \mathcal{P}_2^{(n)}(dv_1 dv_2) \Big) \bigg) d\tau.
\end{aligned} \tag{14}$$

As $V_n \to 0$, the expression $\int \mathbf{E} g_\tau d\mathcal{P}_2^{(n)}$ in the second addend of the right-hand side of (14) may be replaced by $\int \mathbf{E} g_\tau d\mu^{\otimes 2}$.

It follows from the same condition $V_n \to 0$ that $\mathcal{L}(\xi_n) \stackrel{w}{\to} \delta_\mu$ (see Sznitman (1991)) and therefore $\mathbf{E} g_\tau(v_1, v_2) \to f_2(\mu; v_1, v_2)$ for all τ, v_1 and v_2 as $n \to \infty$. Thus we can replace $\mathbf{E} g_\tau$ by f_2 in the second addend of the right-hand side of (14) and (13) is proved.

2. First, $V_n = O(1/n)$ by (12), and (13) takes place. Let

$$\widetilde{F}(\nu) = \widetilde{F}(\nu; \tau, v_1, v_2) = f_2(\mu + \tau(\nu - \mu); v_1, v_2), \quad \widetilde{F} \in \mathbf{C}^3(E).$$

Accurate analysis of the remainder terms during the proofs of Lemma 3.1 and formula (13) shows us that

$$|m_0(\widetilde{F})| = |\mathbf{E} g_\tau(v_1, v_2) - f_2(\mu; v_1, v_2)| \le C/n,$$

where C does not depend on τ, v_1 and v_2, and is determined by supremums of $\widetilde{F}$ and its first three derivatives on the corresponding compact sets. Thus we can replace $\mathbf{E} g_\tau$ by f_2 in the right-hand side of (13).

Taking into account (12) we obtain that

$$\begin{aligned} m_0(F) &= \frac{1}{2n} A(F) + \frac{1}{2n} \int\limits_{D^2} f_2 \, d(\mu \otimes \mu - \nu_n) + o(1/n) \\ &= \frac{1}{2n} \left(\int\limits_{D} f_2(\mu; v, v) \mu(dv) - \int\limits_{D^2} f_2 \, d\nu_n \right) + o(1/n) \end{aligned}$$

and the proof is complete. ∎

The following proposition gives the form for the mean square deviation $s_0(F)$, $s_0^2(F) = \mathbf{E}\Big(F(\xi_n) - F(\mu)\Big)^2$, under the conditions of Theorem 3.1.

Corollary 3.1 1. *Under conditions of the first Theorem* 3.1 *proposition,*

$$s_0^2(F) = \frac{1}{n} \left(\int f_1^2(\mu; v) \mu(dv) - \left(\int f_1(\mu; v) \mu(dv) \right)^2 \right) + o(1/n),$$

where $f_1(\mu; v) = \partial^1 F(\mu; \delta_v)$.

2. *Under conditions of the second Theorem* 3.1 *proposition,* $s_0^2(F) = o(1/n)$.

Proof. As $s_0^2(F) = m_0(F^2) - 2F \cdot m_0(F)$, then by (10)

$$s_0^2(F) = \frac{1}{2n} \Big(A(F^2) - 2F \cdot A(F) \Big) + o(1/n).$$

Denote $\partial^2 F(v_1, v_2) = \partial^2 F(\mu; \delta_{v_1}, \delta_{v_2})$ and $\partial^1 F(v) = \partial^2 F(\mu; \delta_v)$. Then

$$\partial^2 F^2(v_1, v_2) = 2\partial^1 F(v_1)\partial^1 F(v_2) + 2F \cdot \partial^2 F(v_1, v_2)$$

and therefore

$$\begin{aligned} A(F^2) &= 2\int\limits_D \Big(\partial^1 F(v)\Big)^2 \mu(dv) + 2F \cdot \int\limits_D \partial^2 F(v, v)\mu(dv) \\ &-2\Big(\int\limits_D \partial^1 F(v)\mu(dv)\Big)^2 + 2F \cdot \int\limits_{D^2} \partial^2 F(v_1, v_2)\mu(dv_1)\mu(dv_2) \\ &= 2\left(\int\limits_D \Big(\partial^1 F(v)\Big)^2 \mu(dv) - \Big(\int\limits_D \partial^1 F(v)\mu(dv)\Big)^2\right) - 2F \cdot A(F). \end{aligned}$$

The second proposition is evident. ■

---

4 Stratification

4.1 General stratification scheme

The standard sufficient conditions for (10) and (11) are the conditions for $\zeta_i^{(n)}$ to be i.i.d. Let us formulate the sufficient conditions for (12), that gives the better order of convergence than in the case of independence.

Let $\mu \in H$ and suppose that for any $n \geq 1$ there exists such a decomposition $\mu = \sum\limits_{i=1}^{n} \mu_i^{(n)}/n$, that $\mu_i^{(n)} \in H$ and $\max\limits_{i \in I_n} \operatorname{diam}(D_{i,n}) \underset{n\to\infty}{\longrightarrow} 0$, where $D_{i,n} = \operatorname{support}(\mu_i^{(n)})$, $I_n \subset \mathcal{N}_n = \{1, 2, \ldots, n\}$ and $\operatorname{card}(\mathcal{N}_n \setminus I_n)/n \to 0$ as $n \to \infty$. Let $\widehat{\zeta}_i^{(n)}$ be independent and $\mathcal{L}(\widehat{\zeta}_i^{(n)}) = \mu_i^{(n)}$.

Corollary 4.1 *If $F \in \mathbf{C}^5(E)$, then $m_0(F) = o(1/n)$ for $\xi_n = \sum\limits_{i=1}^{n} \delta_{\widehat{\zeta}_i^{(n)}}/n$.*

Proof. Let π_n be such a random permutation of $1, \ldots, n$ that π_n does not depend on $(\widehat{\zeta}_1^{(n)}, \ldots, \widehat{\zeta}_n^{(n)})$ and $\mathcal{L}(\pi_n)$ is a uniform distribution on the whole permutation set. Denote $\zeta_i^{(n)} = \widehat{\zeta}_{\pi_n(i)}^{(n)}$. Surely, $\xi_n = \sum\limits_{i=1}^{n} \delta_{\zeta_i^{(n)}}/n$. Moreover,

$$\mathcal{L}(\zeta_k^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(\widehat{\zeta}_i^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \mu_i^{(n)} = \mu$$

and for $k \neq l$

$$\mathcal{L}\left(\zeta_k^{(n)}, \zeta_l^{(n)}\right) = \sum_{i \neq j} \frac{1}{n(n-1)} \mathcal{L}\left(\widehat{\zeta}_i^{(n)}\right) \otimes \mathcal{L}\left(\widehat{\zeta}_j^{(n)}\right)$$
$$= \frac{1}{n} \sum_{i=1}^{n} \mu_i^{(n)} \otimes \left(\frac{1}{n-1} \sum_{j: j \neq i}^{n} \mu_j^{(n)} \right)$$
$$= \frac{n}{n-1} \mu^{\otimes 2} - \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\mu_i^{(n)}\right)^{\otimes 2} = \frac{n+1}{n} \mu^{\otimes 2} - \frac{1}{n} \nu_n + o(1/n),$$

where

$$\nu_n = \frac{1}{n} \sum_{i=1}^{n} \left(\mu_i^{(n)}\right)^{\otimes 2}$$

All we need now is to prove that for any $\phi \in \mathbf{C}(D^2)$

$$\int_{D^2} \phi \, d\nu_n \underset{n \to \infty}{\longrightarrow} \int_{D} \phi(v,v) \mu(dv).$$

Surely, as $\operatorname{card}(\mathcal{N}_n \setminus I_n)/n = o(1)$,

$$\left| \int_{D^2} \phi \, d\nu_n - \int_{D} \phi(v,v) \mu(dv) \right|$$
$$\leq \frac{1}{n} \sum_{i \in I_n} \int_{D_{i,n}^2} \left| \phi(v_1, v_2) - \phi(v_1, v_1) \right| \mu_i^{(n)}(dv_1) \mu_i^{(n)}(dv_2) + o(1)$$
$$\leq \frac{1}{n} \sum_{i \in I_n} \sup_{v_1, v_2 \in D_{i,n}} \left| \phi(v_1, v_2) - \phi(v_1, v_1) \right| + o(1) = o(1).$$

The proof is complete. ■

4.2 Examples

The simplest example of a smooth functional is the linear functional

$$F(\mu) = \int_{D} f d\mu.$$

If $f \in \mathbf{C}(D)$, then, as it may be easily checked, $F \in \mathbf{C}^k(E)$ for all k, $\partial^1 F(\mu, v) = f(v)$ and $\partial^k F \equiv 0$ for $k \geq 2$.

Surely, the first proposition of the Theorem 3.1 is not interesting in this case, but the second proposition and the Corollary 4.1 are of interest because there is a hope to get more precise results for concrete D here.

Example 1. The Lebesgue measure. The simplest example for the Corollary 4.1 is the example of $D = [0, 1]$, $D_{i,n} = [(i-1)/n, i/n]$, where μ is the Lebesgue measure on $[0, 1]$ and $\mu_i^{(n)}$ stands for the uniform distribution on $D_{i,n}$.

Consider linear functional $F(\mu) = \int_{[0,1]} f d\mu$. Under supposition that f is measurable and bounded, we have $\mathbf{E}F(\xi_n) = F(\mu)$, and

$$\mathbf{D}F(\xi_n) = \frac{1}{n^2} \sum_{i=1}^{n} n \int\limits_{(i-1)/n}^{i/n} (f(x) - J_{i,n})^2 dx,$$

where

$$J_{i,n} = n \int\limits_{(i-1)/n}^{i/n} f(x) dx.$$

It may be easily seen that if f is continuous on $[0, 1]$ then $\mathbf{D}F(\xi_n) = o(n^{-1})$ (this is just the case of the Corollary 4.1).

If f has the first continuous derivative, then the variance can be calculated more explicitly. Let $x_i = x_{i,n} = i/n$ and suppose that $x \in D_{i,n}$. Then

$$f(x) = f(x_{i-1}) + (x - x_{i-1})\, f'(x_{i-1}) + o(n^{-1}),$$

where the remainder term is uniform on $x \in D_{i,n}$ and $1 \leq i \leq n$. Therefore

$$J_{i,n} = f(x_{i-1}) + \frac{1}{2n} f'(x_{i-1}) + o(n^{-1})$$

and

$$n \int\limits_{(i-1)/n}^{i/n} (f(x) - J_{i,n})^2 dx = \frac{1}{12n^2} \Big(f'(x_{i-1})\Big)^2 + o(n^{-2}).$$

Thus

$$\mathbf{D}F(\xi_n) = \frac{1}{12n^4}\sum_{i=1}^{n}\Big(f'(x_{i-1})\Big)^2 + o(n^{-3}) = \frac{\sigma^2(F)}{n^3} + o(n^{-3}),$$

where

$$\sigma^2(F) = \frac{1}{12}\int_0^1 \Big(f'(x)\Big)^2\,dx.$$

Note that such results can not be obtained in the general situation of the Corollary 4.1, as we did not suppose any linear structure for the space D there.

Let us demonstrate the Central Limit Theorem for this example. First,

$$\begin{aligned}
\eta_n = n^{3/2}(F(\xi_n) - F(\mu)) &= \sqrt{n}\sum_{j=1}^{n}\Big(f(\widehat{\zeta}_j^{(n)}) - J_{j,n}\Big)\\
&= \sqrt{n}\left(\sum_{j=1}^{n}\Big(\widehat{\zeta}_j^{(n)} - x_{j-1} - 1/2n\Big)\,f'(x_{j-1}) + o(n^{-1})\right)\\
&= \frac{1}{2\sqrt{n}}\sum_{j=1}^{n}\alpha_j f'(x_{j-1}) + o(n^{-1/2}) = \beta_n + o(n^{-1/2}),
\end{aligned}$$

where α_i are independent and uniformly distributed on $[-1,1]$.

Denote $\varphi_{\beta,n}$ the characteristic function of the random variable β_n. Then

$$\begin{aligned}
\varphi_{\beta,n}(t) &= \prod_{j=1}^{n}\frac{\sin(2n^{-1/2}f'(x_{j-1})t)}{2n^{-1/2}f'(x_{j-1})t}\\
&= \prod_{j=1}^{n}\left(1 - \frac{t^2}{24n}\left(f'(x_{j-1})\right)^2 + o(1/n)\right)\\
&= \exp\left(-t^2\sum_{j=1}^{n}\left(f'(x_{j-1})\right)^2/24n + o(1)\right) \underset{n\to\infty}{\longrightarrow} \exp(-\sigma^2(F)t^2/2).
\end{aligned}$$

Thus $\mathcal{L}(\eta_n) \xrightarrow{w} \mathbf{N}(0,\sigma^2(F))$.

Example 2. General one-dimensional case. Let $\mu \in H$ be arbitrary continuous distribution on $(\mathbf{R},\mathcal{B})$ with the distribution function $M(x)$. Suppose that $M(x) \neq M(y)$ for $x \neq y$. Let f be a bounded continuous function on $\mathbb{R}$, having finite limits at both $-\infty$ and $+\infty$. We are interested in the integral

$$F(\mu) = \int\limits_{\mathbf{R}} f d\mu.$$

Evidently, the problem may be considered at $\overline{\mathbb{R}} = \mathbf{R} \cup \{-\infty\} \cup \{+\infty\}$, which is a compact in the metrics $\rho(x, y) = |M(x) - M(y)|$. The natural extension of f on $\overline{\mathbf{R}}$ is continuous on this compact.

Let $x_{i,n}$ be such points that $\rho(x_{i+1,n}, x_{i,n}) = 1/n$, $x_{0,n} = -\infty$ and $x_{n,n} = +\infty$. Denote $D_i = D_{i,n} = [x_{i-1,n}, x_{i,n}]$ for $1 \le i \le n$. Consider independent random variables $\widehat{\zeta}_i^{(n)}$ with distributions $\mu_i^{(n)} = n\,\mu\Big|_{D_i}$. Then, evidently, $\mathbf{E}F(\xi_n) = F(\mu)$ and due the Corollary 4.1, $\mathbf{D}F(\xi_n) = o(1/n)$.

Note that this stratification scheme is equivalent to the Lebesgue measure example for estimation of the functional

$$F(\mu) = \int\limits_0^1 f \circ M^{-1}(x) dx.$$

Thus all results of the previous paragraph (including Central Limit Theorem) take place if $f \circ M^{-1} \in \mathbf{C}^1([0, 1])$.

Example 3. Discrete distributions. Let $D = \{1, 2, \ldots, N\}$, $\rho(i, j) = \delta_{ij}$ and $\mu(\{i\}) = p_i > 0$, $\sum_i p_i = 1$. Surely, every function $f : D \mapsto \mathbf{R}$ is continuous in metrics ρ. Consider

$$F(\mu) = \int\limits_D f d\mu = \sum_i f_i\, p_i.$$

Let $k_i = [np_i]$ and $m_i = k_1 + \ldots + k_i$. Evidently, $m_N \le n$. Denote $D_{i,n} = \{j\}$ for $m_{j-1} + 1 \le i \le m_i$ and $D_{i,n} = D$ for $m_N + 1 \le i \le n$. As for measures $\mu_i^{(n)}$, we take them concentrated at $D_{i,n}$ for $i \le m_N$ and defined by

$$\mu_i^{(n)}(\{j\}) = \frac{np_j - k_j}{n - m_N}$$

otherwise. Consider independent random variables $\widehat{\zeta}_i^{(n)}$ with $\mathcal{L}(\widehat{\zeta}_i^{(n)}) = \mu_i^{(n)}$ and denote $\xi_n = \sum_i \delta_{\widehat{\zeta}_i^{(n)}}/n$. Surely, $\mathbf{E}F(\xi_n) = F(\mu)$. As $n - m_N = \sum_{i=1}^N \{np_i\} \le N$, $\mathbf{D}F(\xi_n) = o(1/n)$ due to the Corollary 4.1.

The simplest example shows that generally there does not exist any variant of CLT here. Let us take $N = 3$ and $p_i = 1/3$. Then $F(\xi_n) \equiv F(\mu)$ for $n = 3l$,

$$F(\xi_n) = \frac{n-1}{n} F(\mu) + \frac{1}{n} f(\widehat{\zeta}_1^{(n)})$$

for $n = 3l + 1$ and

$$F(\xi_n) = \frac{n-2}{n} F(\mu) + \frac{1}{n} \Big(f(\widehat{\zeta}_1^{(n)}) + f(\widehat{\zeta}_2^{(n)}) \Big)$$

for $n = 3l + 2$, where $\mathbf{P}(\widehat{\zeta}_j^{(n)} = i) = 1/3$, $i = 1, 2, 3$, $j = 1, 2$.

Example 4. Multi-dimensional Lebesgue measure. Suppose $\mu \in H$ is the uniform distribution on $I^d = [0,1]^d$. Consider the case $d = 2$ as the example and show the way to choose measures $\mu_i^{(n)}$, satisfying the conditions of Corollary 4.1. Denote $m = [\sqrt{n}]$, where $[x]$ stands for the entire part of x, and $n_1 = m^2$. Let distributions $\nu_{jk}^{(n)}$, $1 \le j, k \le m$, be uniform on

$$J_{jk} = [(j-1)n^{-1/2}, jn^{-1/2}] \times [(k-1)n^{-1/2}, kn^{-1/2}] \, .$$

For $1 \le i \le n_1$ we choose $\mu_i^{(n)} = \nu_{jk}^{(n)}$ with some $(j,k) = \big(j(i), k(i)\big)$ such that $\mu_i^{(n)} \ne \mu_l^{(n)}$ for $i \ne l$. If $n_1 < i \le n$ we take $\mu_i^{(n)}$ as the uniform distribution on $I^2 \setminus \bigcup_{j_1, j_2 = 1}^{m} J_{j_1 j_2}$. Surely, $\max_{1 \le i \le n_1} \operatorname{diam} \operatorname{support} \mu_i^{(n)} \to 0$ and $n\mu = \sum_{i=1}^{n} \mu_i^{(n)}$. As $n - n_1 = O(n^{-1/2})$, the result of the Corollary 4.1 takes place.

Example 5. Stratification without stratification. Denote $D = (0,1]$ and $d(x,y) = |x - y|$. Consider the metric space (D, ρ), where $\rho(x,y) = \min\{d(x,y), 1 - d(x,y)\}$. Surely, ρ is the circle metrics and a function, continuous on this metrics is a usual continuous function on $(0,1]$ with $f(0+) = f(1)$. The space (D, ρ) is a compact space.

For fixed $n > 2$ consider random vector $\eta_n = (\zeta_1^{(n)}, \ldots, \zeta_n^{(n)})$ which is uniformly distributed in such $B_n \in D^n$ that $\rho(x_i, x_j) > 1/2n$ for all pairs $i \ne j$, $(x_1, \ldots, x_n) \subset B_n$. Evidently, the distribution $\mathcal{L}(\eta_n)$ does not depend on permutations of random variables.

If we put $x \oplus y = x + y \pmod 1$, and $\zeta_{i,c}^{(n)} = \zeta_i^{(n)} \oplus c$, $c > 0$, then the mutual distribution of $\zeta_{i,c}^{(n)}$ coincide with $\mathcal{L}(\eta_n)$ and therefore $\zeta_i^{(n)}$ is uniformly distributed on D for any i.

Let $\mathcal{P}_2^{(n)} = \mathcal{L}\left(\zeta_1^{(n)}, \zeta_2^{(n)}\right)$. Denote U_n the support of $\mathcal{P}_2^{(n)}$. Surely,

$$U_n = \{(x, y) \subset D^2 : \rho(x, y) > 1/2n\}.$$

Moreover $\mathcal{P}_2^{(n)}$ is the uniform distribution on U_n due to the equality $\mathcal{P}_2^{(n)} = \mathcal{L}\left(\zeta_{1,c}^{(n)}, \zeta_{2,c}^{(n)}\right)$.

Let μ be the Lebesgue measure on D. If we demonstrate the equality (12), we would have by the second proposition of the Theorem 3.1, that the variance of the unbiased estimate $F(\xi_n) = \sum_{i=1}^n f(\zeta_i^{(n)})/n$ for $F(\mu) = \int_D f d\mu$ has the order $o(n^{-1})$ if $f \in \mathbf{C}(D, \rho)$.

The two-dimensional Lebesgue measure of U_n is equal to $(n-1)/n$ and

$$\mathcal{P}_2^{(n)} - \frac{n+1}{n}\, \mu \otimes \mu = \mathcal{P}_2^{(n)} - \frac{n}{n-1}\, \mu^{\otimes 2} + o(n^{-1}) = \frac{1}{n-1}\, \nu_n + o(n^{-1}),$$

where ν_n is the uniform distribution on U_n^C. Surely, $\nu_n \xrightarrow{w} \mathcal{L}(\zeta, \zeta)$, $\mathcal{L}(\zeta) = \mu$.

Let us find more explicit expression for the variance. Let $f \in \mathbf{C}\big([0,1], d\big)$. As we may suppose that $\mathbf{E} f(\zeta_i^{(n)}) = 0$, then $\mathbf{E}F(\xi_n) = 0$ and

$$\mathbf{D}F(\xi_n) = \frac{1}{n}\int_0^1 f^2(x)dx + \frac{n-1}{n}\,\mathbf{E} f(\xi_1^{(n)}) f(\xi_2^{(n)})$$

with

$$\frac{n-1}{n}\,\mathbf{E} f(\xi_1^{(n)}) f(\xi_2^{(n)}) = \iint_{U_n} f(x)f(y)dxdy = -\iint_{U_n^C} f(x)f(y)dxdy. \quad (15)$$

Accurate analysis of the last integral in (15) shows that for $f \in \mathbf{C}^2\big([0,1], d\big)$

$$\mathbf{D}F(\xi_n) = \frac{1}{8n^2}\Big(f(0) - f(1)\Big)^2 + \frac{1}{24n^3}\left(\Big(f(0) - f(1)\Big)\Big(f'(0) + f'(1)\Big) + \int_0^1 \Big(f'(x)\Big)^2 dx\right) + o(n^{-3}).$$

Thus the procedure described above is generally worse than that of Example 1 if $f(0) \neq f(1)$ (then $f \notin \mathbf{C}(D, \rho)$), but gives the twice smaller variance if $f(0) = f(1)$.

References

1. Cartan, H. (1967). *Calcul différentiel. Formes différentielles*, Hermann Paris.

2. Sznitman, A.S. (1991). Topics in propagation of chaos, *Lecture Notes in Mathematics*, **1464**, Springer.

3. Bahvalov, N.S. (1964). On optimal estimation of convergence for quadrature processes and Monte Carlo-type integration methods in function classes, In *Computational methods for differential and integral equations and quadrature formulas*, Moscow, Nauka, 5–63 (in Russian).