Unit Stratified Sampling as a Tool for Approximation of Stochastic Optimization Problems*

Martin Šmíd

Institute of Information Theory and Automation of the ASCR, CZ182-08 Prague, Pod Vodárenskou věží 4, Czech Republic

e-mail: smid@utia.cas.cz

Abstract. We apply stratified sampling with equiprobable strata and a single observation drawn from each stratum to the approximate computation of stochastic programming problems. We determine the convergence rate of the approximation error both when computing expectations and when approximating stochastic programming problems.

Keywords: Stochastic programming, approximation, stratified sampling.

JEL classification: C44
AMS classification: 90C15, 01G60

1. Introduction

Since the expectations involved in stochastic programming problems are often impossible to express analytically, various approximation methods are used for their evaluation, one of which being the sample average approximation (Monte Carlo estimation) whose advantages are an easy computability and a convergence rate independent of the dimension of the "chance" (see e.g. [3, 7]). Moreover, various methods reducing the variance of Monte Carlo estimates exist (see [6]) suggesting themselves to be used for the approximation of stochastic programs, too (see e.g. [8]).

In the present paper, we consider a variant of stratified sampling which is a widely used variance reduction technique consisting in sampling from conditional distributions given sub-regions (strata) of the distribution's support. The specific of our variant is that we draw exactly one observation from each stratum and that the strata are equiprobable. Together with a multi-dimensional version of the inverse transform, which we formulate, our approach provides a simple and easily tractable tool to compute expectations (note that, generally, it is not easy to sample from regions of multidimensional distributions). We call our approach *unit stratified sampling* (USS).

*This research was supported by the Czech Science foundation under Grants 402/09/0965, P402/10/0956 and P402/11/0150.

As to our knowledge, unit stratified sampling was first suggested by an early paper [1] by Dupač who proved that the error of the USS approximation of an expectation with respect to a k-dimensional uniform random vector is $O(-(\frac{1}{2}+\frac{1}{2k}))$. In the present paper, we generalize the original method for non-uniform distributions with possibly unbounded support. For easy computation, we formulate a multi-dimensional version of the inverse transform allowing to use uniform random number generators when computing the estimates. We prove that the convergence rate of an approximation error of a Lipschitz function's expectation is $O(-(\frac{1}{2}+\frac{1-a}{k}))$ where a is a distribution-specific constant taking value 0 for the uniform distribution. Finally, we show that the error of an approximation of a stochastic program with a Lipschitz objective function inherits this convergence rate up to an arbitrary small addition.

2. Unit stratified sampling - one-dimensional case

Throughout all the present Section, let X be a random variable with a continuous cumulative distribution function (c.d.f.) F and a finite second moment. Let g be a Lipschitz function with the Lipschitz constant K. We compute the *unit stratified sampling (USS)* estimator of $\mathbb{E}g(X)$ as follows:

1. We partition the support of X into n unbroken regions $R_1, R_2, \ldots, R_n$ such that
$$\mathbb{P}[X \in R_i] = 1/n, \qquad 1 \leq i \leq n.$$
2. We draw n random observations, the i-th one being drawn from the conditional distribution of X given that $[X \in R_i]$
3. We map all the observations by g
4. We average the results.

Equivalently, the USS estimate may be expressed as

$$c_n(g) = \frac{1}{n}\sum_{i=1}^{n} g(X_i), \qquad X_i = F^{-1}(U_i) \tag{1}$$

where $U_1 \sim \mathrm{U}(0, \frac{1}{n}), U_2 \sim \mathrm{U}(\frac{1}{n}, \frac{2}{n}), \ldots, U_n \sim \mathrm{U}(\frac{n-1}{n}, 1)$ are independent ($\mathrm{U}(a,b)$ stands for the uniform distribution on $[a,b]$) and where

$$F^{-1}(\bullet) = \inf\{e \in \mathbb{R} : F(e) \geq \bullet\}$$

denotes the quantile function of F.

Proposition 1. *$c_n(g)$ is an unbiased estimate of $\mathbb{E}g(X)$ with*

$$\operatorname{var} c_n(g) \leq K^2(\operatorname{var} X - \operatorname{var} Y_n) = K^2(\mathbb{E}(X^2 - Y_n^2)) \tag{2}$$

where Y_n is a random variable independent of $X_1, X_2, \ldots, X_n$ such that

$$\mathbb{P}[Y_n = \mathbb{E}(X_i)] = \frac{1}{n}, \qquad 1 \leq i \leq n.$$

The equality is attainable.

Proof. According to [5], p. 238 (see also Proposition 6 in the present text) the distribution of X will not change if $X = F^{-1}(U)$ where $U \sim \mathrm{U}(0,1)$ hence

$$\mathbb{E}c_n(g) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}g(F^{-1}(U_i)) = \frac{1}{n}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n} g(F^{-1}(u))ndu$$
$$= \int_0^1 g(F^{-1}(u))du = \mathbb{E}g(F^{-1}(U)) = \mathbb{E}g(X).$$

Further, by a textbook formula,

$$\operatorname{var} c_n(g) = \frac{1}{n}\sum \operatorname{var} g(X_i).$$

Now, consider a variable

$$Z = \sum_{i=1}^{n} I_{\{\mathbb{E}X_i\}}(Y_n)X_i$$

where I is the indicator function. An easy calculation shows that Z equals to X in distribution and, by the Law of total variance ([13], p. 385),

$$\operatorname{var}(Z) = \mathbb{E}\operatorname{var}(Z|Y_n) + \operatorname{var}(\mathbb{E}(Z|Y_n)) = \frac{1}{n}\sum_{i=1}^{n}\operatorname{var}(X_i) + \operatorname{var}(Y_n). \tag{3}$$

Now, since

$$\operatorname{var} g(X_i) \le K^2 \operatorname{var} X_i$$

(see Lemma 4) with the equality for $g(x) = Kx$, we have

$$\operatorname{var} c_n(g) \le K^2\frac{1}{n}\sum_{i=1}^{n}\operatorname{var}(X_i)$$

which, in combination with (3), proves the "$\le$" in (2). The "$=$" of (2) follows from the fact that $\mathbb{E}Y = \mathbb{E}X$. □

Remark 2. Since, evidently, $Y_n \to X$ in distribution and since the second moment of X is finite, it follows from (2) that $\operatorname{var} c_n(g) \to 0$ which implies (e.g., by Chebyshev inequality) consistency of the USS estimator.

Remark 3. Once the USS estimator is defined by (1), Proposition 1 holds even for discontinuous F.

3. Multidimensional case

Let $\boldsymbol{X}$ be a k-dimensional random vector with a general, possibly discontinuous, distribution such that the second moments of all its marginal distributions are finite. Let $g : \mathbb{R}^k \to \mathbb{R}$ is l_1-K-Lipschitz,[1] i.e.

$$|g(\boldsymbol{x}) - g(\boldsymbol{y})| \le K\sum_{i=1}^{k}|x^i - y^i|, \qquad \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^k.$$

[1] We work with the l_1 norm for technical reasons. A reader wanting to work with the other norms can use the fact that all the norms are equivalent in finite dimensional space.

Lemma 4 (bound of variance of l_1-K-Lipschitz functions).

$$\sup_{g \text{ is } l_1\text{-}K\text{-Lipschitz}} \operatorname{var} g(\boldsymbol{X}) \leq \overline{V}^2 \qquad \overline{V}^2 = K^2 \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \sqrt{v_{i_1,i_1} v_{i_2,i_2}}$$

where $(v_{i_1,i_2})_{i_1=1\ldots k, i_2=1\ldots k}$ is the variance matrix of $\boldsymbol{X}$.

Proof. Let g be l_1-K-Lipschitz. Then

$$\begin{aligned}
\operatorname{var}(g(\boldsymbol{X})) = \mathbb{E}(g(\boldsymbol{X}) - \mathbb{E}g(\boldsymbol{X}))^2 &= \mathbb{E}(g(\boldsymbol{X}) - g(\mathbb{E}\boldsymbol{X}))^2 - \mathbb{E}(g(\mathbb{E}\boldsymbol{X}) - \mathbb{E}g(\boldsymbol{X}))^2 \\
&\leq \mathbb{E}(g(\boldsymbol{X}) - g(\mathbb{E}\boldsymbol{X}))^2 \leq \mathbb{E}(K \sum_{i=1}^{k} |\boldsymbol{X}^i - \mathbb{E}\boldsymbol{X}^i|)^2 \\
= K^2 \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \mathbb{E}|\boldsymbol{X}^{i_1} - \mathbb{E}\boldsymbol{X}^{i_2}||\boldsymbol{X}^{i_2} - \mathbb{E}\boldsymbol{X}^{i_2}| &\leq K^2 \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \sqrt{\operatorname{var} \boldsymbol{X}^{i_1} \operatorname{var} \boldsymbol{X}^{i_2}} = \overline{V}^2
\end{aligned}$$

(we have used the Schwarz inequality at the last $\leq$). □

Remark 5. It follows from the the Lemma and the Schwarz inequality that $\mathbb{E}g(\boldsymbol{X})$ exists and is finite.

3.1. Inverse transform

The next Proposition generalizes the well known random number generation technique called inverse-transform (see [6]) to multiple dimensions. Even though this extension is straightforwared, the author is not aware of its publication in such a general form allowing multi-dimensional, possibly discontinuous, distribution functions.

Proposition 6 (multi-dimensional inverse transform). *Let $\boldsymbol{U}$ be a k-dimensional random vector with the uniform distribution on $[0,1] \times [0,1] \times \cdots \times [0,1]$ and let $\boldsymbol{Y}$ be a k-dimensional real random vector. Denote $G_1(\bullet)$ the c.d.f. of $\boldsymbol{Y}^1$ and, for each $i = 2, 3, \ldots, k$, denote $G_i(\bullet|\boldsymbol{Y}^1, \boldsymbol{Y}^2, \ldots, \boldsymbol{Y}^{i-1})$ the conditional c.d.f. of $\boldsymbol{Y}^i$ given $(\boldsymbol{Y}^1, \boldsymbol{Y}^2, \ldots, \boldsymbol{Y}^{i-1})$. Define*

$$\begin{gathered}
\boldsymbol{t}_{\boldsymbol{Y}} : \prod_{i=1}^{k} [0,1] \to \mathbb{R}^k, \\
\boldsymbol{t}_{\boldsymbol{Y}}^1(\boldsymbol{u}) = G_1^{-1}(\boldsymbol{u}^1), \\
\boldsymbol{t}_{\boldsymbol{Y}}^i(\boldsymbol{u}) = G_i^{-1}(\boldsymbol{u}^i | \boldsymbol{t}_{\boldsymbol{Y}}^1(\boldsymbol{u}), \boldsymbol{t}_{\boldsymbol{Y}}^2(\boldsymbol{u}), \ldots, \boldsymbol{t}_{\boldsymbol{Y}}^{i-1}(\boldsymbol{u})), \qquad i = 2, 3, \ldots, k.
\end{gathered}$$

Then $\boldsymbol{t}_{\boldsymbol{Y}}$ is measurable with respect to the Borel σ-algebra on the k-dimensional unit cube and

$$\boldsymbol{t}_{\boldsymbol{Y}}(\boldsymbol{U}) \stackrel{d}{=} \boldsymbol{Y} \tag{4}$$

where $\stackrel{d}{=}$ denotes the equality in distribution.

For the proof, the following Lemma will be needed:

Lemma 7. *Let G be a one-dimensional c.d.f. Then $G^{-1}(\alpha) \leq x$ if and only if $G(x) \geq \alpha$.*

Proof (Lemma 7). If $G(x) \geq \alpha$ then it must be $G^{-1}(\alpha) = \inf\{e : G(e) \geq \alpha\} \leq x$. If, on the other hand, $G^{-1}(\alpha) = \inf\{e : G(e) \geq \alpha\} \leq x$ then, for each $x_\nu > x$, $x_\nu \to x$, it has to hold that $G(x_\nu) \geq \alpha$ (otherwise $G^{-1}(\alpha) > x$) so $G(x) = \lim_\nu G(x_\nu) \geq \alpha$ by a limit transition. □

Proof (Proposition 6). As $\mathbb{R}^k$ is a complete separable metric space, $\mathcal{L}(\boldsymbol{Y}^i|\boldsymbol{Y}^1, \ldots, \boldsymbol{Y}^{i-1})$ exists according to [12], VI.1.21., hence the definition of $\boldsymbol{t}_{\boldsymbol{Y}}(\boldsymbol{U})$ is correct.

Further, we proceed by induction: From Lemma 7,

$$\left(\boldsymbol{t}_{\boldsymbol{Y}}^1\right)^{-1}((-\infty, x]) = \{u : u \leq G_1(x)\} = [0, G_1(x)]$$

for each $x \in \mathbb{R}$ and, from the one dimensional inverse transform (see [5], p. 238),

$$\boldsymbol{t}_{\boldsymbol{Y}}^1(\boldsymbol{U}^1) \stackrel{\mathrm{d}}{=} \boldsymbol{Y}^1,$$

i.e. $\boldsymbol{t}_{\boldsymbol{Y}}^1$ is measurable and (4) holds for $k = 1$.
Induction step: Let $k > 1$. Assume that vector mapping $(\boldsymbol{t}_{\boldsymbol{Y}}^1, \boldsymbol{t}_{\boldsymbol{Y}}^2, \ldots, \boldsymbol{t}_{\boldsymbol{Y}}^{k-1})'$ is measurable with respect to Borel σ-algebra of $k-1$-dimensional unit cube (hence with respect to Borel σ-algebra of k-dimensional unit cube) and that

$$(\boldsymbol{t}_{\boldsymbol{Y}}^1(\boldsymbol{U}^1), \boldsymbol{t}_{\boldsymbol{Y}}^2(\boldsymbol{U}^1, \boldsymbol{U}^2), \ldots, \boldsymbol{t}_{\boldsymbol{Y}}^{k-1}(\boldsymbol{U}^1, \ldots, \boldsymbol{U}^{k-1})) \stackrel{\mathrm{d}}{=} (\boldsymbol{Y}^1, \ldots, \boldsymbol{Y}^{k-1}). \tag{5}$$

Ad. the measurability. It suffices to show that $(\boldsymbol{t}_{\boldsymbol{Y}}^k)^{-1}((-\infty, t])$ is Borel set. Denote

$$\bar{\boldsymbol{u}}^\nu = (\boldsymbol{u}^1, \boldsymbol{u}^2, \ldots, \boldsymbol{u}^\nu).$$

It holds that

$$\begin{aligned}(\boldsymbol{t}_{\boldsymbol{Y}}^k)^{-1}((-\infty, t]) &= \{\boldsymbol{u} : \boldsymbol{t}_{\boldsymbol{Y}}^k(\boldsymbol{u}) \leq t\} = \{\boldsymbol{u} : G_k^{-1}(\boldsymbol{u}^k | \boldsymbol{t}_{\boldsymbol{Y}}^1(\bar{\boldsymbol{u}}^1), \ldots, \boldsymbol{t}_{\boldsymbol{Y}}^{k-1}(\bar{\boldsymbol{u}}^{k-1})) \leq t\} \\ &\stackrel{\text{Lemma 7}}{=} \{\boldsymbol{u} : G_k(t | \boldsymbol{t}_{\boldsymbol{Y}}^1(\bar{\boldsymbol{u}}^1), \ldots, \boldsymbol{t}_{\boldsymbol{Y}}^{k-1}(\bar{\boldsymbol{u}}^{k-1})) \geq \boldsymbol{u}^k\} \\ &= \{\boldsymbol{u} : G_k(t | \boldsymbol{t}_{\boldsymbol{Y}}^1(\bar{\boldsymbol{u}}^1), \ldots, \boldsymbol{t}_{\boldsymbol{Y}}^{k-1}(\bar{\boldsymbol{u}}^{k-1})) - \boldsymbol{u}^k \geq 0\}. \end{aligned} \tag{6}$$

Since $G_k(t|\bullet)$ is the conditional probability of Borel set $(-\infty, t]$, it is measurable according to [12], VI.1.1. Therefore and due to the facts that $(\boldsymbol{t}^1, \ldots, \boldsymbol{t}^{k-1})$ is measurable by induction hypothesis, that the superposition of two measurable mappings is measurable, that the identity is measurable and that the sum of two measurable functions is measurable, the function defining the set (6) is measurable so the set is Borel.
Ad. (4). Denote $Q = \mathcal{L}(\boldsymbol{t}_{\boldsymbol{Y}}(\boldsymbol{U}))$, $R = \mathcal{L}(\boldsymbol{Y})$,

$$\bar{\boldsymbol{t}}_{\boldsymbol{Y}}(u_1, u_2, \ldots, u_{k-1}) := (\boldsymbol{t}_{\boldsymbol{Y}}^1(u_1), \ldots, \boldsymbol{t}_{\boldsymbol{Y}}^{k-1}(u_1, \ldots, u_{k-1}))',$$

$\bar{Q} = \mathcal{L}(\bar{\boldsymbol{t}}_{\boldsymbol{Y}})$ and $\bar{R} = \mathcal{L}(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_{k-1})$. According to the induction assumption, $\bar{R} = \bar{Q}$. To get $R = Q$ it suffices to show that $Q[C] = R[C]$ for each $C = A \times (-\infty, b)$ where $A \subseteq \mathbb{R}^{k-1}$ is Borel set and $b \in \mathbb{R}$.

Denote H_i the i-dimensional unit cube. Gradually we get

$$
\begin{aligned}
Q[C] &= \int_C dQ(\boldsymbol{\tau}) = \int_{\mathbb{R}^k} I_C(\boldsymbol{\tau})dQ(\boldsymbol{\tau}) \\
&= \int_{H_k} I_C(\boldsymbol{t}_{\boldsymbol{Y}}(u_1,u_2,\ldots,u_k))du_1du_2\ldots du_k \\
&= \int_{H_k} I_{(-\infty,b)}(\boldsymbol{t}^k_{\boldsymbol{Y}}(u_1,u_2,\ldots,u_k))I_A(\bar{\boldsymbol{t}}_{\boldsymbol{Y}}(u_1,u_2,\ldots,u_{k-1}))du_1du_2\ldots du_k \\
&= \int_{H_{k-1}} \left[\int_{[0,1]} I_{(-\infty,b)}(\boldsymbol{t}^k_{\boldsymbol{Y}}(u_1,u_2,\ldots,u_k))du_k\right] \\
&\qquad\qquad \cdot I_A(\bar{\boldsymbol{t}}_{\boldsymbol{Y}}(u_1,u_2,\ldots,u_{k-1}))du_1du_2\ldots du_{k-1} \\
&= \int_{H_{k-1}} \left[\int_{[0,1]} I_{(-\infty,b)}(G_k^{-1}(u_k|\bar{\boldsymbol{t}}_{\boldsymbol{Y}}(u_1,u_2,\ldots,u_{k-1})))du_k\right] \\
&\qquad\qquad \cdot I_A(\bar{\boldsymbol{t}}_{\boldsymbol{Y}}(u_1,u_2,\ldots,u_{k-1}))du_1du_2\ldots du_{k-1} \\
&\overset{\bar{Q}=\bar{R}}{=} \int_{\mathbb{R}^{k-1}} \left[\int_{[0,1]} I_{(-\infty,b)}(G_k^{-1}(u_k|t_1,t_2,\ldots,t_{k-1}))du_k\right] \\
&\qquad\qquad \cdot I_A(t_1,t_2,\ldots,t_{k-1})d\bar{R}(t_1,t_2,\ldots t_{k-1}) \\
&= \int_A \left[\int_{[0,1]} I_{(-\infty,b)}(G_k^{-1}(u_k|t_1,t_2,\ldots,t_{k-1}))du_k\right] d\bar{R}(t_1,t_2,\ldots,t_{k-1}) \\
&\overset{\text{inv. trans.}}{=} \int_A \left[\int_{\mathbb{R}} I_{(-\infty,b)}(t_k)dG_k(t_k|t_1,t_2,\ldots,t_k))\right] d\bar{R}(t_1,t_2,\ldots t_{k-1}) \\
&= \int_A G_k(b|t_1,t_2,\ldots,t_{k-1})d\bar{R}(t_1,t_2,\ldots,t_{k-1}) = R[C]
\end{aligned}
$$

by the definition of conditional probability. □

3.2. USS estimate

Denote $m_n = \lfloor \sqrt[k]{n} \rfloor$. The approximation of $\mathbb{E}g(\boldsymbol{X})$ by the *multidimensional unit stratified sampling* consists in the following steps:

1. We partition the k-dimensional unit cube to m_n^k identical cubes.
2. We choose one observation from the uniform distribution on each of the cubes.
3. We transform our observations using the superposition of the mapping $\boldsymbol{t}_{\boldsymbol{X}}$ with g.
4. We average the results.

Mathematically speaking, our estimator equals to

$$C_n(g) := \frac{1}{m_n^k}\sum_{i_1=1}^{m_n}\sum_{i_2=1}^{m_n}\cdots\sum_{i_k=1}^{m_n} g(\boldsymbol{t}_{\boldsymbol{X}}(\boldsymbol{U}_{i_1,i_2,\ldots,i_k}))$$

where $\boldsymbol{U}_{i_1,i_2,\ldots,i_k}$ is a random vector having uniform distribution on the cube $[\frac{i_1-1}{m_n},\frac{i_1}{m_n}]\times[\frac{i_2-1}{m_n},\frac{i_2}{m_n}]\times\cdots\times[\frac{i_k-1}{m_n},\frac{i_k}{m_n}]$ for each $i_1 = 1,2,\ldots,m_n, i_2 = 1,2,\ldots,m_n,\ldots,i_k = 1,2,\ldots,m_n$, such that all the vectors $\boldsymbol{U}_{i_1,i_2,\ldots,i_k}$ are mutually independent.

Denote

$$\gamma_n(g) := C_n(g) - \mathbb{E}g(\boldsymbol{X}) \tag{7}$$

the error of the approximation. Let $\boldsymbol{U}$ be random vector with k-dimensional uniform distribution defined on the unit cube. Then we can write

$$\begin{aligned}
\gamma_n(g) &\overset{\text{Proposition 6}}{=} C_n(g) - \mathbb{E}(g(\boldsymbol{t}_{\boldsymbol{X}}(\boldsymbol{U}))) = C_n(g) - \int_{[0,1]\times\cdots\times[0,1]} g(\boldsymbol{t}_{\boldsymbol{X}}(\boldsymbol{u}))d\boldsymbol{u} \\
&= \frac{1}{m_n^k}\sum_{i_1=1}^{m_n}\sum_{i_2=1}^{m_n}\cdots\sum_{i_k=1}^{m_n} \\
&\qquad \left[g(\boldsymbol{t}_{\boldsymbol{X}}(\boldsymbol{U}_{i_1,i_2,\ldots,i_k})) - \int_{[\frac{i_1-1}{m_n},\frac{i_1}{m_n}]\times\cdots\times[\frac{i_k-1}{m_n},\frac{i_k}{m_n}]} g(\boldsymbol{t}_{\boldsymbol{X}}(\boldsymbol{u}))m_n^k d\boldsymbol{u}\right] \\
&= \frac{1}{m_n^k}\sum_{i_1=1}^{m_n}\sum_{i_2=1}^{m_n}\cdots\sum_{i_k=1}^{m_n}\left[g(\boldsymbol{t}_{\boldsymbol{X}}(\boldsymbol{U}_{i_1,i_2,\ldots,i_k})) - \mathbb{E}g(\boldsymbol{t}_{\boldsymbol{X}}(\boldsymbol{U}_{i_1,i_2,\ldots,i_k}))\right] \qquad (8)
\end{aligned}$$

immediately giving

$$\mathbb{E}\gamma_n(g) = 0$$

Now, let us examine the convergence rate of $\gamma_n(g)$.

Theorem 8 (convergence rate). *Let g be l_1-K-Lipschitz. Denote $F_1(\bullet)$ the marginal c.d.f. of the $\boldsymbol{X}^1$ and $F_\lambda(\bullet|x_1,\ldots,x_{\lambda-1})$ the conditional c.d.f. of $\boldsymbol{X}^\lambda$ given $(\boldsymbol{X}^1 = x_1,\ldots,, \boldsymbol{X}^{\lambda-1} = x_{\lambda-1})$ for each $1 < \lambda \leq k$. Let*

$$\frac{\partial}{\partial u}F_\lambda^{-1}(u|x_1,x_2,\ldots,x_{\lambda-1})$$

exist for each $u \in (0,1)$, for each $x_1,x_2,\ldots,x_{\lambda-1} \in \mathbb{R}$ and each $1 \leq \lambda \leq k$ and let there exist constants $C > 0$ and $0 \leq a \leq 1$ such that

$$\frac{\partial}{\partial u}F_\lambda^{-1}(u|x_1,x_2,\ldots,x_{\lambda-1}) \leq Cu^{-a} \tag{9}$$

for each $0 < u \leq 1/2$, $x_1,x_2,\ldots,x_{\lambda-1} \in \mathbb{R}$, $1 \leq \lambda \leq k$, and

$$\frac{\partial}{\partial u}F_\lambda^{-1}(u|x_1,x_2,\ldots,x_{\lambda-1}) \leq C(1-u)^{-a} \tag{10}$$

for each $1/2 \leq u < 1$, $x_1,x_2,\ldots,x_{\lambda-1} \in \mathbb{R}$, $1 \leq \lambda \leq k$. Then

$$\gamma_n(g) = O_P(n^{-\left(\frac{1}{2}+\frac{1-a}{k}\right)}).$$

(i.e. for each $\epsilon > 0$ there exists a constant M_ϵ such that $\limsup_{n\to\infty} P[|\gamma_n(g)| \geq n^{-(\frac{1}{2}+\frac{1-a}{k})}M_\epsilon] \leq \epsilon$).

Remark 9. The following distributions fulfil the assumptions of Theorem 8:
(i) Uniform distribution with $a = 0$
(ii) Any distribution with a compact support and a density f such that there exists $c > 0$ fulfilling $f(x) \geq c$ for each x from the support, with $a = 0$
(iii) Exponential and normal distributions with $a = 1$ but not with any $a < 1$.

Proof (Remark). (i) and (ii) are straightforward, for (iii), see [10] Sec. 1.2.3. □

In order to prove the Theorem, let us state the upper bound of the variance of the approximation error first:

Proposition 10 (bounds of variance). *Let $n \geq 2^k$. Then, under the assumptions of Theorem 8,*

$$\operatorname{var}(\gamma_n(g)) \leq \begin{cases} \left(\frac{1}{3}\frac{(m_n-2)^k}{m_n^k} + \frac{1}{(2-a)(1-a)}\frac{m_n^k-(m_n-2)^k}{m_n^k}\right) m_n^{2a-2-k} K^2 k^2 C^2 & \text{if } a < 1 \\ \left(2 - \frac{5(m_n-2)^k}{3m_n^k}\right) m_n^{-k} K^2 k^2 C^2 & \text{if } a = 1. \end{cases} \tag{11}$$

Proof (Proposition 10). (i) Introduction. Re-index random vectors $\boldsymbol{U}_{1,1,\ldots,1}, \boldsymbol{U}_{1,1,\ldots,2}, \ldots, \boldsymbol{U}_{m_n,m_n,\ldots,m_n}$ as $\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots, \boldsymbol{U}_{m_n^k}$ and denote H_i the support of $\mathcal{L}(\boldsymbol{U}_i)$ for each $i = 1, 2, \ldots, m_n^k$. By (8),

$$\gamma_n(g) = m_n^{-k} \sum_{i=1}^{m_n^k} Z_i \tag{12}$$

where $Z_i = g(\boldsymbol{t_X}(\boldsymbol{U}_i)) - Eg(\boldsymbol{t_X}(\boldsymbol{U}_i))$.

(ii) Ad. $0 \leq a < 1$. We start with the second moments of $\boldsymbol{t}_{\boldsymbol{X}}^\lambda(\boldsymbol{U}_i)$. Fix i and λ and assume that the projection of H_i into the dimension λ is $[q/m_n, (q+1)/m_n]$ for some integer q fulfilling $0 \leq q < m_n/2$.

Denote $H_i^{\lambda-1}$ the projection of H_i into the first $\lambda - 1$ dimensions. Using the fact that $\min_c \mathbb{E}(X-c)^2$ happens for $c = EX$ for each random variable X we get that

$$\begin{aligned} &\operatorname{var}(\boldsymbol{t}_{\boldsymbol{X}}^\lambda(\boldsymbol{U}_i)) \\ &\quad = \int_{H_i^{\lambda-1}} \int_{q/m_n}^{(q+1)/m_n} [F_\lambda^{-1}(u_\lambda | \boldsymbol{t}_{\boldsymbol{X}}^1(u_1), \ldots, \boldsymbol{t}_{\boldsymbol{X}}^{\lambda-1}(u_1, \ldots, u_{\lambda-1})) - \mathbb{E}(\boldsymbol{t}_{\boldsymbol{X}}^\lambda(\boldsymbol{U}_i))]^2 \\ &\qquad\qquad m_n^\lambda du_\lambda du_{\lambda-1} \ldots du_1 \\ &\quad \leq \int_{H_i^{\lambda-1}} J(\boldsymbol{t}_{\boldsymbol{X}}^1(u_1), \ldots, \boldsymbol{t}_{\boldsymbol{X}}^{\lambda-1}(u_1, \ldots, u_{\lambda-1})) m_n^{\lambda-1} du_{\lambda-1} \ldots du_1 \end{aligned} \tag{13}$$

where

$$J(t_1, \ldots, t_{\lambda-1}) = \int_{q/m_n}^{(q+1)/m_n} [F_\lambda^{-1}(u | t_1, \ldots, t_{\lambda-1}) - F_\lambda^{-1}((q+1)/m_n | t_1, \ldots, t_{\lambda-1})]^2 m_n du. \tag{14}$$

For brevity, we shall write J instead of $J(\boldsymbol{t}_{\boldsymbol{X}}^1(u_1), \ldots, \boldsymbol{t}_{\boldsymbol{X}}^{\lambda-1}(u_1, \ldots, u_{\lambda-1}))$ and $F_\lambda^{-1}(\bullet)$ instead of $F_\lambda^{-1}(\bullet | \boldsymbol{t}_{\boldsymbol{X}}^1(u_1), \ldots, \boldsymbol{t}_{\boldsymbol{X}}^{\lambda-1}(u_1, \ldots, u_{\lambda-1}))$ in the following text.

If $q = 0$ then we may use inequality

$$F_\lambda^{-1}(v) - F_\lambda^{-1}(u) = \int_u^v (F_\lambda^{-1})'(x) dx \leq \int_u^v C x^{-a} dx = \frac{C}{1-a}(v^{1-a} - u^{1-a}) \tag{15}$$

to estimate

$$\begin{aligned}
J &= \int_0^{1/m_n} [F_\lambda^{-1}(1/m_n) - F_\lambda^{-1}(u)]^2 m_n du \\
&\leq \left(\frac{C}{1-a}\right)^2 \int_0^{1/m_n} [(1/m_n)^{1-a} - u^{1-a}]^2 m_n du \\
&= \left(\frac{C}{1-a}\right)^2 \int_0^{1/m_n} [(1/m_n)^{2-2a} - 2(1/m_n)^{1-a}u^{1-a} + u^{2-2a}] m_n du \\
&= \left(\frac{C}{1-a}\right)^2 m_n \left[u(1/m_n)^{2-2a} - 2(1/m_n)^{1-a}\frac{u^{2-a}}{2-a} + \frac{u^{3-2a}}{3-2a}\right]_0^{1/m_n} \\
&= \left(\frac{C}{1-a}\right)^2 m_n \left[1 - \frac{2}{2-a} + \frac{1}{3-2a}\right](1/m_n)^{3-2a} \\
&= \left(\frac{C}{1-a}\right)^2 \left[1 - \frac{2}{2-a} + \frac{1}{3-2a}\right] m_n^{2a-2} \\
&= \left(\frac{C}{1-a}\right)^2 \left[\frac{-a}{2-a} + \frac{1}{3-2a}\right] m_n^{2a-2} \\
&\leq \left(\frac{C}{1-a}\right)^2 \left[\frac{-a}{2-a} + \frac{1}{2-a}\right] m_n^{2a-2} \\
&= \left(\frac{C}{1-a}\right)^2 \left(\frac{1-a}{2-a}\right) m_n^{2a-2} = C^2 \frac{1}{(2-a)(1-a)} m_n^{2a-2} \qquad (16)
\end{aligned}$$

(we have used fact that $3-2a \geq 2-a$ for $a \leq 1$ at the last "$\leq$") which implies that

$$\begin{aligned}
\mathrm{var}(\boldsymbol{t}_{\boldsymbol{x}}^\lambda(\boldsymbol{U}_i)) &\overset{(13),(16)}{\leq} C^2 \frac{1}{(2-a)(1-a)} m_n^{2a-2} \int_{H_i^{\lambda-1}} m_n^{\lambda-1} du_{\lambda-1}\dots du_1 \\
&= \frac{C^2}{(2-a)(1-a)} m_n^{2a-2} \qquad (17)
\end{aligned}$$

for $q = 0$.

Let $1 \leq q \leq m_n/2 - 1$. Using the Mean Value Theorem we get that, for each $0 < u < v < \frac{1}{2}$,

$$F_\lambda^{-1}(v) - F_\lambda^{-1}(u) = (F_\lambda^{-1})'(b)(v-u) \qquad (18)$$

for some $u \leq b \leq v$. Therefore and due to the assumptions of the present Proposition, it holds that

$$F_\lambda^{-1}(v) - F_\lambda^{-1}(u) \leq Cb^{-a}(v-u) \leq Cu^{-a}(v-u). \qquad (19)$$

Using it we get

$$
\begin{aligned}
J &= \int_{q/m_n}^{(q+1)/m_n} [F_\lambda^{-1}((q+1)/m_n) - F_\lambda^{-1}(u)]^2 m_n du \\
&\leq \int_{q/m_n}^{(q+1)/m_n} \left[C u^{-a} \left(\frac{q+1}{m_n} - u \right) \right]^2 m_n du \\
&\leq \int_{q/m_n}^{(q+1)/m_n} \left[C \left(\frac{q}{m_n} \right)^{-a} \left(\frac{q+1}{m_n} - u \right) \right]^2 m_n du \\
&= C^2 \left(\frac{q}{m_n} \right)^{-2a} \int_0^{1/m_n} v^2 m_n dv \\
&= C^2 \left(\frac{q}{m_n} \right)^{-2a} m_n \frac{m_n^{-3}}{3} \\
&\leq C^2 \left(\frac{1}{m_n} \right)^{-2a} m_n \frac{m_n^{-3}}{3} = \frac{C^2}{3} m_n^{2a-2}
\end{aligned}
$$

so that

$$
\mathrm{var}(\boldsymbol{t}_{\mathbf{x}}^{\lambda}(\boldsymbol{U}_i)) \overset{(13)}{\leq} \frac{C^2}{3} m_n^{2a-2} \tag{20}
$$

for $1 \leq q \leq m_n/2$.

Finally, if m_n is odd and $q = m_n/2 - 1/2$ then, similarly to (13),

$$
\mathrm{var}(\boldsymbol{t}_{\mathbf{x}}^{\lambda}(\boldsymbol{U}_i)) \leq \int_{H_i^{\lambda-1}} \int_{1/2-1/(2m_n)}^{1/2+1/(2m_n)} [F_\lambda^{-1}(u_\lambda) - F_\lambda^{-1}(1/2)]^2 m_n du_\lambda du_{\lambda-1} \dots du_1.
$$

Analogically to (20), we get that

$$
\int_{1/2-1/(2m_n)}^{1/2} [F_\lambda^{-1}(u) - F_\lambda^{-1}(1/2)]^2 m_n du \leq \frac{1}{2} \frac{C^2}{3} m_n^{2a-2}
$$

and, using the symmetry, that

$$
\int_{1/2}^{1/2+1/(2m_n)} [F_\lambda^{-1}(u) - F_\lambda^{-1}(1/2)]^2 m_n du \leq \frac{1}{2} \frac{C^2}{3} m_n^{2a-2}.
$$

so that (20) holds also for m_n odd and $q = m_n/2 - 1/2$.

It can be easily proved using the symmetry that (20) holds for $m_n/2 \leq q \leq m_n - 2$ and that (17) holds for $q = m_n - 1$. Therefore and due to the previous calculations

$$
\mathrm{var}(\boldsymbol{t}_{\mathbf{x}}^{\lambda}(\boldsymbol{U}_i)) \leq \begin{cases} \frac{1}{(2-a)(1-a)} C^2 m_n^{2a-2} & \text{if } q = 0 \text{ or } q = m_n - 1 \\ \frac{1}{3} C^2 m_n^{2a-2} & \text{otherwise.} \end{cases} \tag{21}
$$

It is clear that if $q = 0$ or $q = m_n - 1$ then the cube H_i touches the boundary of the unit cube. Hence, if H_i does not touch the boundary of the unit cube then $0 < q < m_n - 1$ so that we get that

$$
\mathrm{var}(\boldsymbol{t}_{\mathbf{x}}^{\lambda}(\boldsymbol{U}_i)) \leq \frac{1}{3} C^2 m_n^{2a-2}
$$

if H_i does not touch the boundary of the unit cube, and

$$\operatorname{var}(\boldsymbol{t}^\lambda_{\boldsymbol{x}}(\boldsymbol{U}_i)) \leq \max\left(\frac{1}{(2-a)(1-a)}, \frac{1}{3}\right) C^2 m_n^{2a-2} = \frac{1}{(2-a)(1-a)} C^2 m_n^{2a-2}$$

if H_i touches the boundary of the unit cube. Now we can use Lemma 4 to obtain

$$\begin{aligned}
\operatorname{var}(Z_i) &\leq K^2 \sum_{i_1=1}^{k}\sum_{i_2=1}^{k} \sqrt{\operatorname{var}(\boldsymbol{t}_P^{i_1}(\boldsymbol{U}_i))\operatorname{var}(\boldsymbol{t}_P^{i_2}(\boldsymbol{U}_i))} \\
&\leq \begin{cases} K^2k^2\frac{1}{(2-a)(1-a)}C^2m_n^{2a-2} & \text{if } H_i \text{ touches the boundary of the unit cube} \\ K^2k^2\frac{1}{3}C^2m_n^{2a-2} & \text{otherwise.}\end{cases}
\end{aligned} \tag{22}$$

Since exactly $(m_n-2)^k$ cubes H_i do not touch the boundary of the unit cube in any dimension, we may estimate

$$\begin{aligned}
\sum_{i=1}^{m_n^k} \operatorname{var} Z_i &\leq (m_n-2)^k K^2k^2\frac{1}{3}C^2m_n^{2a-2} + [m_n^k-(m_n-2)^k]K^2k^2\frac{1}{(2-a)(1-a)}C^2m_n^{2a-2} \\
&= \left(\frac{1}{3}(m_n-2)^k + \frac{1}{(2-a)(1-a)}[m_n^k-(m_n-2)^k]\right) m_n^{2a-2}K^2k^2C^2
\end{aligned} \tag{23}$$

so that

$$\begin{aligned}
\operatorname{var}(\gamma_n(g)) &= \operatorname{var}\left(m_n^{-k}\sum_{i=1}^{m_n^k} Z_i\right) = m_n^{-2k}\sum_{i=1}^{m_n^k}\operatorname{var}(Z_i) \\
&\overset{(23)}{\leq} \left(\frac{1}{3}(m_n-2)^k + \frac{1}{(2-a)(1-a)}[m_n^k-(m_n-2)^k]\right) m_n^{2a-2-2k}K^2k^2C^2
\end{aligned}$$

which proves (11) for $a<1$.
(iii) Ad. $a=1$. We shall proceed analogously to *(ii)*: Let $1 \leq q \leq m_n/2-1$. It follows from the Mean Value Theorem and from (9) that for each $0<u<v\leq 1/2$ there exist $u \leq b \leq v$ such that

$$F_\lambda^{-1}(v) - F_\lambda^{-1}(u) = (F_\lambda^{-1})'(b)(v-u) \leq Cb^{-1}(v-u) \leq Cu^{-1}(v-u). \tag{24}$$

Therefore

$$\begin{aligned}
J &= \int_{q/m_n}^{(q+1)/m_n} [F_\lambda^{-1}((q+1)/m_n) - F_\lambda^{-1}(u)]^2 m_n du \\
&\overset{(24)}{\leq} \int_{q/m_n}^{(q+1)/m_n} \left[Cu^{-1}\left(u-\frac{q}{m_n}\right)\right]^2 m_n du \\
&\leq \int_{q/m_n}^{(q+1)/m_n} \left[C\left(\frac{1}{m_n}\right)^{-1}\left(u-\frac{q}{m_n}\right)\right]^2 m_n du \\
&= C^2\left(\frac{1}{m_n}\right)^{-2}\int_0^{m_n^{-1}} v^2 m_n dv = C^2m_n^2m_n\left[v^3/3\right]_0^{m_n^{-1}} = C^2/3
\end{aligned}$$

(J is defined by (14)).

Let $q = 0$. Since it holds that

$$F_\lambda^{-1}(v) - F_\lambda^{-1}(u) = \int_u^v (F_\lambda^{-1})'(x)dx \le \int_u^v Cx^{-1}dx = C(\ln v - \ln u) \qquad (25)$$

we have

$$\begin{aligned} J &= \int_0^{1/m_n} [F^{-1}(1/m_n) - F_\lambda^{-1}(u)]^2 m_n du \\ &\le C^2 \int_0^{1/m_n} [\ln(1/m_n) - \ln u]^2 m_n du \\ &\overset{u=\frac{\exp v}{m_n}}{=} C^2 \int_{-\infty}^0 [\ln(1/m_n) - (v + \ln(1/m_n))]^2 m_n \frac{\exp v}{m_n} dv \\ &= C^2 \int_{-\infty}^0 v^2 \exp v\, dv \\ &= C^2 \left(\left[v^2 \exp v\right]_{-\infty}^0 - [2v \exp v]_{-\infty}^0 + 2\int_{-\infty}^0 \exp v\, dv \right) \\ &= 2C^2. \end{aligned}$$

Hence, after handling the middle interval in case of odd m_n and using the symmetry the same way as in *(ii)*,

$$\mathrm{var}(\boldsymbol{t}_{\mathbf{x}}^\lambda(\boldsymbol{U}_i) \overset{(13)}{\le} \begin{cases} 2C^2 & \text{if } q = 0 \text{ or } q = m_n - 1 \\ \frac{1}{3}C^2 & \text{otherwise.} \end{cases}$$

and, consequently,

$$\mathrm{var}(Z_i) \le \begin{cases} K^2k^2 2C^2 & \text{if the cube } H_i \text{ touches the boundary of the unit cube} \\ K^2k^2\frac{1}{3}C^2 & \text{otherwise} \end{cases} \qquad (26)$$

so that

$$\begin{aligned} \mathrm{var}\left(\gamma_n(g)\right) &= \mathrm{var}\left(m_n^{-k} \sum_{i=1}^{m_n^k} Z_i \right) = m_n^{-2k} \sum_{i=1}^{m_n^k} \mathrm{var}(Z_i) \\ &\overset{(26)}{\le} m_n^{-2k} \left(\frac{1}{3}(m_n - 2)^k + 2[m_n^k - (m_n - 2)^k] \right) K^2k^2C^2 \\ &= m_n^{-2k} \left(2m_n^k - \frac{5}{3}(m_n - 2)^k \right) K^2k^2C^2 \\ &= m_n^{-k} \left(2 - \frac{5(m_n - 2)^k}{3m_n^k} \right) K^2k^2C^2. \end{aligned}$$

which proves (11) for $a = 1$. □

Proof (Theorem 8). If $a = 1$ then $\frac{2-2a+k}{2k} = \frac{1}{2}$ and we get using the Chebyshev inequality

$$\begin{aligned} P(|n^{1/2}\gamma_n(g)| \ge \epsilon) &\le \frac{\mathrm{var}(n^{1/2}\gamma_n(g))}{\epsilon^2} \overset{(11)}{\le} \frac{nm_n^{-k}}{\epsilon^2} \left(2 - \frac{5(m_n - 2)^k}{3m_n^k} \right) A \\ &\le \frac{(m_n + 1)^k}{\epsilon^2 m_n^k} 2A \overset{n\to\infty}{\longrightarrow} \epsilon^{-2} 2A \end{aligned}$$

for some constant $A \in \mathbb{R}$.

If $a < 1$ then there exists constant $B > 0$ such that

$$\begin{aligned}\operatorname{var}\left(n^{\frac{2-2a+k}{2k}}\gamma_n(g)\right) &= n^{\frac{2-2a+k}{k}}\operatorname{var}\left(\gamma_n(g)\right)\\ &\overset{(11)}{\leq} \frac{\left(\frac{1}{3}\frac{(m_n-2)^k}{m_n^k} + \frac{1}{(2-a)(1-a)}\frac{m_n^k-(m_n-2)^k}{m_n^k}\right) m_n^{2a-2-k}B}{(m_n+1)^{2a-2-k}}\\ &\overset{n\to\infty}{\longrightarrow} B/3\end{aligned}$$

hence, we may use the Chebyshev inequality to get the assertion of the Theorem. □

4. USS approximation of stochastic programming problems

Consider a *stochastic programming problem*, i.e., to find (one of the solutions) $\hat{\xi} \in \mathcal{X}$ of a problem

$$\min_{\xi \in \mathcal{X}} \mathbb{E}g(\xi, \boldsymbol{X}) \tag{27}$$

where $\boldsymbol{X}$ is a k-dimensional random vector, and that the problem (27) is approximated by

$$\min_{\xi \in \mathcal{X}} C_n(g(\xi, \bullet)) \tag{28}$$

(recall that $C_n(g(\xi, \bullet)$ is the USS approximation of $\mathbb{E}g(\xi, \boldsymbol{X})$).

Assume that

(a) $\mathcal{X}$ is a compact subset of $\mathbb{R}^r$

(b) for each $\xi \in \mathcal{X}$, $g(\xi, \bullet)$ is l_1-Lipschitz with a constant K, not depending on ξ

(c) for each $\boldsymbol{x} \in \mathbb{R}^k$, $g(\bullet, \boldsymbol{x})$ is l_2-Lipschitz with a constant L, not depending on $\boldsymbol{x}$

(d) for each $\xi \in \mathcal{X}$ and $\boldsymbol{x} \in \mathbb{R}^k$, $|g(\xi, \boldsymbol{x}) - \mathbb{E}g(\xi, \boldsymbol{X})| < H$ for some $H > 0$ not depending on $\boldsymbol{x}$ and ξ.

Remark 11. The assumption (d) if fulfilled if the support of $\boldsymbol{X}$ is compact of if the function g is bounded.

As we do not suppose the random variables $U_\bullet$, used to compute the approximation, to be able anticipate the realization of the "chance" in reality, we naturally assume $U_\bullet$ to be independent of $\boldsymbol{X}$.

4.1. Expected loss

We evaluate the error of the approximation by the *expected loss*

$$\eta(\dot{\xi}) := W(\dot{\xi}) - W(\hat{\xi}), \qquad W(\xi) = \mathbb{E}g(\xi, \boldsymbol{X}) \tag{29}$$

where $\hat{\xi}$ is a solution of (27) and $\dot{\xi}$ is a (chosen) solution of (28) (both the solutions are assumed to exist). The expected loss is a very natural measure of the inaccuracy of the approximation because it evaluates the loss caused by the usage of the "approximate" solution instead of the "true" one (see [11] for a more extensive discussion on the expected loss). Note that, thanks to the optimality of $\hat{\xi}$, the expected loss is always non-negative. The following result is well known (see e.g. [4]):

Lemma 12. *For each optimal solution $\dot{\xi}$ of (28) and each optimal solution $\hat{\xi}$ of (27),*

$$\eta(\dot{\xi}) \leq 2 \sup_{\xi \in \mathcal{X}} |C_n(g(\xi, \bullet)) - \mathbb{E}(g(\xi, \bullet))|$$

(note that the relation is valid for any realization of the random variables).

Proof. see [4].

4.2. Convergence rate of the USS approximation

Theorem 13. *Let $\boldsymbol{X}$ fulfil the assumptions given in Theorem 8 and let $a \geq 1 - k/2$. Then*

$$\eta(\dot{\xi}_n) = o_P\left(n^{-\left(\frac{1}{2}+\frac{1-a}{k}\right)+\delta}\right)$$

for any $\delta > 0$ where, for each $n \in \mathbb{N}$, $\dot{\xi}_n$ is an arbitrary solution of (28) (we write $X_n = o_P(n^a)$ if there exists $M > 0$ such that $\lim_{n\to\infty} P[|X_n| \geq n^a M] = 0$).

Before we prove the Theorem, let us cite suitable version of a "large deviation inequality" which we then use in a subsequent Lemma.

Proposition 14 (*[9], "Bennet's Inequality 3.", pp. 951-2,*[2] *and Proposition 1., p. 441). Let $X_1, X_2, \ldots, X_n$ be independent random variables with $X_i \leq b$, $\mathbb{E}X_i = 0$ and $\operatorname{var} X_i = \sigma_i^2$, for $i = 1, 2, \ldots, n$. Let $\sigma^2 = (\sigma_1^2 + \cdots + \sigma_n^2)/n$ and $\bar{X} = (X_1 + \cdots + X_n)/n$. Then for all $x \geq 0$*

$$\mathbb{P}\left[\sqrt{n}\bar{X} \geq x\right] \leq \exp\left\{-\frac{x^2}{2\sigma^2}\psi\left(\frac{xb}{\sigma^2\sqrt{n}}\right)\right\}$$

where

$$\psi(y) = (2/y^2)[(1+y)\log(1+y) - y].$$

The function ψ has the following properties:

(a) *$\psi(y)$ in non-increasing for $y \geq -1$,*

(b) *$y\psi(y)$ is non-decreasing for $y \geq -1$,*

(c) *$\psi(y) \geq \frac{1}{1+y/3}$ for $y \geq -1$.*

Lemma 15. *Let the c.d.f.'s of $\boldsymbol{X}$ fulfil the assumptions of the Theorem 8. Denote*

$$\gamma_n(\xi) := C_n(g(\xi, \bullet)) - \mathbb{E}g(\xi, \boldsymbol{X}).$$

If $a \geq 1 - k/2$ then, for each $\eta > 0$, there exists a constant $C_\eta > 0$, independent of ξ and n, such that

$$\mathbb{P}\left[n^{\frac{1}{2}+\frac{1-a}{k}-\delta}|\gamma_n(\xi)| \geq \eta\right] \leq 2\exp\left\{-C_\eta n^\delta\right\}$$

for each $n > 1$, $\xi \in \mathcal{X}$ and $\delta > 0$.

---

[2] There is a misprint in the original text. The correct version follows from the proof.

Proof (Lemma 15). Denote $\kappa := \frac{1-a}{k}$ and $N = \lfloor \sqrt[k]{n} \rfloor^k$. Recall that

$$\gamma_n(\xi) = \frac{1}{N}\sum_{i=1}^{N} Z_{n,i}, \qquad Z_{n,i} = g(\xi, \boldsymbol{t_x}(\boldsymbol{U}_{n,i})) - \mathbb{E}g(\xi, \boldsymbol{t_x}(\boldsymbol{U}_{n,i}))$$

where $\boldsymbol{U}_{n,1}, \boldsymbol{U}_{n,2}, \dots, \boldsymbol{U}_{n,N}$ are some independent uniform random variables and where

$$\mathbb{E}Z_{n,i} = 0, \qquad \operatorname{var} Z_{n,i} \leq KBN^{-2\kappa}. \tag{30}$$

for some constant B independent of i and ξ (see (12), (26) and (22), recall that K is the Lipschitz constant introduced by the assumption (b) of the present Section). By putting $\bar{X} = \gamma_n(\xi)$, $\lambda = \eta n^{-\kappa+\delta}$ and $b = H$ into Proposition 14 we get

$$\mathbb{P}\left(n^{\kappa+1/2-\delta}\gamma_n(\xi) \geq \eta\right) \leq \exp\left\{-\frac{n^{-2\kappa+2\delta}\eta^2}{2\sigma^2}\psi\left(\frac{n^{-\kappa+\delta}\eta H}{\sigma^2\sqrt{n}}\right)\right\}$$

where $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} \operatorname{var} Z_{n,i}$ and ψ is defined at Proposition 14. Since

$$\sigma^2 \overset{(30)}{\leq} N^{-2\kappa}V \leq (n^{1/k}-1)^{-2k\kappa}V,$$

where $V = KB$, and since $x\psi(x)$ is non-decreasing function according to (b) of Proposition 14, we have

$$\begin{aligned}
\mathbb{P}\left(n^{\kappa+1/2-\delta}\gamma_n(\xi) \geq \eta\right) &\leq \exp\left\{-\frac{n^{-2\kappa+2\delta}\eta^2}{2[n^{1/k}-1]^{-2k\kappa}V}\psi\left(\frac{n^{-\kappa+\delta}\eta H}{(n-1)^{-2\kappa}V\sqrt{n}}\right)\right\} \\
&\overset{\text{(a) of Pr. 14}}{\leq} \exp\left\{-\frac{n^{-2\kappa+2\delta}\eta^2}{2[n^{1/k}-1]^{-2k\kappa}V}\psi\left(\frac{n^{-\kappa+\delta}\eta H}{n^{-2\kappa}V\sqrt{n}}\right)\right\} \\
&= \exp\left\{-\left(\frac{n^{1/k}-1}{n^{1/k}}\right)^{2k\kappa} n^{2\delta}\frac{\eta^2}{2V}\psi\left(n^{\kappa+\delta-1/2}\frac{\eta H}{V}\right)\right\} \\
&\overset{\text{(c) of Th. 14}}{\leq} \exp\left\{-\left(1-\frac{1}{n^{1/k}}\right)^{2k\kappa} n^{2\delta}\frac{\eta^2}{2V\left[1+n^{\kappa+\delta-1/2}\frac{\eta H}{3V}\right]}\right\} \\
&= \exp\left\{-\left(1-\frac{1}{n^{1/k}}\right)^{2k\kappa} n^{\delta}\frac{\eta^2}{2V\left[n^{-\delta}+n^{\kappa-1/2}\frac{\eta H}{3V}\right]}\right\} \\
&\leq \exp\left\{-\left(1-\frac{1}{n^{1/k}}\right)^{2k\kappa} n^{\delta}\frac{\eta^2}{2V\left[1+\frac{\eta H}{3V}\right]}\right\}
\end{aligned}$$

because $\kappa - 1/2 \leq 0$ according to the assumption of the present Lemma. Hence

$$\begin{aligned}
\mathbb{P}\left(n^{\kappa+1/2-\delta}|\gamma_n(\xi)| \geq \eta\right) &\leq 2\exp\left\{-\left(1-\frac{1}{n^{1/k}}\right)^{2k\kappa} n^{\delta}\frac{\eta^2}{2\left[V+\frac{\eta H}{3}\right]}\right\} \\
&\leq 2\exp\left\{-n^{\delta}C_\eta\right\}
\end{aligned}$$

where $C_\eta = \left(1-\frac{1}{\sqrt[k]{2}}\right)^{2k\kappa}\frac{\eta^2}{2\left[V+\frac{\eta H}{3}\right]}$. $\square$

Proof (Theorem 13). Let $\eta > 0$. We shall proceed similarly to [3]. Denote $A = \frac{2-2a+k}{2k}$. Since $g(\bullet, \boldsymbol{x})$ is Lipschitz with the constant L, the function $\gamma_n(\bullet)$ is Lipschitz with the constant $2L$ so that

$$\theta_n(\bullet) := n^{A-\delta}\gamma_n(\bullet)$$

is Lipschitz with the constant $2n^{A-\delta}L$.

Using a procedure, identical to the proof of Theorem 2 from [2],[3] we may show that for each $n \in \mathbb{N}$ there exists $\xi_1, \xi_2, \ldots, \xi_{N(\eta,n)}$ with $N(\eta, n) \leq (D_\eta 2Ln^{A-\delta} + 1)^k$ for some constant D_η, independent of n, such that

$$\mathbb{P}\left[\max_{\xi\in\mathcal{X}} |\theta_n(\xi)| \geq \eta\right] \leq \mathbb{P}\left[\max_{i=1,2,\ldots,N(\eta,n)} |\theta_n(\xi_i)| \geq \frac{\eta}{3}\right].$$

Further, we may estimate

$$\begin{aligned}
\mathbb{P}\left[\max_{i=1,2,\ldots,N(\eta,n)} |\theta_n(\xi_i)| \geq \frac{\eta}{3}\right] \quad &\leq \quad \sum_{i=1}^{N(\eta,n)} \mathbb{P}\left[|\theta_n(\boldsymbol{x}_i)| \geq \frac{\eta}{3}\right] \\
&\overset{\text{Lemma 15}}{\leq} \quad \sum_{i=1}^{N(\eta,n)} 2\exp\left\{-n^\delta C_{\frac{\eta}{3}}\right\} = N(\eta,n) 2\exp\left\{-n^\delta C_{\frac{\eta}{3}}\right\} \\
&\leq \quad (D_\eta 2Ln^{A-\delta} + 1)^k 2\exp\left\{-n^\delta C_{\frac{\eta}{3}}\right\} \overset{n\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Since

$$0 \leq n^{A-\delta}\eta(\dot{\xi}_n) \overset{\text{Lemma 12}}{\leq} n^{A-\delta} 2\max_{\xi\in\mathcal{X}} |\gamma_n(\xi)| = 2\max_{\xi\in\mathcal{X}} |\theta_n(\xi)|$$

we have that

$$\mathbb{P}\left[n^{A-\delta}\eta(\dot{\xi}_n) \geq \eta\right] \overset{n\to\infty}{\longrightarrow} 0.$$

□

5. Conclusion

We have applied one of the variance-reduction techniques to the approximate computation of stochastic programming problems. As the convergence rate of a Monte Carlo estimate is $O(n^{-1/2})$, our method is able, given some conditions, to speed up the rate of convergence of the approximation error significantly. Even if we have discussed only one-stage problems here, the method may be applied to the multi-stage case, too; we have omitted these results here for space reasons, interested readers may, however, refer [10], Sec. 3.2.3.

References

[1] V. Dupač. Stochastic numerical methods (in czech). *Časopis pro pěstování matematiky*, 81:55–67, 1956.

[3] The author works with an i.i.d. random sample but the part of the proof used here does not depend on this fact.

[2] V. Kaňková. An approximative solution of a stochastic optimization problem. In *Trans. Eighth Prague Conference*, pages 234–244. Academia, Prague, 1978.

[3] V. Kaňková and P. Lachout. Convergence rate of empirical estimates in stochastic programming. *Informatica*, 3(4):497–523, 1992.

[4] G. Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Math. Program. ser. B*, 89(2):251–271, 2001.

[5] D. Pollard. *A User's Guide to Measure Theoretic Probability*. Cambridge Univ. Press, Cambridge, 2002.

[6] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo Method*. John Wiley & Sons, 2007.

[7] A. Shapiro. Stochastic programming by monte carlo simulation methods. *Stochastic Programming E-print series (dochost.rz.hu-berlin.de/speps/)*, 2000.

[8] A. Shapiro. Monte carlo sampling methods. In A. Ruszczynski and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 353 – 425. Elsevier, 2003.

[9] G. R. Shorak and J. A. Wellner. *Empirical Processes with Applications to Statistics*. J. Wiley, New York, 1986.

[10] M. Šmíd. *On Approximation of Stochastic Programming Problems*. PhD thesis, Charles University, department of probability and statistics, 2004.

[11] M. Šmíd. Expected loss in the discretization of multistage stochastic programming problems - estimation and convergence rate. *Annals of Operations Research*, 165(1):29–45, 2009. DOI=10.1007/s10479-008-0355-9.

[12] J. Štěpán. *Probability Theory (in Czech)*. Academia, Praha, 1987.

[13] N.A. Weiss, P.T. Holmes, and M. Hardy. *A course in probability*. Pearson Addison Wesley, 2006.