Charles University in Prague,
Faculty of Mathematics and Physics,
Department of Probability and Mathematical Statistics

ON APPROXIMATION OF STOCHASTIC PROGRAMMING PROBLEMS

Doctoral Thesis

Mgr. Martin Šmíd

Advisor:
RNDr. Vlasta Kaňková, CSc.

Prague, 2004

Branch m5 - Econometrics and Operations Research

Univerzita Karlova v Praze, Matematicko-fyzikální fakulta
Katedra pravděpodobnosti a matematické statistiky

PŘIBLIŽNÉ ŘEŠENÍ ÚLOH STOCHASTICKÉHO PROGRAMOVÁNÍ

Disertační práce

Mgr. Martin Šmíd

Školitel:
RNDr. Vlasta Kaňková, CSc.

Praha, 2004

Obor m5 - Ekonometrie a operační výzkum

Acknowledgement

First of all, I would like to thank my advisor, RNDr. Vlasta Kaňková, CSc., for a patient leadership and for valuable advises concerning my work.

Then, I would also like to thank all the members of the Department of Probability and Mathematical Statistics and to many of my colleagues from the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic for their effective help. Especially, I would like to mention Prof. RNDr. Jitka Dupačová, DrSc., Prof. RNDr. Václav Dupač, DrSc., Doc. RNDr. Jiří Outrata, DrSc., Prof. RNDr. Josef Štěpán, DrSc., and Doc. Ing. Miloslav Vošvrda, CSc.

Finally, I want to thank my partner Karolina for the corrections of the English text and for her patience.

Prague, 12th August 2004 .....

Mgr. Martin Šmíd

Contents

Introduction

Stochastic Programming

As reported by [4], the history of stochastic programming started at 1955 when the first theoretical paper by G. B. Dantzig was published. In the next decade, stochastic programming found its most popular application - the portfolio selection problem - for which its author H. Markowitz got the Nobel price.

From then, stochastic programming went through a significant development - there are many researchers being engaged in the discipline all around the world. An international conference devoted to stochastic programming is held each three years. Several monographs on stochastic programming were published (e.g. [2], [17] or [38]). The bibliography of stochastic programming kept by M. H. van der Vlerk counts up to 4000 citations (see [47]). Some parts of the current state of the art of the discipline may be found in [40].

During the last fifty years, stochastic programming found many successful applications mostly in finance, but also in transportation, energetics, hydrology, military planning and many other areas.

Stochastic Programming Problem

A rather general definition of the stochastic programming problem is

$$\inf_{x\in X, P[h(x,\zeta)\geq 0]\geq 1-\alpha} Eg(x,\zeta) \tag{1}$$

where X is a set, ζ is a random element, $h(\bullet,\bullet)$ and $g(\bullet,\bullet)$ are some functions, $\alpha \in [0,1]$ is some constant and E denotes the (mathematical) expectation. However, since it is difficult to deal with the problems of type (1) and, anyway, the problems with probabilistic constraints are not used much in practice, we will concentrate on the problems of the type

$$\inf_{x\in X} Eg(x,\zeta) \tag{2}$$

where X does not depend on the probability measure, in the present work.[1]

A special (but frequently used) class of stochastic programming problems are multistage programming problems. Here, the function g is defined by another multistage stochastic minimization problem (see formula (3.1) of the present work for the exact definition).

[1] Problems of type (1) may be converted to (2) using extended arithmetics.

Solution of Problems with Discrete Distribution

If the distribution of the random element in problem (2) is discrete with finite number of atoms then the expectation in (2) has the form of a finite sum, hence the solution may be found using methods of (deterministic) optimization.

However, if (2) is a multistage problem (with a discrete distribution at the first-stage problem as well as at all its nested problems) then we would need to solve many (typically nonlinear) problems to get the values of g. Therefore, the multistage problems with a discrete distribution are usually transformed to the non-recursive form[2] before they are solved (the reformulation usually leads to nonlinear optimization problems with a very large number of constraints).

If (2) is a *linear* multistage programming problem (see [8] for the definition), then its solution leads to a large (deterministic) linear programming problem. Hence, it may be solved using the simplex algorithm, by means of an interior point method or, most preferably, using one of the algorithms exploiting the special structure of the problem (the most popular are the methods based on Benders decomposition).

For a survey concerning the solution of multistage stochastic programming problems with a discrete distribution, see [10], chp. II.8. and references therein.

Solution of "Unsolvable" Problems

In many cases, however, the stochastic programming problems may not be solved using the methods mentioned above - it usually happens when the distribution of ζ in (2) is either continuous (then we are typically unable to evaluate the expectation) or discrete with a large number of atoms (then the solution of the problem exceeds the capacity of the present computers). In both situations, the problem has to be solved approximately.

Two approaches to the solution of these problems have appeared: the methods based on an approximation of the gradient (see [13], for instance) and the approach consisting in an approximation of the original distribution by a simpler one. The present work deals with the second approach. Particularly, three different methods of the original distribution approximation are examined and compared. Two of them have been widely discussed in the literature: the replacement of the original distribution by a discrete one (we call it discretization) and the replacement of the expectation $Eg(x,\zeta)$ by $n^{-1}\sum_{i=1}^{n} g(x,\zeta_i)$ where $\zeta_1, \zeta_2, \ldots, \zeta_i$ is a random sample from $\mathcal{L}(\zeta)$ either generated by a random number generator (Monte Carlo sample) or observed from the past (empirical estimate). The third method discussed in our work - a combination the discretization and the Monte Carlo sampling - seems to be a new method of stochastic programming approximation.

Discretization

The accuracy of the discretization of the stochastic programming problem (applied to the two-stage linear programming problem) was first studied by Kall (in [15]). His results were further generalized by Kaňková in [21] (the work deals with two stage non-linear programming) and in [25] (the results are extended to the multistage stochastic programming problem with Markov interstage dependence). To improve the accuracy of the approximation, Kall, Ruszczyński and Frauendorfer designed a sequential approximating method (see [16]). In order to get the optimal

[2] See [8] for the discussion on the multistage programming problem re-formulations.

discretization, Pflug (in [36]) suggests to approximate the original distribution by the discrete one minimizing the Wasserstein distance to the original distribution. To make problems with huge discrete distribution solvable, Dupačová, Gröve and Römisch (in [9]) introduce a technique of optimal reduction of large scenarios (i.e. discrete distributions with a special structure, see definition 1.2. of the present work) to smaller one.

Monte Carlo (Empirical estimating)

The idea of empirical estimation of stochastic programming problems appeared independently at [48] of Wets and [19] of Kaňková. The subsequent work [20] of Kaňková gives a probabilistic upper bound of the approximation error (this result is further generalized by Kaniovski, King and Wets in [18]). Kaňková further extends her results to the case of time dependent samples (in [22]) and to the case of multistage stochastic programming problems with Markov dependence of data (in [26]). Dupačová and Wets (in [12]) obtain the asymptotic normality of stochastic programming problem's empirical estimates. The convergence of the Monte Carlo approximation of multistage problems is studied by Shapiro (in [43]).

Another approach

A new approach consisting in using of the deterministic quasi-random sequences to approximate the distribution of the original problem has been introduced by Pennanen and Koivu (see [35]).

Remark

More details on the current state of the art and further citations may be found in the text of the present work.

The Present Work

Chapter 1 of the present work summarizes and extends results concerning the approximate computation of expectations. Chapter 2 is devoted to one stage stochastic programming. Chapter 3 is studying multistage stochastic programming problems. Chapter 4 contains the conclusion. In the Appendix, some results, cited in the present work, are listed.

Notation

Definition.

- The symbol := stands for the words "is defined by".

Special sets.

- Under R we shall understand the set of the real numbers. By R^k we shall mean Euclidean space of dimension k (i.e. cartesian product of k spaces R).
- $\mathrm{N} = \{1, 2, \ldots\}$ (the set of the natural numbers).
- The symbol $\mathcal{C}(\mathcal{X})$ denotes the linear space of the continuous functions on $\mathcal{X}$ equipped by the sup norm.

Sums.

- We use the convention that $\sum_{i=1}^{0} a_i = 0$

Vectors and matrices.

- Typeface of vector symbols will be small bold ($\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{\beta}$ for instance), symbols for matrices will be capital bold ($\boldsymbol{A}$, $\boldsymbol{V}$).
- The i-th component of the vector $\boldsymbol{a}$ is denoted by $\boldsymbol{a}^i$
- The notation $[a_{i,j}]_{i=1,\ldots,m,j=1,\ldots,n}$ means the matrix with $a_{i,j}$ in the i-th row and in the j-th column.
- The transposition of vectors and matrices is denoted by the symbol $'$,

  by $|\boldsymbol{C}|$.

Intervals and special functions.

- (Bounded) closed intervals are denoted by brackets, e.g. $[0, 1]$, open ones by parentheses, e.g. $(0, 1)$.
- The indicator function of a set M - i. e. the function being equal to one on the set M and to zero outside the set M - is denoted by I_M.
- The cartesian product of the sets A_i, $i = 1, 2, \ldots, n$, is denoted by $\prod_{i=1}^{n} A_i$ or $A_1 \times A_2 \times \ldots \times A_n$.

Limits.

- The notation $a_\nu \to a-$ means that the sequence $\{a_\nu\}_{\nu=1}^{\infty}$ converges to a from the left.
- Similarly, $a_\nu \to a+$ means the convergence from the right.
- The notation $a_\nu \to a$ denotes the convergence.

Sections and restrictions.

- Writing bullet ($\bullet$) instead of an argument of a function means that we are speaking of the section of a function with "bulleted" arguments varying.
- If $h : A \to B$ and $S \subseteq A$ then $h|_S$ denotes the restriction of the mapping h to the set S.

Norms and Lipschitz functions.

- Notation $\| \, \|_p$, where $p \in \{1, 2, \ldots\}$, means l_p norm in Euclidean space (i.e. $\|\boldsymbol{x}\|_p = \sqrt[p]{\sum_{i=1}^{k} |\boldsymbol{x}^i|^p}$).
- We say that the function $f : S \to \mathrm{R}$, $S \subset \mathrm{R}^k$ is *l_p-Lipschitz with constant $K > 0$* or shortly *l_p-K-Lipschitz* if $|f(x) - f(y)| \leq K\|x - y\|_p$ for each $x, y \in S$.

Optimization.

- By $\arg\min_{y \in Y} f(y)$ we denote the set $\{z \in Y : f(z) = \min_{y \in Y} f(y)\}$

Rounding.

- The symbol $\lfloor x \rfloor$ denotes the floor of number x, i. e. the greatest integer less or equal to x.
- The symbol $\lceil x \rceil$ stands for ceil of x, i. e. the smallest integer greater or equal then x.

Probability theory

- The words *probability distribution* are sometimes replaced by the abbreviation *p.d.*
- The *distribution function* of a random variable X defined as $F(x) := P(x < X)$ shall be mentioned as *d.f.* of X.
- The word *density* always means density with respect to the Lesbegue measure.
- By $\{x_i, p_i\}_{i=1}^{n}$ we mean the discrete probability distribution with the atoms x_i and the corresponding probabilities p_i. We allow the atoms of the distribution to coincide.
- If F is a one dimensional distribution function, then F^{-1} denotes its quantile (defined by the formula $F^{-1}(y) = \inf\{e \in \mathrm{R} : F(e) \geq y\}$).
- The convergence in distribution is denoted by $\xrightarrow{\mathcal{D}}$.
- We denote the normal distribution with the mean μ and the variance σ^2 by $N(\mu, \sigma^2)$.
- The words *independent identically distributed* are abbreviated as *i.i.d.*

- The symbol $\mathcal{L}(X)$ means the distribution of the random element X.
- Notion $\sigma(X)$ means the σ-algebra generated by random element X.
- The formula $P \otimes Q$ denotes the product measure of the measures P and Q. The formula $\mathcal{A} \otimes \mathcal{B}$ denotes the product σ-algebra of the σ-algebras $\mathcal{A}$ and $\mathcal{B}$.
- By writing $X \leq Y$ for some random variables X and Y we mean that $X(\omega) \leq Y(\omega)$ for each elementary event ω.

Conditional probabilities

- Let X, Y be random vectors with a joint distribution P. By the words *conditional distribution of X given Y* (denoted by $\mathcal{L}(X|Y)$) we mean a regular version of the random measure defined by the conditional probability $P(\bullet|\sigma(Y))$ (see [45], p. 331 for the definition of conditional probability and p. 338 for the definition of random measure).
- If $X_1, X_2, \ldots, X_m$ are random elements then, for readability, we sometimes write $\bar{X}_i$ instead of X_1, X_2, …, X_m. Further, if $P_1(\bullet)$ is the marginal distribution of X_1 and if $P_i(\bullet|X_1, X_2, \ldots, X_{i-1})$ is the conditional distribution of X_i given the vector (X_1, X_2, …, X_{i-1}), $i = 2, 3, \ldots, m$, we sometimes take $P_1(\bullet|\bar{X}_0) := P_1(\bullet)$ and $E(\bar{X}_1|\bar{X}_0) := E(\bar{X}_1)$. Moreover, if $F_i(\bullet|\bar{X}_{i-1})$ denotes a conditional distribution function of X_i given $\bar{X}_{i-1}$ and F_1 is the "unconditional" distribution function of X_1, we take $F_1(\bullet|\bar{X}_0) := F_1(\bullet)$.

Integrals and expectations.

- Instead of the lengthy notation

$$\int_\Omega g(X(\omega))dP(\omega),$$

  where Ω is the probability space, we shall write simply $E_{P(\omega)}g(X(\omega))$ or $E_P g$.
- Similarly, $E_P g(x)$ shall be used instead of $\int_\Omega g(x, \omega)dP(\omega)$.
- The variance $E_P(X - EX)^2$ shall be denoted simply by $\mathrm{var}_P X$, the symbol $\mathrm{cov}_P(X, Y)$ shall stand for the covariance.
- If there is no possible confusion then the symbol specifying the probability measure at E, var or cov shall be omitted.

Rate of convergence.

- We write $f(\bullet) = O(g(x))$ as $x \to s$ if $\limsup_{x\to s} |f(x)/g(x)| < \infty$.
- We write $f(x) = o(g(x))$ as $x \to s$ if $\limsup_{x\to s} |f(x)/g(x)| = 0$.

Convergence rate of random variables.

- Let X_n be a sequence of random variables. We write $X_n = O_P(n^a)$ if for each $\epsilon > 0$ there exists a constant M_ϵ such that $\limsup_{n\to\infty} P[|X_n| \geq n^a M_\epsilon] \leq \epsilon$.
- We write $X_n = o_P(n^a)$ as $n \to \infty$ if $\lim_{n\to\infty} P[|X_n| \geq n^a M] = 0$ for each $M > 0$.

Chapter 1

Approximate Computation of Expectation

1.1 Introduction

Consider the expression

$$E_P g = \int_{\mathrm{R}^k} g(\boldsymbol{x}) dP(\boldsymbol{x})$$

where $g : \mathrm{R}^k \to \mathrm{R}$ is l_1-K-Lipschitz[1] function, P is a probability distribution defined on Borel σ-algebra of R^k and k is some positive integer. Moreover, assume that the marginal distributions of P have finite first absolute moments.

Since it follows from the Lipschitz property that

$$g(\mathbf{0}) - K \sum_{i=1}^{k} |\boldsymbol{x}^i| \leq g(\boldsymbol{x}) \leq g(\mathbf{0}) + K \sum_{i=1}^{k} |\boldsymbol{x}^i|$$

it also holds that

$$g(\mathbf{0}) - K \sum_{i=1}^{k} E_{P_i(\boldsymbol{x}^i)} |\boldsymbol{x}^i| \leq E_{P(\boldsymbol{x})} g(\boldsymbol{x}) \leq g(\mathbf{0}) + K \sum_{i=1}^{k} E_{P_i(\boldsymbol{x}^i)} |\boldsymbol{x}^i|$$

(P_i denote the marginal distributions) so that the expectation $E_P g$ exists and is finite under our assumptions.

1.2 Methods of Approximation of Expectation

Three different approaches to the approximate computation of $E_P g$ are considered in the present Chapter: discretization (i.e. the replacement of P by a suitable discrete distribution), Monte Carlo estimation (i.e. the approximation of $E_P g$ by an average of i.i.d. realizations of g) and Quasi Monte Carlo (i.e. the partition of P's support into suitable regions, generating a realization of the conditional distribution given each region and averaging the values).

[1] We work with the l_1 norm for technical reasons. The one who wants to work with the other norms can use the fact that all the norms are equivalent in finite dimensional space, see [29], paragraph A.1.7.

1.2.1 Discretization

By (simple) discretization we mean the estimation of $E_P g$ by $E_\Pi g$ where $\Pi = \{\boldsymbol{x}_i, p_i\}_{i=1}^n$ is the probability distribution concentrated in atoms $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ with the corresponding probabilities $p_1, p_2, \ldots, p_n$.

It should be mentioned that there exist more sophisticated methods of integral's approximate computation then the simple discretization (e.g. various quadrature rules, see [5] for a survey). However, these methods require the differentiability (of g) which is often missing in stochastic programming problems. Therefore, we do not deal with these methods in the present work.

Concerning simple discretization itself, assertions similar to ones presented in our work exist (see [5], chp. 5.). However, these results do not cover the situation of unbounded support of P, common in many applications.

Denote

$$e(\Pi, g) = |E_\Pi g - E_P g| \tag{1.1}$$

the error risen by the discretization of P by Π and

$$e(\Pi) = \sup_{g \text{ is } l_1\text{-}K\text{-Lipschitz}} e(\Pi, g) \tag{1.2}$$

its upper bound. For one-dimensional probability distributions P and Π, it is easy to compute $e(\Pi)$ using either an analytical formula (if possible) or a suitable numerical method:

Lemma 1.1. (Pflug [36]) If $k = 1$ then

$$e(\Pi) = K \cdot d_W(P, \Pi) \tag{1.3}$$

where

$$d_W(P, \Pi) = \sup_{g \text{ is } l_1\text{-1-Lipschitz}} |E_\Pi g - E_P g|$$

and it holds that

$$d_W(P, \Pi) = \sum_{i=0}^{n} \int_{x_i}^{x_{i+1}} |F(x) - \sum_{j=1}^{i} p_j| dx \tag{1.4}$$

(F denotes the distribution function of P, we suppose that $x_1 \leq x_2 \leq \ldots \leq x_n$ and we take $x_0 = -\infty$, $x_{n+1} = \infty$ and $\sum_{j=1}^{0} p_j = 0$).

Proof. The formula (1.3) is straightforward.

Ad (1.4): According to [46], it holds that

$$d_W(P, \Pi) = \int |F(x) - G(x)| dx$$

where G is distribution function of Π. When we note that $G|_{(x_i, x_{i+1}]} = \sum_{j=1}^{i} p_j$ we get (1.4). □

Remark 1.1. The quantity $d_W(P, \Pi)$ is called Wasserstein distance of measures P and Π (see [36], Definition 1, and [39], chapter 5.3., for details).

It could be difficult to compute $e(\Pi)$ for $k > 1$. However, it is easy if Π belongs to a special class of discretizations:

Definition 1.1. Let $n = m_1 \cdot m_2 \cdot \ldots \cdot m_k$ for some positive integers $m_1, m_2, \ldots, m_k$. Distribution $\Pi = \{\boldsymbol{x}_i, p_i\}_{i=1}^n$ is *grid discretization of P with dimensions m_1, m_2, ..., m_k* if there exist

$$-\infty = c_{1,0} < x_{1,1} \leq c_{1,1} \leq x_{1,2} \leq \ldots \leq x_{1,m_1} < c_{1,m_1} = \infty,$$

$$-\infty = c_{2,0} < x_{2,1} \leq c_{2,1} \leq x_{2,2} \leq \ldots \leq x_{2,m_2} < c_{2,m_2} = \infty,$$

$$\ldots$$

$$-\infty = c_{k,0} < x_{k,1} \leq c_{k,1} \leq x_{k,k} \leq \ldots \leq x_{k,m_k} < c_{k,m_k} = \infty$$

fulfilling the following conditions: For each $i \leq i \leq n$ there exist $1 \leq j_1 \leq m_1, 1 \leq j_2 \leq m_2, \ldots, 1 \leq j_k \leq m_k$ such that

$$\boldsymbol{x}_i = (x_{1,j_1}, x_{2,j_2}, \ldots, x_{k,j_k})$$

and

$$p_i = P(M_{j_1,j_2,\ldots,j_k}),$$

where

$$M_{j_1,j_2,\ldots,j_k} = [c_{1,j_1-1}, c_{1,j_1}) \times [c_{2,j_2-1}, c_{2,j_2}) \times \ldots \times [c_{k,j_k-1}, c_{k,j_k})$$

(we take $P(\{-\infty\}) = 0$).

Remark 1.2. (i) Denote P_i the i-th marginal distribution of P. If Π is grid discretization of P with dimensions $m_1, m_2, \ldots, m_k$ then

$$\Pi_i = \{x_{i,j}, P_i([c_{i,j-1}, c_{i,j}))\}_{j=1}^{m_i}$$

is the i-th marginal distribution of Π for each $1 \leq i \leq k$.
(ii) Π_i is grid discretization of P_i with dimension m_i for each $1 \leq i \leq k$.
(iii) Conversely, if

$$\Pi_i = \{x_{i,j}, P_i([c_{i,j-1}, c_{i,j}))\}_{j=1}^{m_i}$$

is grid discretization of P_i for each $i = 1, 2, \ldots, k$ then

$$\Pi = \{(x_{1,j_1}, x_{2,j_1}, \ldots, x_{k,j_k}), P(M_{j_1,j_2,\ldots,j_k})\}_{j_1=1,2,\ldots,m_1, j_2=1,2,\ldots,m_2,\ldots,j_k=1,2,\ldots,m_k}$$

is grid discretization of P with dimensions $m_1, m_2, \ldots, m_k$ (with probabilities defined by $c_{i,j}$, $1 \leq j \leq m_i - 1$, $1 \leq i \leq k$) such that $\Pi_1, \Pi_2, \ldots, \Pi_k$ are its marginal distributions.

Lemma 1.2. Let Π be a grid discretization of P with dimensions $m_1, m_2, \ldots, m_k$. Then it holds that

$$e(\Pi) = \sum_{i=1}^{k} K \cdot d_W(\Pi_i, P_i)$$

where P_i and Π_i, $i = 1, 2, \ldots, k$ are marginal distributions of P, Π respectively.

Proof. For each l_1-K-Lipschitz function g it holds that

$$\begin{aligned} |E_P g - E_\Pi g| &= \left| \int_{\mathrm{R}^k} g dP - \sum_{j_1=1, j_2=1, \ldots, j_k=1}^{m_1, m_2, \ldots, m_k} g(x_{1,j_1}, x_{2,j_2}, \ldots, x_{k,j_k}) p_{j_1,j_2,\ldots,j_k} \right| \\ &= \left| \sum_{j_1,j_2,\ldots,j_k=1}^{m_1,m_2,\ldots,m_k} \left[\int_{M_{j_1,j_2,\ldots,j_k}} g dP - g(x_{1,j_1}, x_{2,j_2}, \ldots, x_{k,j_k}) \int_{M_{j_1,j_2,\ldots,j_k}} dP \right] \right| \end{aligned}$$

$$
\begin{aligned}
&= \Big| \sum_{j_1,j_2,\ldots,j_k=1}^{m_1,m_2,\ldots,m_k} \int_{M_{j_1,j_2,\ldots,j_k}} [g(\xi_1,\xi_2,\ldots,\xi_k) \\
&\qquad\qquad -g(x_{1,j_1},x_{2,j_2},\ldots,x_{k,j_k})] dP(\xi_1,\xi_2,\ldots,\xi_k)\Big| \\
&\leq \sum_{j_1,j_2,\ldots,j_k=1}^{m_1,m_2,\ldots,m_k} \int_{M_{j_1,j_2,\ldots,j_k}} K \sum_{i=1}^{k} |\xi_i - x_{i,j_i}| dP(\xi_1,\xi_2,\ldots,\xi_k) \\
&= \sum_{i=1}^{k} \left[K \sum_{j_1,j_2,\ldots,j_k=1}^{m_1,m_2,\ldots,m_k} \int_{M_{j_1,j_2,\ldots,j_k}} |\xi_i - x_{i,j_i}| dP(\xi_1,\xi_2,\ldots,\xi_k) \right] \\
&= \sum_{i=1}^{k} \left[K \sum_{j_i=1}^{m_i} \int_{\bigcup_{j_1,\ldots,j_{i-1},j_{i+1},\ldots,j_m} M_{j_1,j_2,\ldots,j_k}} |\xi_i - x_{i,j_i}| dP(\xi_1,\xi_2,\ldots,\xi_k) \right] \\
&= \sum_{i=1}^{k} \left[K \sum_{j=1}^{m_i} \int_{\mathrm{R}\times\ldots\times\mathrm{R}\times[c_{i,j-1},c_{i,j})\times\mathrm{R}\times\ldots\times\mathrm{R}} |\xi_i - x_{i,j}| dP(\xi_1,\xi_2,\ldots,\xi_k) \right] \\
&= \sum_{i=1}^{k} K d_i \qquad (1.5)
\end{aligned}
$$

where

$$d_i = \sum_{j=1}^{m_i} \int_{[c_{i,j-1},c_{i,j})} |\xi - x_{i,j}| dF_i(\xi),$$

F_i is the d.f. of P_i (we have used the triangular inequality and the Lipschitz property at the "$\leq$"). If we put

$$\bar{g} = K \sum_{i=1}^{k} \bar{g}_i, \quad \bar{g}_i : \mathrm{R} \to \mathrm{R}, \quad \bar{g}_i(\xi) = |\xi - x_{i,j}| + a_{i,j} \quad \text{for each } \xi \in [c_{i,j-1}, c_{i,j}]$$

where the constants $a_{i,j}$ are chosen so that g_i is continuous, then we get "=" instead of "$\leq$" in (1.5). Therefore and since $\bar{g}$ is l_1-K-Lipschitz we have

$$\sup_{h \text{ is } l_1\text{-}K\text{-Lipschitz}} |E_P h - E_\Pi h| \;=\; |E_P \bar{g} - E_\Pi \bar{g}| = \sum_{i=1}^{k} K d_i. \qquad (1.6)$$

Moreover, it holds that

$$d_W(P_i,\Pi_i) \overset{(1.3)}{=} \sup_{h \text{ is } l_1\text{-1-Lipschitz}} |E_{P_i} h - E_{\Pi_i} h| \overset{(1.6) \text{ with } k=1}{=} d_i \qquad (1.7)$$

for each $i = 1, 2, \ldots, k$. By the combination of (1.6) and (1.7) we get the assertion of the Lemma.
□

The following Theorem estimates how close to P we may get with m^k-atom grid distribution.

Theorem 1.1. Let m be a positive integer. Denote $F_1, F_2, \ldots, F_k$ the marginal distribution functions of P and assume that F_i is continuous for each $i = 1, 2, \ldots, k$. Let $\bar{\Pi}$ a be grid discretization of P with dimensions $m, m, \ldots, m$ defined by

$$x_{i,j} := F_i^{-1}\left(\frac{2j-1}{2m}\right) \quad i = 1, 2, \ldots, k \quad j = 1, 2, \ldots, m, \qquad (1.8)$$

and

$$c_{i,j} := F_i^{-1}\left(\frac{j}{m}\right) \quad i = 1, 2, \ldots, k \quad j = 1, 2, \ldots, m-1. \tag{1.9}$$

Denote

$$\underline{t}_i(m) = \int_{-\infty}^{F_i^{-1}(\frac{1}{2m})} F_i(x)dx, \quad \bar{t}_i(m) = \int_{F_i^{-1}(1-\frac{1}{2m})}^{\infty} [1 - F_i(x)]dx, \quad i = 1, 2, \ldots, k.$$

Then it holds that

$$e(\bar{\Pi}) \leq \overline{e}_{m^k}, \qquad \overline{e}_{m^k} = K \sum_{i=1}^{k} \left[\underline{t}_i(m) + \frac{1}{2m}\left(F_i^{-1}\left(1 - \frac{1}{2m}\right) - F_i^{-1}\left(\frac{1}{2m}\right)\right) + \bar{t}_i(m)\right] \tag{1.10}$$

Remark 1.3. $\underline{t}_i(m)$ and $\bar{t}_i(m)$ exist and are finite for each $i = 1, 2, \ldots, k$, $m \in \mathrm{N}$.

Proof of the Remark 1.3. The proof is well known: Using the integration by parts, we get

$$\int_{-\infty}^{0} F_i(x)dx = [xF_i(x)]_{-\infty}^{0} - \int_{-\infty}^{0} xdF_i(x).$$

The finiteness of the second term follows from the finiteness of the P_i's first moment. As to the first term, it holds that

$$0 \leq [xF_i(x)]_{-\infty}^{0} = -\lim_{x\to-\infty} xF_i(x) = -\lim_{x\to-\infty} \int_{-\infty}^{x} xdF_i(y) \leq -\lim_{x\to-\infty} \int_{-\infty}^{x} ydF_i(y) = 0$$

due to the finiteness of the first moment. Hence, $\underline{t}_i(m)$ has to be finite. The proof for $\bar{t}_i(m)$ is similar.

□

Proof of the Theorem 1.1. It holds that $F_i(F_i^{-1}(\alpha)) \leq \alpha$ (if $F_i(F_i^{-1}(\alpha)) > \alpha$ then, thanks to the continuity of F_i from the left, there would exist $x < F_i^{-1}(\alpha)$ such that $F(x) > \alpha$ which would contradict the definition of F_i^{-1}). Moreover, we have that $F_i(F_i^{-1}(\alpha)+) \geq \alpha$ (the definition of quantile implies that for each $x > F_i^{-1}(\alpha)$ it holds that $F_i(x) \geq \alpha$ so that for each sequence $x_\nu > F_i^{-1}(\alpha)$ it should hold $\lim_\nu F_i(x_\nu) \geq \alpha$). Using it and the continuity of F_i we get that

$$F_i(F_i^{-1}(\alpha)) = \alpha \tag{1.11}$$

for each $\alpha \in (0, 1)$, $i = 1, 2, \ldots, k$.

Therefore, being $1 \leq i \leq k$ and $1 \leq j \leq m-1$ it holds that, for each $\xi \in [x_{i,j}, c_{i,j}]$,

$$F_i(c_{i,j}) - F_i(\xi) \leq F_i(c_{i,j}) - F_i(x_{i,j}) \overset{(1.11)}{=} \frac{j}{m} - \frac{2j-1}{2m} = \frac{1}{2m}$$

and, for each $\xi \in [c_{i,j}, x_{i,j+1}]$,

$$F_i(\xi) - F_i(c_{i,j}) \leq F_i(x_{i,j+1}) - F_i(c_{i,j}) \overset{(1.11)}{=} \frac{2j+1}{2m} - \frac{j}{m} = \frac{1}{2m}.$$

Thus, $|F_i(\xi) - F_i(c_{i,j})| \leq \frac{1}{2m}$ for each $\xi \in [x_{i,j}, x_{i,j+1}]$, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, m-1$ and it holds that

$$\int_{x_{i,j}}^{x_{i,i+1}} |F_i(\xi) - F_i(c_{i,j})|d\xi \leq (x_{i,j+1} - x_{i,j})\frac{1}{2m}. \tag{1.12}$$

Using it we get

$$\begin{aligned} d_W(\bar{\Pi}_i, P_i) &\overset{(1.3)}{=} \underline{t}_i(m) + \sum_{j=1}^{m-1} \int_{x_{i,j}}^{x_{i,i+1}} |F_i(\xi) - \sum_{\nu=1}^{j} [F_i(c_{i,\nu}) - F(c_{i,\nu-1})]| d\xi + \bar{t}_i(m) \\ &= \underline{t}_i(m) + \sum_{j=1}^{m-1} \int_{x_{i,j}}^{x_{i,i+1}} |F_i(\xi) - F_i(c_{i,j})| d\xi + \bar{t}_i(m) \\ &\overset{(1.12)}{\leq} \underline{t}_i(m) + \frac{1}{2m}(x_{i,m-1} - x_{i,1}) + \bar{t}_i(m) \end{aligned} \tag{1.13}$$

for $i = 1, 2, \ldots, k$ ($\bar{\Pi}_i$ and P_i denote marginal distribution of $\bar{\Pi}$, P respectively) so that

$$e(\bar{\Pi}) \overset{\text{Lemma 1.2.}}{=} K \sum_{i=1}^{k} d_W(\bar{\Pi}_i, P_i) \overset{(1.13)}{\leq} K \sum_{i=1}^{k} \left[\underline{t}_i(m) + \frac{1}{2m}(x_{i,m-1} - x_{i,1}) + \bar{t}_i(m) \right].$$

□

The following assertion states a lower bound of the discretization error.

Theorem 1.2. Let $n = m^k$ for some integer $m > 1$. Assume that p.d. P has a density f. Then it holds, for each n-atom p.d. Π, that

$$e(\Pi) \geq \underline{e}_n > 0 \tag{1.14}$$

where

$$\underline{e}_n := \sup_{\delta > 0} \underline{e}_{n,\delta},$$

$$\underline{e}_{n,\delta} = K\delta \max\left(\frac{1}{4m} - \frac{1}{4m^{2k^2-k+1}}, \frac{1}{2m} - \frac{(m^{2k-1}+1)^{k+1} - 1}{(k+1)2m^{2k^2+k}} \right) \sum_{i=1}^{\infty} \sqrt[k]{M_{i\delta}^{k+1} k!}$$

and

$$M_z = \mu\{\boldsymbol{y} \in \mathrm{R}^k : f(\boldsymbol{y}) \geq z\}$$

(μ denotes Lesbegue measure).

Before we prove the Theorem, we state two Lemma's.

Lemma 1.3. Let $A \subseteq \mathrm{R}^k$ be a Borel set. Then it holds, for each nonnegative measurable function h and $z > 0$, that

$$h(\boldsymbol{x}) \geq \frac{1}{z} \sum_{j=1}^{\infty} I_{\{\boldsymbol{y}: h(\boldsymbol{y}) \geq \frac{j}{z}\}}(\boldsymbol{x}). \tag{1.15}$$

and

$$\int_A h(\boldsymbol{x}) d\boldsymbol{x} \geq \frac{1}{z} \sum_{j=1}^{\infty} \mu\left(A \cap \left\{ \boldsymbol{y} : h(\boldsymbol{y}) \geq \frac{j}{z} \right\} \right). \tag{1.16}$$

Proof. It holds that

$$h(\boldsymbol{x}) \geq \underline{h}(\boldsymbol{x}) \quad \text{where} \quad \underline{h}(\boldsymbol{x}) = \frac{\lfloor h(\boldsymbol{x}) z \rfloor}{z} = \frac{1}{z} \sum_{j=1}^{\infty} I_{\{\boldsymbol{y}: h(\boldsymbol{y}) z \geq j\}}(\boldsymbol{x}) = \frac{1}{z} \sum_{j=1}^{\infty} I_{\{\boldsymbol{y}: h(\boldsymbol{y}) \geq \frac{j}{z}\}}(\boldsymbol{x}).$$

which proves (1.15). By integrating $\underline{h}$ over A we get (1.16).
□

Lemma 1.4. Let $\boldsymbol{z}$ be a k-dimensional vector and let $v > 0$ be a constant. It holds that

$$\mu\{\boldsymbol{y} \in \mathrm{R}^k : \|\boldsymbol{y} - \boldsymbol{z}\|_1 < v\} = \frac{2^k v^k}{k!}.$$

where μ denotes the Lesbegue measure.

Proof. It suffices to take $\boldsymbol{z} = (0, 0, \ldots, 0)'$ (the translation preserves Lesbegue measure). We prove the assertion by induction.
If $k = 1$ then the assertion is trivial.
Denote $\psi_k : \psi_k(v) = \mu\{\boldsymbol{y} \in \mathrm{R}^k : \|\boldsymbol{y}\|_1 < v\}$, $k = 1, 2, \ldots$ and assume that $\psi_{k-1}(w) = \frac{2^{k-1} w^{k-1}}{(k-1)!}$ for each $w > 0$. Then it holds that

$$\begin{aligned}
\psi_k(v) &= \int_{\left\{\sum_{i=1}^k |y_i| \le v\right\}} dy_k dy_{k-1} \ldots dy_1 = \int_{\left\{|y_k| \le v \wedge \sum_{i=1}^{k-1} |y_i| \le v - |y_k|\right\}} dy_k dy_{k-1} \ldots dy_1 \\
&= \int_{\{|y_k| \le v\}} \left(\int_{\left\{\sum_{i=1}^{k-1} |y_i| \le v - |y_k|\right\}} dy_{k-1} \ldots dy_1 \right) dy_k \\
&= \int_{-v}^{v} \psi_{k-1}(v - |y_k|) dy_k = 2 \int_0^v \psi_{k-1}(v - y_k) dy_k \overset{u=v-y_k}{=} 2 \int_0^v \psi_{k-1}(u) du. \qquad (1.17)
\end{aligned}$$

By evaluating (1.17) we get the assertion of the Lemma.
□

Proof of Theorem 1.2. *Ad.* $\underline{e}_n > 0$. We show, by contradiction, the existence of such $\delta > 0$ that $\underline{e}_{n,\delta} > 0$: If $e_{n,\delta} = 0$ for each $\delta > 0$ then necessarily $M_\delta = 0$ for each $\delta > 0$. If $M_\delta = 0$ for each $\delta > 0$ then, for each measurable set A with finite Lesbegue measure, it would hold that

$$\begin{aligned}
P(A) &= \int_{A \cap \{\boldsymbol{y}: f(\boldsymbol{y}) \ge \delta\}} f(\boldsymbol{y}) d\boldsymbol{y} + \int_{A \cap \{\boldsymbol{y}: f(\boldsymbol{y}) < \delta\}} f(\boldsymbol{y}) d\boldsymbol{y} \\
&\le \int_{\{\boldsymbol{y}: f(\boldsymbol{y}) \ge \delta\}} f(\boldsymbol{y}) d\boldsymbol{y} + \int_{A \cap \{\boldsymbol{y}: f(\boldsymbol{y}) < \delta\}} f(\boldsymbol{y}) d\boldsymbol{y} \\
&\overset{\mu\{\boldsymbol{y}: f(\boldsymbol{y}) \ge \delta\} = 0}{=} \int_{A \cap \{\boldsymbol{y}: f(\boldsymbol{y}) < \delta\}} f(\boldsymbol{y}) d\boldsymbol{y} \le \delta \mu(A)
\end{aligned}$$

for each $\delta > 0$, hence $P(A) \le \lim_{\delta \to 0} \delta \mu(A) = 0$. Since R^k can be covered by a countable system of sets with finite Lesbegue measure it would hold that $P(\mathrm{R}^k) = 0$ which is contradiction.
Ad. $e(\Pi) \ge \underline{e}_n$. Let $\Pi = \{\boldsymbol{x}_i, p_i\}_{i=1}^n$ be an n-atom probability distribution and let $\delta > 0$ be a real constant.
Consider the function $\tilde{g} : \tilde{g}(\boldsymbol{x}) = K \min_{i=1,2,\ldots,n} \|\boldsymbol{x} - \boldsymbol{x}_i\|_1$. It is obvious that $\tilde{g}$ is l_1-K-Lipschitz and it follows from (1.2), from basic properties of Lesbegue integral and from Lemma 1.3. that

$$\begin{aligned}
e(\Pi) &\ge \left| \int_{\mathrm{R}^k} \tilde{g} dP - \int_{\mathrm{R}^k} \tilde{g} d\Pi \right| = \left| \int_{\mathrm{R}^k} \tilde{g} dP - 0 \right| = \int_{\mathrm{R}^k} \tilde{g}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} \\
&\overset{(1.15)}{\ge} \int_{\mathrm{R}^k} \tilde{g}(\boldsymbol{x}) \left(\sum_{i=1}^{\infty} \delta I_{\{\boldsymbol{y}: f(\boldsymbol{y}) \ge i\delta\}}(x) \right) d\boldsymbol{x} = \delta \sum_{i=1}^{\infty} \int_{\{\boldsymbol{y}: f(\boldsymbol{y}) \ge i\delta\}} \tilde{g}(\boldsymbol{x}) d\boldsymbol{x}. \qquad (1.18)
\end{aligned}$$

Using Lemma 1.4. and the definition of $\tilde{g}$ we get that

$$\begin{aligned}\mu\{\boldsymbol{y}\in\mathrm{R}^k:\tilde{g}(\boldsymbol{y})<\rho\} &= \mu\left(\bigcup_{j=1}^{n}\{\boldsymbol{y}\in\mathrm{R}^k:K\|\boldsymbol{y}-\boldsymbol{x}_j\|_1<\rho\}\right)\\ &\le \sum_{j=1}^{n}\mu\left(\{\boldsymbol{y}\in\mathrm{R}^k:K\|\boldsymbol{y}-\boldsymbol{x}_j\|_1<\rho\}\right)\\ &\overset{\text{Lemma 1.4.}}{=} n\frac{2^k\rho^k}{K^k k!}\end{aligned}$$

for each $\rho>0$ so that

$$\begin{aligned}\mu(\{\boldsymbol{y}:f(\boldsymbol{y})\ge i\delta\}\cap\{\boldsymbol{y}:\tilde{g}(\boldsymbol{y})\ge\rho\}) &\ge \max\left(0,\mu(\{\boldsymbol{y}:f(\boldsymbol{y})\ge i\delta\})-\mu(\{\boldsymbol{y}:\tilde{g}(\boldsymbol{y})<\rho\})\right)\\ &\ge \max\left(0,M_{i\delta}-n\frac{2^k\rho^k}{K^k k!}\right) \qquad (1.19)\end{aligned}$$

for each $\rho>0$ and $i=1,2,\ldots$. (We have used the fact that $\mu(A\cap B)=\mu(A\cap B)+\mu(\neg B)-\mu(\neg B)=\mu((A\cap B)\cup\neg B)-\mu(\neg B)\ge\mu((A\cap B)\cup(A\cap\neg B))-\mu(\neg B)=\mu(A)-\mu(\neg B)$.) By putting $A=\{\boldsymbol{y}:f(\boldsymbol{y})>i\delta\}$, $h=\tilde{g}$ and $z=\frac{2m^{2k}}{K\sqrt[k]{M_{i\delta}k!}}$ in (1.16) and using (1.19) we get

$$\begin{aligned}\int_{\{\boldsymbol{y}:f(\boldsymbol{y})>i\delta\}}\tilde{g}(\boldsymbol{x})d\boldsymbol{x} &\ge \frac{K\sqrt[k]{M_{i\delta}k!}}{2m^{2k}}\sum_{j=1}^{\infty}\mu(\{\boldsymbol{y}:f(\boldsymbol{y})>i\delta\}\cap\{\boldsymbol{y}:\tilde{g}(\boldsymbol{y})\ge j\frac{K\sqrt[k]{M_{i\delta}k!}}{2m^{2k}}\})\\ &\ge \frac{K\sqrt[k]{M_{i\delta}k!}}{2m^{2k}}\sum_{j=1}^{\infty}\max\left(0,M_{i\delta}-n\frac{2^k\left(\frac{jK\sqrt[k]{M_{i\delta}k!}}{2m^{2k}}\right)^k}{K^k k!}\right)\\ &= \frac{K\sqrt[k]{M_{i\delta}k!}}{2m^{2k}}\sum_{j=1}^{\infty}\max\left(0,M_{i\delta}-m^k\frac{M_{i\delta}j^k}{m^{2k^2}}\right)\\ &= \frac{K\sqrt[k]{M_{i\delta}^{k+1}k!}}{2m^{2k}}\sum_{j=1}^{\infty}\max\left(0,1-\frac{j^k}{m^{2k^2-k}}\right)\\ &= \frac{K\sqrt[k]{M_{i\delta}^{k+1}k!}}{2m^{2k}}\sum_{j=1}^{m^{2k-1}}\left(1-\frac{j^k}{m^{2k^2-k}}\right)\\ &= K\sqrt[k]{M_{i\delta}^{k+1}k!}\left(\frac{1}{2m}-\frac{1}{2m^{2k^2+k}}\sum_{j=1}^{m^{2k-1}}j^k\right). \qquad (1.20)\end{aligned}$$

Let us estimate the sum two different ways (each being suitable for a different m and k). Since $\sum_{j=1}^{M}j^k\le M(1+M^k)/2$ for each $M\in\mathrm{N}$ (we have estimated the convex function j^k by a linear function) we have

$$\sum_{j=1}^{m^{2k-1}}j^k\le\frac{m^{2k-1}}{2}(1+m^{k(2k-1)})$$

so that

$$\begin{aligned}\frac{1}{2m}-\frac{1}{2m^{2k^2+k}}\sum_{j=1}^{m^{2k-1}}j^k &\ge \frac{1}{2m}-\frac{1}{2m^{2k^2+k}}\frac{m^{2k-1}}{2}\left(1+m^{k(2k-1)}\right)\\ &= \frac{1}{2m}-\frac{1}{4m^{2k^2-k+1}}-\frac{1}{4m}=\frac{1}{4m}-\frac{1}{4m^{2k^2-k+1}}. \qquad (1.21)\end{aligned}$$

Further, since

$$\sum_{j=1}^{M} j^k \;\leq\; \int_1^{M+1} j^k dj = \left[\frac{j^{k+1}}{k+1}\right]_1^{M+1} = \frac{(M+1)^{k+1}-1}{k+1}$$

for each $M \in \mathrm{N}$, we have

$$\sum_{j=1}^{m^{2k-1}} j^k \leq \frac{(m^{2k-1}+1)^{k+1}-1}{k+1}$$

so that

$$\frac{1}{2m} - \frac{1}{2m^{2k^2+k}} \sum_{j=1}^{m^{2k-1}} j^k \;\geq\; \frac{1}{2m} - \frac{(m^{2k-1}+1)^{k+1}-1}{(k+1)2m^{2k^2+k}} \tag{1.22}$$

Hence

$$\begin{aligned} e(\Pi) &\overset{(1.18)}{\geq} \delta \sum_{i=1}^{\infty} \int_{\{\boldsymbol{y}: f(\boldsymbol{y}) \geq i\delta\}} \tilde{g}(\boldsymbol{x}) d\boldsymbol{x} \overset{(1.20)}{\geq} \delta \sum_{i=1}^{\infty} K \sqrt[k]{M_{i\delta}^{k+1} k!} \left(\frac{1}{2m} - \frac{1}{2m^{2k^2+k}} \sum_{j=1}^{m^{2k-1}} j^k \right) \\ &\overset{(1.21),(1.22)}{\geq} \delta \sum_{i=1}^{\infty} K \sqrt[k]{M_{i\delta}^{k+1} k!} \max\left(\frac{1}{4m} - \frac{1}{4m^{2k^2-k+1}}, \frac{1}{2m} - \frac{(m^{2k-1}+1)^{k+1}-1}{(k+1)2m^{2k^2+k}} \right). \end{aligned}$$

□

Rate of Convergence

It is known (see [5], Theorem on p. 267) that if $h : \mathrm{R}^k \to \mathrm{R}$ is a function with finite variation and $H \subset \mathrm{R}^k$ is a bounded set then the error of approximation of $\int_H h(\boldsymbol{x}) d\boldsymbol{x}$ by $\frac{1}{n}\sum_{i=1}^n h(\boldsymbol{x}_i)$, where $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are suitably chosen points, is $O(n^{-1/k})$. The present work generalizes this result.

Theorem 1.3. Denote F_i, $i = 1, 2, \ldots, k$, the marginal distribution functions of P. If there exists $a > 1$ such that $F_i(x) = O(|x|^{-a})$ as $x \to -\infty$ and $1 - F_i(x) = O(|x|^{-a})$ as $x \to \infty$ for each $i = 1, 2, \ldots, k$ then a sequence $\{\Pi^n\}_{n=1}^{\infty}$ of at most n-atom distributions Π^n can be constructed such that

$$e(\Pi^n) = O\left(n^{\frac{1}{k}(-1+1/a)}\right) \quad \text{as } n \to \infty.$$

Before we prove the Theorem, we state a useful result.

Lemma 1.5. Let S be distribution of one-dimensional real random variable defined by a distribution function H, let $C > 0$ be some constant and let $m > 2$ be some positive integer. Then there exists grid discretization Γ of S with at most m atoms such that

$$d_W(\Gamma, S) \leq \int_{-\infty}^{-C} H(\xi) d\xi + \frac{2C}{m-2} + \int_C^{\infty} [1 - H(\xi)] d\xi. \tag{1.23}$$

Proof. Define

$$H^-(\xi) := \begin{cases} H(\xi) & \xi \leq -C \\ H(-C) & \xi > -C, \end{cases}$$

$$H^+(\xi) := \begin{cases} 0 & \xi \leq C \\ H(\xi) - H(C) & \xi > C, \end{cases}$$

and

$$\tilde{H}(\xi) := \begin{cases} 0 & \xi \le -C \\ H(\xi) - H(-C) & -C < \xi \le C \\ H(C) - H(-C) & \xi > C. \end{cases}$$

By a simple calculation we get that $H^- + \tilde{H} + H^+ = H$.

Further, define

$$G^-(\xi) := H(-C) I_{(-C,\infty)}(\xi), \quad \xi \in \mathrm{R},$$
$$G^+(\xi) := [1 - H(C)] I_{(C,\infty)}(\xi), \quad \xi \in \mathrm{R}$$

and

$$\begin{aligned} \tilde{G}(\xi) \;:=\; & \tilde{H}(\tilde{H}^{-1}(D)) I_{(-C,\tilde{H}^{-1}(D)]}(\xi) + \sum_{i=2}^{m-2} \tilde{H}(\tilde{H}^{-1}(iD)) I_{(\tilde{H}^{-1}((i-1)D),\tilde{H}^{-1}(iD)]}(\xi) \\ & + \tilde{H}(C) I_{(\tilde{H}^{-1}((m-2)D),\infty)}(\xi), \quad \xi \in \mathrm{R} \end{aligned}$$

where $D = \frac{\tilde{H}(C)}{m-2}$ and $\tilde{H}^{-1}(y) = \inf\{e \in \mathrm{R} : \tilde{H}(e) \ge y\}$. It is possible to show that $\tilde{H}(\tilde{H}^{-1}(d)) \le d$ and that $\tilde{H}(\tilde{H}^{-1}(d)+) \ge d$ for each $d \in (0, \tilde{H}(C)]$ (see the proof of Theorem 1.1.). Using it, we can estimate: If $\xi \in (-C, \tilde{H}^{-1}(D)]$ then

$$\begin{aligned} |\tilde{G}(\xi) - \tilde{H}(\xi)| &= |\tilde{H}(\tilde{H}^{-1}(D)) - \tilde{H}(\xi)| = \tilde{H}(\tilde{H}^{-1}(D)) - \tilde{H}(\xi) \\ &\le \tilde{H}(\tilde{H}^{-1}(D)) \le D. \end{aligned} \tag{1.24}$$

If $\xi \in (\tilde{H}^{-1}((i-1)D), \tilde{H}^{-1}(iD)]$, $2 \le i \le m-2$, then

$$\begin{aligned} |\tilde{G}(\xi) - \tilde{H}(\xi)| &= |\tilde{H}(\tilde{H}^{-1}(iD)) - \tilde{H}(\xi)| = \tilde{H}(\tilde{H}^{-1}(iD)) - \tilde{H}(\xi) \\ &\le \tilde{H}(\tilde{H}^{-1}(iD)) - \tilde{H}(\tilde{H}^{-1}((i-1)D)+) \le iD - (i-1)D = D. \end{aligned} \tag{1.25}$$

Finally, if $\xi \in (\tilde{H}^{-1}((m-2)D), C)$ then

$$\begin{aligned} |\tilde{G}(\xi) - \tilde{H}(\xi)| &= |\tilde{H}(C) - \tilde{H}(\xi)| = \tilde{H}(C) - \tilde{H}(\xi) \\ &\le \tilde{H}(C) - \tilde{H}(\tilde{H}^{-1}((m-2)D)+) \\ &\overset{\text{def. of } D}{=} \tilde{H}(C) - \tilde{H}(\tilde{H}^{-1}(\tilde{H}(C))+) \le \tilde{H}(C) - \tilde{H}(C) = 0. \end{aligned} \tag{1.26}$$

Moreover, it follows from the definitions of $\tilde{G}$ and $\tilde{H}$ that

$$|\tilde{G}(\xi) - \tilde{H}(\xi)| = |0 - 0| = 0 \tag{1.27}$$

for $\xi \in (-\infty, -C] \cup (C, \infty)$. When we summarize (1.24), (1.25), (1.26) and (1.27) we get

$$|\tilde{G}(\xi) - \tilde{H}(\xi)| \le D \cdot I_{(-C,\tilde{H}^{-1}((m-2)D)]}(\xi)$$

and, using the fact that $\tilde{H}^{-1}((m-2)D) = \tilde{H}^{-1}(\tilde{H}(C)) \le C$,

$$|\tilde{G}(\xi) - \tilde{H}(\xi)| \le D \cdot I_{(-C,C]}(\xi). \tag{1.28}$$

Define $G := G^- + \tilde{G} + G^+$. Since $\lim_{\xi \to -\infty} G(\xi) = 0$, $\lim_{\xi \to \infty} G(\xi) = 1$ and G has at most m jumps (at $x_1 := -C, x_2 := \tilde{H}^{-1}(D), x_3 := \tilde{H}^{-1}(2D), \ldots, x_{m-1} := \tilde{H}^{-1}((m-2)D).x_m := C$,

some of the values may coincide), G is distribution function of an at most m-atom discrete probability distribution Γ. Using [46], triangular inequality, formula (1.28) and the definitions of H^-,H^+,G^- and G^+, we gradually obtain

$$\begin{aligned} d_W(\Gamma, S) &= \int |G(\xi) - H(\xi)|d\xi \\ &\leq \int |G^-(\xi) - H^-(\xi)|d\xi + \int |\tilde{G}(\xi) - \tilde{H}(\xi)|d\xi + \int |G^+(\xi) - H^+(\xi)|d\xi \\ &\leq \int |G^-(\xi) - H^-(\xi)|d\xi + \int [D \cdot I_{(-C,C]}(\xi)]d\xi + \int |G^+(\xi) - H^+(\xi)|d\xi \\ &= \int_{-\infty}^{-C} H(\xi)d\xi + 2CD + \int_C^{\infty} [1 - H(\xi)]d\xi \\ &= \int_{-\infty}^{-C} H(\xi)d\xi + 2C\frac{\tilde{H}(C)}{m-2} + \int_C^{\infty} [1 - H(\xi)]d\xi \\ &\overset{\tilde{H}(C)\leq 1}{\leq} \int_{-\infty}^{-C} H(\xi)d\xi + \frac{2C}{m-2} + \int_C^{\infty} [1 - H(\xi)]d\xi. \end{aligned}$$

It remains to prove that the distribution Γ is grid discretization, i.e. that there exist c_i, $i = 1, 2, \ldots, m-1$, such that $x_1 \leq c_1 \leq x_2 \leq c_2 \leq x_3 \leq \ldots \leq c_{m-1} \leq x_m$ and $\Gamma\{x_i\} = H(c_i) - H(c_{i-1})$, i.e.

$$G(x_{i+1}) - G(x_i) = H(c_i) - H(c_{i-1}) \tag{1.29}$$

for each $i = 1, 2, \ldots, m$ (we take $x_{m+1} = \infty$, $G(x_{m+1}) = 1$, $c_0 = -\infty$, $H(c_0) = 0$, $c_m = \infty$ and $H(c_m) = 1$). Clearly, (1.29) holds if

$$G(x_{i+1}) = H(c_i)$$

for each $i = 0, 1, 2, \ldots, m-1$. But if we put $c_1 = \tilde{H}^{-1}(D), c_2 = \tilde{H}^{-1}(2D), \ldots, c_{m-2} = \tilde{H}^{-1}((m-2)D)$ and $c_{m-1} = C$ we see that

$$G(x_1) = 0 = H(c_0),$$

$$G(x_{i+1}) = G(\tilde{H}^{-1}(iD)) = H(-C) + \tilde{H}(\tilde{H}^{-1}(iD)) = H(\tilde{H}^{-1}(iD)) = H(c_i)$$

for $i = 1, 2, \ldots, m-2$ and

$$G(x_m) = G(C) = H(-C) + \tilde{H}(C) = H(C) = H(c_{m-1}).$$

The three equalities above prove (1.29).
□

Proof of the Theorem 1.3. Denote P_i, $i = 1, 2, \ldots, k$, the marginal distributions of P. By using Lemma 1.5. we get that for each $1 \leq i \leq k$ and $m \in \mathrm{N}$, there exists at most m-atom discrete distribution Π_i^m such that

$$d_W(\Pi_i^m, P_i) \leq \bar{d}_{m,i}, \qquad \overline{d}_{m,i} = \int_{-\infty}^{-m^{1/a}} F_i(\xi)d\xi + \frac{2m^{1/a}}{m-2} + \int_{m^{1/a}}^{\infty} [1 - F_i(\xi)]d\xi \tag{1.30}$$

(we have put $C = m^{1/a}$ in Lemma 1.5.).

Since, according to the assumptions of the Theorem, $F_i(\xi) = O(|\xi|^{-a})$ as $\xi \to -\infty$, i.e.

$$\limsup_{\xi \to \infty} \frac{F_i(-\xi)}{\xi^{-a}} = A$$

for some $A < \infty$, it has to exist $\epsilon > 0$ and ξ_0 such that $F_i(-\xi) \le (A+\epsilon)\xi^{-a}$ for each $\xi \ge \xi_0$. Therefore,

$$\begin{aligned}\limsup_{m\to\infty} \frac{\int_{-\infty}^{-m^{1/a}} F_i(\xi)d\xi}{m^{-1+1/a}} &= \limsup_{m\to\infty} \frac{\int_{m^{1/a}}^{\infty} F_i(-\xi)d\xi}{m^{-1+1/a}} \le \limsup_{m\to\infty} \frac{\int_{m^{1/a}}^{\infty}(A+\epsilon)\xi^{-a}d\xi}{m^{-1+1/a}} \\ &= \lim_{m\to\infty} \frac{(A+\epsilon)\left[\frac{\xi^{-a+1}}{-a+1}\right]_{m^{1/a}}^{\infty}}{m^{-1+1/a}} = \lim_{m\to\infty} \frac{-\frac{A+\epsilon}{-a+1}m^{-1+1/a}}{m^{-1+1/a}} \\ &= \frac{A+\epsilon}{a-1} < \infty.\end{aligned}$$

i.e. $\int_{-\infty}^{-m^{1/a}} F_i(\xi)d\xi = O(m^{-1+1/a})$. Similarly we get that $\int_{m^{1/a}}^{\infty}[1-F_i(\xi)]d\xi = O(m^{-1+1/a})$.
Clearly

$$\frac{2m^{1/a}}{m-2} = O(m^{-1+1/a})$$

so that, due to (1.30),

$$\bar{d}_{m,i} = O(m^{-1+1/a}) + O(m^{-1+1/a}) + O(m^{-1+1/a}) = O(m^{-1+1/a}). \tag{1.31}$$

Now we are able to construct the sequence of distributions whose existence is asserted by the Theorem: Let $n \in \mathrm{N}$ and denote Π^n the grid discretization defined by $\Pi_i^{\lfloor n^{1/k}\rfloor}$, $i = 1, 2, \ldots, k$ (see Remark 1.2., (iii)). Obviously, Π^n has at most n atoms. Moreover, it holds that

$$e(\Pi^n) \overset{\text{Lemma 1.2.}}{=} \sum_{i=1}^{k} K \cdot d_W(\Pi_i^{\lfloor n^{1/k}\rfloor}, P_i) \overset{(1.30)}{\le} K\sum_{i=1}^{k} \bar{d}_{\lfloor n^{1/k}\rfloor,i} \overset{(1.31)}{=} O\left(\lfloor n^{\frac{1}{k}}\rfloor^{-1+1/a} A\right) = O\left(n^{\frac{-1+1/a}{k}}\right).$$

□

Corollary. Let P have marginal densities $f_i(x) = e^{-\phi_i(x)}$, $i = 1, 2, \ldots, k$, such that for some constants $C > 0$, $D > 0$ and for each $|x| \ge D$, it holds that $\phi_i(x) \ge C|x|$. Then there exists a sequence of at most n-atom distributions Π^n such that

$$e(\Pi^n) = o(n^{-\frac{1}{k}+\delta}) \quad \text{as } n \to \infty \tag{1.32}$$

for each arbitrarily small $\delta > 0$.

Proof. Since it holds that

$$\begin{aligned}\lim_{m\to\infty} \frac{F_i(-m)}{m^{-\frac{2}{k\delta}}} &= \lim_{m\to\infty} \frac{\int_{-\infty}^{-m} e^{-\phi_i(x)}dx}{m^{-\frac{2}{k\delta}}} \le \lim_{m\to\infty} \frac{\int_{-\infty}^{-m} e^{-C|x|}dx}{m^{-\frac{2}{k\delta}}} = \lim_{m\to\infty} \frac{\int_{-\infty}^{-m} e^{Cx}dx}{m^{-\frac{2}{k\delta}}} \\ &= \lim_{m\to\infty} \frac{\left[\frac{1}{C}e^{Cx}\right]_{-\infty}^{-m}}{m^{-\frac{2}{k\delta}}} = \lim_{m\to\infty} \frac{(1/C)e^{-Cm}}{m^{-\frac{2}{k\delta}}} = 0,\end{aligned}$$

we have that $F_i(m) = O(|m|^{-\frac{2}{k\delta}})$ as $m \to -\infty$. Similarly,

$$\lim_{m\to\infty} \frac{1-F_i(m)}{m^{-\frac{2}{k\delta}}} = \lim_{m\to\infty} \frac{\int_m^{\infty} f_i(x)dx}{m^{-\frac{2}{k\delta}}} = 0$$

so that $F_i(m) = O(m^{-\frac{2}{k\delta}})$ as $m \to \infty$. Hence, it follows from the Theorem 1.3. that there exists a sequence of at most n-atom distributions Π_n such that

$$e(\Pi^n) = O\left(n^{\frac{1}{k}\left(-1+\frac{k\delta}{2}\right)}\right) = O\left(n^{-\frac{1}{k}+\frac{\delta}{2}}\right) = o\left(n^{-\frac{1}{k}+\delta}\right).$$

for each $0 < \delta \leq 1/k$.
□

Remark 1.4. Note that densities of the type $Ce^{-h(x)}$ are of the type $e^{-\phi(x)}$ as well ($\phi(x) = h(x) - \ln C$).

Remark 1.5. If P has density then for each sequence Π^n of n-atom probability distributions it holds that $e(\Pi^n) \neq o(n^{-1/k})$

Proof of the Remark. Let Π'^n be a $\lceil n^{1/k} \rceil^k$-atom distribution coinciding with Π_n (we can construct it by adding atoms with zero probability). It holds that

$$\begin{aligned} e(\Pi^n) &= e(\Pi'^n) \overset{(1.14)}{\geq} C\left(\frac{1}{4\lceil n^{1/k}\rceil} - \frac{1}{4\lceil n^{1/k}\rceil^{2k^2-k+1}}\right) \\ &\geq C\left(\frac{1}{4(n^{1/k}+1)} - \frac{1}{4(n^{1/k})^{2k^2-k+1}}\right) \end{aligned}$$

for some fixed C. Since

$$\lim_{n\to\infty} \frac{\frac{1}{4(n^{1/k}+1)} - \frac{1}{4(n^{1/k})^{2k^2-k+1}}}{n^{-1/k}} > 0$$

it cannot be $e(\Pi^n) = o(n^{-1/k})$.
□

1.2.2 Monte Carlo

Another way of evaluating $E_P g$ is Monte-Carlo estimate $\frac{1}{n}\sum_{i=1}^n g(\boldsymbol{X}_i)$ where $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ is an i.i.d. sample from P. The error of approximation then equals to

$$\epsilon_n(g) = |\mathcal{E}_n(g)|, \qquad \mathcal{E}_n(g) := \frac{1}{n}\sum_{i=1}^n g(\boldsymbol{X}_i) - E_P g.$$

Monte Carlo is a well known and widely used method of approximate evaluation of expectations. Its great advantage consists in the fact that one need not necessarily know the distribution P - it suffices to have a sufficiently large sample drawn from the distribution.

Our work formulates some upper bounds of approximation errors quantile

$$\epsilon_n(\alpha, g) := \inf\{e \in \mathrm{R} : P[\epsilon_n(g) < e] \geq 1 - \alpha\}.$$

in case that g is l_1-Lipschitz.

Upper Estimate of $\epsilon_n(\alpha, g)$ Based on Large Deviations

First we state an upper bound of $\epsilon_n(\alpha, g)$ based on the theory of large deviations. Our estimate can hardly be used when n is large (it does not tend to zero as $n \to \infty$), however, it may be useful for a small n.

Theorem 1.4. Denote $\tilde{\boldsymbol{x}}$ the vector of medians of P's marginal distributions and let the probability measure P be such that

$$E_P \exp\{\theta K \|\boldsymbol{X}_1\|_1\} \tag{1.33}$$

exists and is finite for $0 \le \theta \le \theta_0$ where $\theta_0 > 0$. Then it holds, for each $0 \le \theta \le \theta_0$, that

$$\epsilon_n(\alpha, g) \le \hat{\epsilon}_n(\alpha) \tag{1.34}$$

where

$$\begin{aligned} \hat{\epsilon}_n(\alpha) \;:=\; & K \textstyle\sum_{i=1}^k E|\boldsymbol{X}^i - \tilde{\boldsymbol{x}}^i| \\ & + \inf_{0 \le \theta \le \theta_0} \left[\frac{1}{\theta} \ln\left(E \exp\left\{ \theta K \textstyle\sum_{i=1}^k \left| \boldsymbol{X}^i - \tilde{\boldsymbol{x}}^i \right| \right\} \right) - \frac{1}{\theta n} \ln(\alpha) \right]. \end{aligned}$$

Proof. Let $\boldsymbol{X}$ be a random vector with the distribution P and g be l_1-K-Lipschitz function. Using the triangular inequality (*t.i.*), Lipschitz property of g (*L.p.*), Theorem A.1. and the fact that $\boldsymbol{X}_i$, $i = 1, 2, \ldots, k$ are i.i.d. we obtain

$$\begin{aligned} E \exp\{\theta \textstyle\sum_{i=1}^n |g(\boldsymbol{X}_i) - g(\tilde{\boldsymbol{x}})|\} &\overset{L.p.}{\le} E \exp\{\theta \textstyle\sum_{i=1}^n K \|\boldsymbol{X}_i - \tilde{\boldsymbol{x}}\|_1\} \\ &\overset{t.i.}{\le} E \exp\{\theta \textstyle\sum_{i=1}^n K (\|\boldsymbol{X}_i\|_1 + \|\tilde{\boldsymbol{x}}\|_1)\} \\ &= [\exp\{\theta K \|\tilde{\boldsymbol{x}}\|_1\} \times E \exp\{\theta K \|\boldsymbol{X}\|_1\}]^n \\ &\overset{(1.33)}{<} \infty \end{aligned} \tag{1.35}$$

and

$$\begin{aligned} P[|\mathcal{E}_n(g)| \ge x] &= P\left[\left| \left(\textstyle\sum_{i=1}^n g(\boldsymbol{X}_i) \right) - nEg(\boldsymbol{X}) \right| \ge nx \right] \\ &\overset{t.i.}{\le} P\left[\left| \textstyle\sum_{i=1}^n g(\boldsymbol{X}_i) - ng(\tilde{\boldsymbol{x}}) \right| + |ng(\tilde{\boldsymbol{x}}) - nEg(\boldsymbol{X})| \ge nx \right] \\ &= P\left[\left| \textstyle\sum_{i=1}^n g(\boldsymbol{X}_i) - ng(\tilde{\boldsymbol{x}}) \right| + n\,|E[g(\tilde{\boldsymbol{x}}) - g(\boldsymbol{X})]| \ge nx \right] \\ &\overset{t.i.}{\le} P\left[\textstyle\sum_{i=1}^n |g(\boldsymbol{X}_i) - ng(\tilde{\boldsymbol{x}})| + nE\,|g(\tilde{\boldsymbol{x}}) - g(\boldsymbol{X})| \ge nx \right] \\ &= P\left[\textstyle\sum_{i=1}^n |g(\boldsymbol{X}_i) - ng(\tilde{\boldsymbol{x}})| \ge n\,(x - E\,|g(\tilde{\boldsymbol{x}}) - g(\boldsymbol{X})|) \right] \\ &\overset{\text{Theorem A.1.,(1.35)}}{\le} E\left[\exp\left\{ \theta \left(\textstyle\sum_{i=1}^n |g(\boldsymbol{X}_i) - g(\tilde{\boldsymbol{x}})| \right) \right\} \right] \\ &\qquad \times \exp\{-\theta n [x - E|g(\boldsymbol{X}) - g(\tilde{\boldsymbol{x}})|]\} \\ &\overset{i.i.d.}{=} [E\left(\exp\left\{\theta\left[|g(\boldsymbol{X}) - g(\tilde{\boldsymbol{x}})|\right]\right\}\right) \times \exp\{-\theta x\} \\ &\qquad \times \exp\{\theta E|g(\tilde{\boldsymbol{x}}) - g(\boldsymbol{X})|\}]^n \\ &\overset{L.p.}{\le} \left[E \exp\left\{ \theta K \textstyle\sum_{i=1}^k \left| \boldsymbol{X}_1^i - \tilde{\boldsymbol{x}}^i \right| \right\} \times \exp\{-\theta x\} \right. \\ &\qquad \left. \times \exp\left\{ \theta K \textstyle\sum_{i=1}^k E|\boldsymbol{X}_1^i - \tilde{\boldsymbol{x}}^i| \right\} \right]^n \end{aligned} \tag{1.36}$$

for all $x \in \mathrm{R}$. By putting $x = \hat{\epsilon}_n(\alpha, \theta)$ into (1.36) where

$$\hat{\epsilon}_n(\alpha,\theta) := K \sum_{i=1}^k E|\boldsymbol{X}^i - \tilde{\boldsymbol{x}}^i| + \frac{1}{\theta} \ln\left(E \exp\left\{\theta K \sum_{i=1}^k \left|\boldsymbol{X}^i - \tilde{\boldsymbol{x}}^i\right|\right\}\right) - \frac{1}{\theta n} \ln(\alpha)$$

we get that

$$\begin{aligned} P[\mathcal{E}_n(g) \geq \hat{\epsilon}_n(\alpha,\theta)] &\overset{(1.36)}{\leq} \left[E \exp\left\{\theta K \sum_{i=1}^k \left|\boldsymbol{X}^i - \tilde{\boldsymbol{x}}^i\right|\right\}\right. \\ &\quad \left. \times \exp\{-\theta \hat{e}_n(\alpha,\theta)\} \times \exp\left\{\theta K \sum_{i=1}^k E|\boldsymbol{X}^i - \tilde{\boldsymbol{x}}^i|\right\}\right]^n \\ &= \left[E \exp\left\{\theta K \sum_{i=1}^k \left|\boldsymbol{X}^i - \tilde{\boldsymbol{x}}^i\right|\right\}\right. \\ &\quad \times \exp\left\{-\theta \frac{1}{\theta} \ln\left(E \exp\left\{\theta K \sum_{i=1}^k \left|\boldsymbol{X}^i - \tilde{\boldsymbol{x}}^i\right|\right\}\right)\right\} \\ &\quad \times \exp\left\{-\theta \frac{-1}{\theta n} \ln(\alpha)\right\} \\ &\quad \times \exp\left\{-\theta K \sum_{i=1}^k E|\boldsymbol{X}^i - \tilde{\boldsymbol{x}}^i|\right\} \\ &\quad \left. \times \exp\left\{\theta K \sum_{i=1}^k E|\boldsymbol{X}^i - \tilde{\boldsymbol{x}}^i|\right\}\right]^n \\ &= \alpha. \end{aligned} \tag{1.37}$$

Using it we obtain

$$\begin{aligned} \epsilon_n(\alpha, g) &= \inf\{e \in \mathrm{R} : P[|\mathcal{E}_n(g)| < e] \geq 1 - \alpha\} \\ &= \inf\{e \in \mathrm{R} : P[|\mathcal{E}_n(g)| \geq e] \leq \alpha\} \overset{(1.37)}{\leq} \hat{\epsilon}_n(\alpha, \theta) \end{aligned}$$

for each $0 < \theta < \theta_0$ which proves (1.34).
□

Upper Estimate of $\epsilon_n(\alpha, g)$ Based on Refinement of Central Limit Theorem

Since the upper bound stated by Theorem 1.4. is suitable only for the small n, another upper bound is needed. We start with several auxiliary results.

Lemma 1.6. (Bounds of variance of l_1-K-Lipschitz functions) Let marginal distributions of P have finite variances. Then

$$\underline{V}^2 \leq \sup_{g \text{ is } l_1\text{-}K\text{-Lipschitz}} \mathrm{var}_P\, g \leq \overline{V}^2 \tag{1.38}$$

where

$$\underline{V}^2 = K^2 \sum_{i_1=1}^k \sum_{i_2=1}^k v_{i_1,i_2} \quad \text{and} \quad \overline{V}^2 = K^2 \sum_{i_1=1}^k \sum_{i_2=1}^k \sqrt{v_{i_1,i_1} v_{i_2,i_2}}. \tag{1.39}$$

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

Proof. Let random vector $\boldsymbol{X}$ have distribution P.
Ad. the first "$\leq$". Put $g(\boldsymbol{x}) = K\sum_{i=1}^{k}\boldsymbol{x}^i$. Since g is l_1-K-Lipschitz function and since

$$\mathrm{var}(g(\boldsymbol{X})) = \underline{V}^2$$

the assertion is proved.
Ad the second "$\leq$". Let g be l_1-K-Lipschitz. Then it holds that

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

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

Lemma 1.7. (Maximal third moment of l_1-K-Lipschitz functions) Denote $e_i := E|\boldsymbol{X}^i - E\boldsymbol{X}^i|$, $i = 1, 2, \ldots, k$ and let

$$M_i^3 := E|\boldsymbol{X}^i - E\boldsymbol{X}^i|^3 < \infty, \qquad i = 1, 2, \ldots, k. \tag{1.40}$$

Then it holds that

$$\sup_{g \text{ is } l_1\text{-}K\text{-Lipschitz}} E|g(\boldsymbol{X}) - Eg(\boldsymbol{X})|^3 \leq \overline{M}^3$$

where

$$\overline{M} = K\left[\sqrt[3]{\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\sum_{i_3=1}^{k}M_{i_1}M_{i_2}M_{i_3}} + \sum_{i=1}^{k}e_i\right] \tag{1.41}$$

Proof. Let random vector $\boldsymbol{X}$ have distribution P. We use the Lipschitz inequality ($L.I.$), the triangular inequality ($T.I.$), the Minkowski inequality $(E|X+Y|^p)^{\frac{1}{p}} \leq (E|X|^p)^{\frac{1}{p}} + (E|Y|^p)^{\frac{1}{p}}$ for $p \geq 1$ ($M.I.$), the Hőlder inequality ($H.I.$) and the Schwarz inequality ($S.I.$) in our proof (see [45], p. 143, for $M.I.$, $H.I.$ and $S.I.$).

It holds, for arbitrary random variables X, Y and Z, that

$$\begin{aligned}
E|XYZ| &\overset{H.I.}{\leq} \left(E|X|^3\right)^{\frac{1}{3}}\left(E|Y|^{\frac{3}{2}}|Z|^{\frac{3}{2}}\right)^{\frac{2}{3}} \\
&\overset{S.I.}{\leq} \left(E|X|^3\right)^{\frac{1}{3}}\left(\left[E\left(|Y|^{\frac{3}{2}}\right)^2 E\left(|Z|^{\frac{3}{2}}\right)^2\right]^{\frac{1}{2}}\right)^{\frac{2}{3}} \\
&= \left(E|X|^3\right)^{\frac{1}{3}}\left(E|Y|^3\right)^{\frac{1}{3}}\left(E|Z|^3\right)^{\frac{1}{3}}
\end{aligned} \tag{1.42}$$

so that

$$\begin{aligned}
\sqrt[3]{E|g(\boldsymbol{X})-Eg(\boldsymbol{X})|^3} &= \sqrt[3]{E|g(\boldsymbol{X})-g(E\boldsymbol{X})+g(E\boldsymbol{X})-Eg(\boldsymbol{X})|^3} \\
&\overset{M.I.}{\leq} \sqrt[3]{E|g(\boldsymbol{X})-g(E\boldsymbol{X})|^3}+\sqrt[3]{|g(E\boldsymbol{X})-Eg(\boldsymbol{X})|^3} \\
&= \sqrt[3]{E|g(\boldsymbol{X})-g(E\boldsymbol{X})|^3}+|E(g(E\boldsymbol{X})-g(\boldsymbol{X}))| \\
&\overset{T.I.}{\leq} \sqrt[3]{E|g(\boldsymbol{X})-g(E\boldsymbol{X})|^3}+E|g(E\boldsymbol{X})-g(\boldsymbol{X})| \\
&\overset{L.I.}{\leq} K\sqrt[3]{E\left[\sum_{i=1}^{k}|\boldsymbol{X}^i-E\boldsymbol{X}^i|\right]^3}+K\sum_{i=1}^{k}E|\boldsymbol{X}^i-E\boldsymbol{X}^i| \\
&= K\sqrt[3]{\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\sum_{i_3=1}^{k}E|\boldsymbol{X}^{i_1}-E\boldsymbol{X}^{i_1}||\boldsymbol{X}^{i_2}-E\boldsymbol{X}^{i_2}||\boldsymbol{X}^{i_3}-E\boldsymbol{X}^{i_3}|} \\
&\quad +K\sum_{i=1}^{k}E|\boldsymbol{X}^i-E\boldsymbol{X}^i| \\
&\overset{(1.42)}{\leq} K\sqrt[3]{\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\sum_{i_3=1}^{k}\sqrt[3]{E|\boldsymbol{X}^{i_1}-E\boldsymbol{X}^{i_1}|^3}\sqrt[3]{E|\boldsymbol{X}^{i_2}-E\boldsymbol{X}^{i_2}|^3}\sqrt[3]{E|\boldsymbol{X}^{i_3}-E\boldsymbol{X}^{i_3}|^3}} \\
&\quad +K\sum_{i=1}^{k}E|\boldsymbol{X}^i-E\boldsymbol{X}^i| = K\left[\sqrt[3]{\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\sum_{i_3=1}^{k}M_{i_1}M_{i_2}M_{i_3}}+\sum_{i=1}^{k}e_i\right].
\end{aligned}$$

□

The following Lemma provides an upper bound of $P[\epsilon_n(g)\geq\Delta]$.[2]

Lemma 1.8. Let the marginals of P have finite third moments. Then, for each l_1-K-Lipschitz function g, for each constant $\Delta > 3\overline{M}n^{-2/3}e^{1/3}$ and for each $\delta\in\mathrm{R}$ such that

$$\frac{\overline{M}\sqrt[3]{e}}{\sqrt[6]{n}}\leq\delta\leq\frac{\sqrt{n}\Delta}{3} \tag{1.43}$$

it holds that

$$P[\epsilon_n(g)\geq\Delta]\leq U_n(\Delta,\delta) \tag{1.44}$$

where

$$U_n(\Delta,\delta)=2\Phi\left(-\frac{\sqrt{n}\Delta-3\delta}{\overline{V}}\right)+\frac{90}{1-e^{-1}}\cdot\frac{\overline{M}^3}{\sqrt{n}\delta^3}\left[1-\ln\left(\frac{\overline{M}^3}{\sqrt{n}\delta^3}\right)\right], \tag{1.45}$$

Φ is the standard normal distribution function and the values $\overline{V}$ and $\overline{M}$ are defined by Lemma 1.6., Lemma 1.7. respectively. Moreover, there exists $n_0\in\mathrm{N}$,

$$\lim_{n\geq n_0, n\to\infty}U_n(\Delta,\delta)=0.$$

[2]One may ask why we do not use the well known Berry Esseén inequality instead of constructing a new bound. The reason is that one has to divide by the standard deviation of the converging variables to compute the B-E bound. However, it is usually impossible to compute the standard deviation if we are not able to compute the expectation. We even cannot make an upper estimate of the B-E bound: to do so we would have to use a lower estimate of the standard deviation which we cannot have if we want a bound valid simultaneously for all the Lipschitz functions.

We shall use the following result in the proof of the Lemma.

Lemma 1.9. Let $\xi_1, \ldots, \xi_n$ be independent random variables with $E\xi_i = 0$ for each i and $\beta = \sum_i E|\xi_i|^3$ finite. Let $S := \xi_1 + \ldots + \xi_n$. Then there exists random variable $T \sim N(0, \operatorname{var} S)$ such that, for each Borel $A \subset \mathrm{R}$,

$$P\left[S \in A\right] \leq P[T \in A^{3\delta}] + C_0 B\left(1 + |\ln(1/B)|\right)$$

where $B := \beta\delta^{-3}$ and $C_0 = \frac{90}{1-e^{-1}}$ (A^ϵ denotes the ϵ-neighborhood of the set A).

Proof. See [37], p. 245, proof of Theorem $< 10 >$, formula $< 15 >$ and the last formula of the proof.

□

Proof of Lemma 1.8. If $\operatorname{var}(g)$ then the assertion is trivial. Assume that $\operatorname{var}(g) > 0$. Denote

$$M_n^3 := \sum_{i=1}^n E\left|n^{-1/2}(g(\boldsymbol{X}_i) - E_P g)\right|^3.$$

It follows from Lemma 1.9. that, for each $d > 0$,

$$P\left[\sqrt{n}\epsilon_n(g) \geq d\right] = P\left[|\sqrt{n}\mathcal{E}_n(g)| \geq d\right] \leq P\left[|T| \geq d - 3\delta\right] + C_0\psi(\delta^{-3}M_n^3) \tag{1.46}$$

where $\psi(B) = B(1 + |\ln(1/B)|)$ and T is a random variable with distribution $N(0, \operatorname{var}(g))$ (we have put $\xi_i = n^{-1/2}(g(\boldsymbol{X}_i) - E_P g)$ and $A = (-\infty, -d] \cup [d, \infty)$ in Lemma 1.9.).

Since

$$\begin{aligned} M_n^3 &= \sum_{i=1}^n E\left|\frac{1}{\sqrt{n}}(g(\boldsymbol{X}_i) - Eg(\boldsymbol{X}_i))\right|^3 = \frac{n}{n^{3/2}} E\left|g(\boldsymbol{X}_1) - Eg(\boldsymbol{X}_1)\right|^3 \\ &= \frac{1}{\sqrt{n}} E\left|g(\boldsymbol{X}_1) - Eg(\boldsymbol{X}_1)\right|^3 \overset{\text{Lemma 1.7.}}{\leq} \frac{\overline{M}^3}{\sqrt{n}} \end{aligned}$$

and due to (1.43) it holds that

$$\delta^{-3}M_n^3 \leq \delta^{-3}n^{-1/2}\overline{M}^3 \leq e^{-1}. \tag{1.47}$$

Since ψ is nondecreasing on $(0, 1]$ (if $B < 1$ then it holds that $\psi'(B) = [B(1 + |\ln(1/B)|)]'$ $= [B(1 + \ln(1/B))]' = [B(1 - \ln(B))]' = 1 - \ln(B) - \frac{B}{B} = -\ln(B) \geq 0$) we have

$$\psi(\delta^{-3}M_n^3) \overset{(1.47)}{\leq} \psi\left(\frac{\overline{M}^3}{\delta^3\sqrt{n}}\right). \tag{1.48}$$

By using it and by putting $d = \sqrt{n}\Delta$ we get

$$\begin{aligned} &P\left[|\sqrt{n}\mathcal{E}_n(g)| > \sqrt{n}\Delta\right] - C_0\psi\left(\overline{M}^3\delta^{-3}n^{-1/2}\right) \\ &\overset{(1.46),(1.48)}{\leq} P\left[|T| \geq \sqrt{n}\Delta - 3\delta\right] = P\left[|N(0,1)| \geq \frac{\sqrt{n}\Delta - 3\delta}{\sqrt{\operatorname{var}(g)}}\right] \\ &\overset{\delta \leq \frac{\sqrt{n}\Delta}{3}}{=} \left[1 - \Phi\left(\frac{\sqrt{n}\Delta - 3\delta}{\sqrt{\operatorname{var}(g)}}\right)\right] + \Phi\left(-\frac{\sqrt{n}\Delta - 3\delta}{\sqrt{\operatorname{var}(g)}}\right) \\ &= 2\Phi\left(-\frac{\sqrt{n}\Delta - 3\delta}{\sqrt{\operatorname{var}(g)}}\right) \overset{\text{Lemma 1.6.}}{\leq} 2\Phi\left(-\frac{\sqrt{n}\Delta - 3\delta}{\overline{V}}\right) \end{aligned}$$

which proves (1.44).

Ad. the convergence. $U_n(\Delta,\delta)$ can be rewritten as

$$2\Phi(-n^{1/2}C+D)+En^{-1/2}-Fn^{-1/2}\ln(Gn^{-1/2})$$

where $C,D,E,F,G\geq 0$. The first and the second summand trivially go to zero, the third one is easy to prove to do so:

$$\lim_{n\to\infty} n^{-1/2}\ln(Gn^{-1/2})=\lim_{x\to 0} x\ln(Gx)=\lim_{x\to 0}\frac{1/x}{-1/x^2}=0$$

(we have used the well known l'Hospital rule).
□

When we fix n and Δ, the smallest of the upper bounds, stated by the previous Lemma, is

$$\underline{U}_n(\Delta):=\min_{\overline{M}n^{-1/6}e^{1/3}\leq\delta\leq n^{1/2}\Delta/3} U_n(\Delta,\delta). \tag{1.49}$$

Lemma 1.10. Under the assumptions of the previous Lemma,

$$P[\epsilon_n(g)\geq\Delta]\leq\underline{U}_n(\Delta) \tag{1.50}$$

where $\underline{U}_n(\Delta)$ is defined by (1.49). Moreover, the problem (1.49) is convex (i. e. both its feasibility set and its objective function are convex).

Proof. The formula (1.50) follows directly from Lemma 1.8.

Ad. the convexity of (1.49): Since the sum of two convex functions is convex, to prove the convexity of $U_n(\Delta,\bullet)$ it suffices to show the convexity of both $\psi(\overline{M}^3n^{-1/2}\delta^{-3})$ and

$$U_1(\delta):=\Phi\left(-\frac{\sqrt{n}\Delta-3\delta}{\overline{V}}\right)$$

(ψ is defined in the proof of the previous Lemma).

Ad. U_1: $\partial U_1/\partial\delta=D\varphi((3\delta-\sqrt{n}\Delta)/\overline{V})$, where D is a positive constant and φ is the standard normal density. Since $\partial U_1/\partial\delta$ is nondecreasing on $(-\infty,\sqrt{n}\Delta/3]$, the function U_1 is convex therein.

Ad $\psi(\overline{M}^3n^{-1/2}\delta^{-3})$: If h is a differentiable function such that $h<1$ then

$$[\psi(h)]''=(h(1-\ln h)]''=[h'-(h'\ln h+h')]'=-[h'\ln h]'=-\left[h''\ln h+(h')^2\frac{1}{h}\right]$$

When we put $h(\delta)=C\delta^{-3}$ where $C=\overline{M}^3n^{-1/2}$ we get that

$$\begin{aligned}[\psi(C\delta^{-3})]'' &= -\left[12C\delta^{-5}\ln(C\delta^{-3})+9C^2\delta^{-8}\frac{1}{C\delta^{-3}}\right]=-[12C\delta^{-5}\ln(C\delta^{-3})+9C\delta^{-5}]\\ &= -C\delta^{-5}[12\ln(C\delta^{-3})+9]\end{aligned}$$

whenever $C\delta^{-3}<1$. Clearly, $[\psi(C\delta^{-3})]''\geq 0$ whenever $C\delta^{-3}\leq e^{-1}$. Since the latter inequality holds true for each $\delta\in[\overline{M}n^{-1/6}e^{1/3},\infty)$, the function $\psi(C\delta^{-3})$ has to be convex on the feasibility set of (1.49).
□

Remark 1.6. To find the solution of (1.49) it suffices to check the boundary points of the feasibility set of (1.49) and the global minimum of the function $U_n(\Delta, \bullet)$ which is uniquely determined by equation

$$0 = \frac{6}{\overline{V}}\varphi\left(-\frac{\sqrt{n}\Delta - 3\delta}{\overline{V}}\right) + \frac{90}{1-e^{-1}} \cdot \frac{3\overline{M}^3}{\sqrt{n}\delta^4} \ln\left(\frac{\overline{M}^3}{\sqrt{n}\delta^3}\right)$$

(we have used the fact, proved in the proof of Lemma 1.8., that $\psi'(B) = -\ln(B)$ for each $B < 1$, φ denotes the standard normal distributions density). Even if we are unable to solve the equation analytically. the numerical solution is easy.

Finally, we can state an upper bound of the Monte Carlo error's quantile.

Theorem 1.5. Let the marginal distributions of the distribution P have finite third moments, let $\alpha \in (0,1)$ and let g be l_1-K-Lipschitz function. Then it holds that

$$\epsilon_n(\alpha, g) \leq \tilde{\epsilon}_n(\alpha) \tag{1.51}$$

where

$$\begin{aligned} \tilde{\epsilon}_n(\alpha) \quad := \quad & \min_{\Delta,\delta} \Delta && (1.52)\\ & \textit{subject to} \\ & U_n(\Delta,\delta) \leq \alpha, && (1.53)\\ & \overline{M} n^{-1/6} e^{1/3} \leq \delta, && (1.54)\\ & \delta \leq n^{1/2}\Delta/3 && (1.55) \end{aligned}$$

(U_n is defined by (1.45)). Moreover, the problem (1.52) is convex and it holds that

$$\lim_{n\to\infty} \tilde{\epsilon}_n(\alpha) = 0.$$

Proof. Fix α and n. First we show that the right hand side of (1.52) is well defined: Since

$$\lim_{D\to\infty} U_n(D, n^{1/2}D/6) = 0,$$

there exists $D_1 \in \mathrm{R}$ such that the vector $(\Delta,\delta) = (D, n^{1/2}D/6)$ fulfils (1.53) for each $D > D_1$. Since $(D, n^{1/2}D/6)$ trivially fulfils (1.55) for each $D \geq 0$ and since there exists $D_2 \in \mathrm{R}$ such that $(D, n^{1/2}D/6)$ fulfils (1.54) for each $D \geq D_2$, the problem (1.52) has to have at least one feasible solution. Moreover, $\tilde{\epsilon}_n(\alpha) \leq \max(D_1, D_2)$, i.e. the minimum exists.

Further, we show the convexity of the problem described by (1.52), (1.53), (1.54), (1.55). The function U_n can be decomposed as

$$U_n(\Delta,\delta) = 2\Phi(L(\Delta,\delta)) + \psi(\overline{M}^3 n^{-1/2}\delta^{-3})$$

where Φ is the standard normal distribution function, ψ is defined in the proof of Lemma 1.8. and L is some linear mapping from R^2 into R. It was shown in the proof of Lemma 1.10.that $\psi(\overline{M}^3 n^{-1/2}\delta^{-3})$ is convex in δ on the set formed by (1.54) and (1.55). We prove that $\Phi(L(\bullet,\bullet))$ is convex on the same set. As mentioned above, Φ is convex for negative arguments. However, the negativity of L is assured by (1.55). Since the superposition of a linear and a convex functions is convex, $\Phi(L(\Delta,\delta))$ is convex on the set formed by (1.55). Hence, U_n is

convex on the (convex) set formed by (1.54) and by (1.55). Therefore, the whole feasibility set is convex. Because the objective function is linear, the problem (1.52) is convex.

Finally, we prove (1.51): Denote by $(\hat{\Delta}, \hat{\delta})$ the optimal solution of the problem (1.52). Since it follows from (1.54) and from (1.55) that $\hat{\Delta} \geq 3\overline{M}n^{-2/3}e^{1/3}$ and that $\hat{\delta}$ fulfils (1.43), we may use Lemma 1.8. to get

$$P[\epsilon_n(g) \geq \tilde{\epsilon}_n(\alpha)] \;=\; P[\epsilon_n(g) \geq \hat{\Delta}] \overset{\text{Lemma 1.8.,(1.54),(1.55)}}{\leq} U_n(\hat{\Delta}, \hat{\delta}) \overset{(1.53)}{\leq} \alpha. \tag{1.56}$$

Therefore

$$\begin{aligned} \epsilon_n(\alpha, g) &= \inf\{e \in \mathrm{R} : P[|\mathcal{E}_n(g)| < e] \geq 1 - \alpha\} \\ &= \inf\{e \in \mathrm{R} : P[|\mathcal{E}_n(g)| \geq e] \leq \alpha\} \overset{(1.56)}{\leq} \tilde{e}(\alpha). \end{aligned}$$

Ad. the convergence. Let $\epsilon > 0$, $0 < \Delta \leq \epsilon$ and $\delta > 0$. It follows from (1.54), from (1.55) and from the fact that $\lim_{n\to\infty} U_n(\Delta, \delta) = 0$ (see Lemma 1.8.) that there exists n_0 such that (Δ, δ) is feasible solution of (1.52) for $n \geq n_0$. Therefore, $\tilde{\epsilon}_n(\alpha) \leq \epsilon$ for $n \geq n_0$.
□

Limit Behavior of $\epsilon_n(\alpha, g)$

The following Theorem describes the limit of approximation error's quantile.

Theorem 1.6. Denote Φ^{-1} the inverse of the standard normal distribution's d.f. and assume that the distribution P has finite variance matrix $(v_{i_1,i_2})_{i_1=1\ldots k, i_2=1\ldots k}$. Let g be l_1-K-Lipschitz function. Then

$$\lim_{n\to\infty} \sqrt{n}\, \epsilon_n(\alpha, g) \leq \bar{\epsilon}(\alpha) \quad \text{where} \quad \bar{\epsilon}(\alpha) := \overline{V}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{1.57}$$

where $\overline{V}$ is defined by (1.39).

We shall use the following result in the proof of the Theorem.

Lemma 1.11. Let random variables Y_n with d.f.'s F_n, $n = 1, 2, \ldots$, converge in distribution to a random variable Y with an increasing d.f. F. Then

$$\lim_{n\to\infty} F_n^{-1}(\beta) \longrightarrow F^{-1}(\beta)$$

for each $0 < \beta < 1$.

Proof. The assertion follows directly from [14], par. 5.37.
□

Proof of the Theorem 1.6. Since, according to Lemma 1.6. and to the assumptions of the present Theorem,

$$\mathrm{var}(\sqrt{n}\mathcal{E}_n(g)) = \mathrm{var}\, g \leq \overline{V} < \infty$$

we may apply the Central Limit Theorem to get that

$$\sqrt{n}\mathcal{E}_n(g) \xrightarrow{\mathcal{D}} N(0, \mathrm{var}(g)).$$

By application of the Lemma 1.11. we get that

$$\sqrt{n}\epsilon_n(\alpha, g) \overset{n\to\infty}{\longrightarrow} \sqrt{\mathrm{var}(g)}\Phi^{-1}(1-\alpha/2) \leq \overline{V}\Phi^{-1}(1-\alpha/2).$$

□

Corollary. If P has finite variance matrix then

$$\epsilon_n(\alpha, g) = O(n^{-1/2}) \quad \text{as } n \to \infty.$$

for each l_1-Lipschitz function g.

Proof. The assertion follows directly from Theorem 1.6.
□

1.2.3 Quasi Monte Carlo

Another way of the approximate computation of $E_P g$ is using a combination of discretization and Monte Carlo. In the present work, we call the technique Quasi Monte Carlo but it should be noted here that the name Quasi Monte Carlo is being used for a wide range of techniques, most of them using quasi-random sequences instead of random ones (see [33]).

The author is grateful to professor Dupač who worked in this area already fifty years ago (see [7]) and who inspired him to incorporate Quasi Monte Carlo into the present work.[3]

Before defining the Quasi Monte Carlo estimate, we shall formulate a useful assertion. It is probably well known, however, the author could not find it in the literature, available to him. The assertion provides a way how to get any $k-$dimensional random vector by a transformation of a vector with a k-dimensional uniform distribution.

Lemma 1.12. Let $\boldsymbol{U}$ be a k-dimensional random vector with uniform distribution on $[0,1] \times [0,1] \times \ldots \times [0,1]$ and let $\boldsymbol{X}$ be a k-dimensional real random vector. Let $G_1(\bullet)$ be d. f. of $\mathcal{L}(\boldsymbol{X}^1)$ and, for each $i = 2, 3, \ldots, k$, let $G_i(\bullet|\boldsymbol{X}^1, \boldsymbol{X}^2, \ldots, \boldsymbol{X}^{i-1})$ be a d. f. of $\mathcal{L}(\boldsymbol{X}^i|\boldsymbol{X}^1, \boldsymbol{X}^2, \ldots, \boldsymbol{X}^{i-1})$. Denote $Q = \mathcal{L}(\boldsymbol{X})$ and define

$$\boldsymbol{t}_Q : \prod_{i=1}^{k}[0,1] \to \mathrm{R}^k,$$

$$\boldsymbol{t}_Q^1(\boldsymbol{u}) = G^{-1}(\boldsymbol{u}^1),$$

$$\boldsymbol{t}_Q^i(\boldsymbol{u}) = G_i^{-1}(\boldsymbol{u}^i|\boldsymbol{t}_Q^1(\boldsymbol{u}), \boldsymbol{t}_Q^2(\boldsymbol{u}), \ldots, \boldsymbol{t}_Q^{i-1}(\boldsymbol{u}))$$

for $1 < i < k$. Then $\boldsymbol{t}_Q$ is measurable with respect to the Borel σ-algebra of the k-dimensional unit cube and

$$\mathcal{L}(\boldsymbol{t}_Q(\boldsymbol{U})) = Q. \tag{1.58}$$

Before we proceed, we formulate some helpful results. The first one is well known:

Lemma 1.13. If U is a random variable having one-dimensional uniform distribution and if G is a distribution function then the random variable $G^{-1}(U)$ has distribution function G.

Proof. See [1], Theorem 4 on p. 15 and the notes below.
□

[3] Our technique is analogical to stratified sampling, particulary the case when exactly one observation is drawn from each stratum.

Lemma 1.14. If G is distribution function then $G^{-1}(\alpha) \leq x$ if and only if $G(x+) \geq \alpha$.

Proof. If $G(x+) \geq \alpha$ then there exists $x_\nu > x$, $x_\nu \to x$ such that $G(x_\nu) \geq \alpha$ for each ν so that $G^{-1}(\alpha) = \inf\{e : G(e) \geq \alpha\} \leq x$. If $G^{-1}(\alpha) = \inf\{e : G(e) \geq \alpha\} \leq x$ then, for each $x_\nu > x$, it has to hold that $G(x_\nu) \geq \alpha$ (otherwise $G^{-1}(\alpha) > x$) so that $G(x+) = \lim_\nu G(x_\nu) \geq \alpha$ by the limit transition.
□

Proof of Lemma 1.12. Since R^k is a complete separable metric space, $\mathcal{L}(\boldsymbol{X}^i|\boldsymbol{X}^1,\ldots,\boldsymbol{X}^{i-1})$ exists according to [45], VI.1.21., hence the definition of $\boldsymbol{t}_Q(\boldsymbol{U})$ is correct.

Proof of measurability and of (1.58). We proceed by induction: We get from Lemma 1.14.

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

for each $x \in \mathrm{R}$ and, from Lemma 1.13.,

$$\mathcal{L}(\boldsymbol{t}_Q^1(\boldsymbol{U}^1)) = \mathcal{L}(\boldsymbol{X}^1),$$

i.e. $\boldsymbol{t}_Q^1$ is measurable and (1.58) holds for $k = 1$.

Induction step: Let $k > 1$. Assume that vector mapping $(\boldsymbol{t}_Q^1, \boldsymbol{t}_Q^2, \ldots, \boldsymbol{t}_Q^{k-1})'$ 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

$$\mathcal{L}((\boldsymbol{t}_Q^1(\boldsymbol{U}^1), \boldsymbol{t}_Q^2(\boldsymbol{U}^1,\boldsymbol{U}^2), \ldots, \boldsymbol{t}_Q^{k-1}(\boldsymbol{U}^1,\ldots,\boldsymbol{U}^{k-1}))) = \mathcal{L}((\boldsymbol{X}^1,\ldots,\boldsymbol{X}^{k-1})). \tag{1.59}$$

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

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

It holds that

$$\begin{aligned}
(\boldsymbol{t}_Q^k)^{-1}((-\infty,t]) &= \{\boldsymbol{u} : \boldsymbol{t}_Q^k(\boldsymbol{u}) \leq t\} = \{\boldsymbol{u} : G_k^{-1}(\boldsymbol{u}^k|\boldsymbol{t}_Q^1(\bar{\boldsymbol{u}}^1), \ldots, \boldsymbol{t}_Q^{k-1}(\bar{\boldsymbol{u}}^{k-1})) \leq t\} \\
&\overset{\text{Lemma 1.14.}}{=} \{\boldsymbol{u} : G_k(t + |\boldsymbol{t}_Q^1(\bar{\boldsymbol{u}}^1), \ldots, \boldsymbol{t}_Q^{k-1}(\bar{\boldsymbol{u}}^{k-1})) \geq \boldsymbol{u}^k\} \\
&= \{\boldsymbol{u} : G_k(t + |\boldsymbol{t}_Q^1(\bar{\boldsymbol{u}}^1), \ldots, \boldsymbol{t}_Q^{k-1}(\bar{\boldsymbol{u}}^{k-1})) - \boldsymbol{u}^k \geq 0\}.
\end{aligned} \tag{1.60}$$

Since $G_k(t + |\bullet)$ is the conditional probability of Borel set $(0,t]$, it is measurable according to [45], VI.1.1. Therefore and due to the facts that $(\boldsymbol{t}^1,\ldots,\boldsymbol{t}^{k-1})$ is measurable by induction assumption, that the superposition of two measurable mappings is measurable (see [30], A.3.5 (c)), that identity is measurable and that the sum of two measurable functions is measurable (see [30] A.3.5.(a)) the function defining the set (1.60) is measurable hence the set is Borel.

Ad. (1.58). Denote $R = \mathcal{L}(\boldsymbol{t}_Q(\boldsymbol{U}))$,

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

and $\bar{Q} = \mathcal{L}(\bar{\boldsymbol{t}}_Q)$. According to the induction assumption, the marginal distribution of the $R's$ first $k-1$ coordinates is identical to $\bar{Q}$. To get $R = Q$ it suffices to show that $Q[C] = R[C]$ for each $C = A \times (-\infty, b)$ where $A \subseteq \mathrm{R}^{k-1}$ is Borel set and $b \in \mathrm{R}$.

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

$$\begin{aligned}
R[C] &= \int_C dR(\boldsymbol{\tau}) = \int_{\mathrm{R}^k} I_C(\boldsymbol{\tau}) dR(\boldsymbol{\tau}) \overset{R=\mathcal{L}(\boldsymbol{t}_Q)}{=} \int_{H_k} I_C(\boldsymbol{t}_Q(u_1, u_2, \ldots, u_k)) du_1 du_2 \ldots du_k \\
&= \int_{H_k} I_{(-\infty,b)}(\boldsymbol{t}_Q^k(u_1, u_2, \ldots, u_k)) I_A(\bar{\boldsymbol{t}}_Q(u_1, u_2, \ldots, u_{k-1})) du_1 du_2 \ldots du_k \\
&= \int_{H_{k-1}} \left[\int_{[0,1]} I_{(-\infty,b)}(\boldsymbol{t}_Q^k(u_1, u_2, \ldots, u_k)) du_k \right] \\
&\qquad \cdot I_A(\bar{\boldsymbol{t}}_Q(u_1, u_2, \ldots, u_{k-1})) du_1 du_2 \ldots du_{k-1} \\
&= \int_{H_{k-1}} \left[\int_{[0,1]} I_{(-\infty,b)}(G_k^{-1}(u_k | \bar{\boldsymbol{t}}_Q(u_1, u_2, \ldots, u_{k-1}))) du_k \right] \\
&\qquad \cdot I_A(\bar{\boldsymbol{t}}_Q(u_1, u_2, \ldots, u_{k-1})) du_1 du_2 \ldots du_{k-1} \\
&\overset{\bar{Q}=\mathcal{L}(\bar{t}_Q)}{=} \int_{\mathrm{R}^{k-1}} \left[\int_{[0,1]} I_{(-\infty,b)}(G_k^{-1}(u_k | t_1, t_2, \ldots, t_{k-1})) du_k \right] \\
&\qquad \cdot I_A(t_1, t_2, \ldots, t_{k-1}) d\bar{Q}(t_1, t_2, \ldots t_{k-1}) \\
&= \int_A \left[\int_{[0,1]} I_{(-\infty,b)}(G_k^{-1}(u_k | t_1, t_2, \ldots, t_{k-1})) du_k \right] d\bar{Q}(t_1, t_2, \ldots t_{k-1}) \\
&\overset{\text{Lemma 1.13.}}{=} \int_A \left[\int_{\mathrm{R}} I_{(-\infty,b)}(t_k) dG_k(t_k | t_1, t_2, \ldots, t_k)) \right] d\bar{Q}(t_1, t_2, \ldots t_{k-1}) \\
&= \int_A G_k(b | t_1, t_2, \ldots, t_{k-1}) d\bar{Q}(t_1, t_2, \ldots, t_{k-1}) = Q[C]
\end{aligned}$$

by the definition of conditional probability.
□

Denote $m_n = \lfloor \sqrt[k]{n} \rfloor$. Our approximation method consists in the following steps:

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

The resulting value serves us as an estimator of $E_p g$. Mathematically speaking, our estimator equals to

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

where $\boldsymbol{U}_{i_1,i_2,\ldots,i_k}$ is a random vector having uniform distribution on the cube $[\frac{i_1-1}{m_n}, \frac{i_1}{m_n}] \times [\frac{i_2-1}{m_n}, \frac{i_2}{m_n}] \times \ldots \times [\frac{i_k-1}{m_n}, \frac{i_k}{m_n}]$ 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

$$\Gamma_n(g) := C_n(g) - E_P g \tag{1.61}$$

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

$$
\begin{aligned}
\Gamma_n(g) &= C_n(g) - E_P g \overset{\text{Lemma 1.12.}}{=} C_n(g) - E(g(\boldsymbol{t}_P(\boldsymbol{U})) \\
&= C_n(g) - \int_{[0,1]\times\ldots\times[0,1]} g(\boldsymbol{t}_P(\boldsymbol{u}))d\boldsymbol{u} \\
&= \frac{1}{m_n^k}\sum_{i_1=1}^{m_n}\sum_{i_2=1}^{m_n}\cdots\sum_{i_k=1}^{m_n} g(\boldsymbol{t}_P(\boldsymbol{U}_{i_1,i_2,\ldots,i_k})) \\
&\qquad -\frac{1}{m_n^k}\sum_{i_1=1}^{m_n}\sum_{i_2=1}^{m_n}\cdots\sum_{i_k=1}^{m_n}\int_{[\frac{i_1-1}{m_n},\frac{i_1}{m_n}]\times\ldots\times[\frac{i_k-1}{m_n},\frac{i_k}{m_n}]} g(\boldsymbol{t}_P(\boldsymbol{u}))m_n^k d\boldsymbol{u} \\
&= \frac{1}{m_n^k}\sum_{i_1=1}^{m_n}\sum_{i_2=1}^{m_n}\cdots\sum_{i_k=1}^{m_n} \\
&\qquad \left[g(\boldsymbol{t}_P(\boldsymbol{U}_{i_1,i_2,\ldots,i_k})) - \int_{[\frac{i_1-1}{m_n},\frac{i_1}{m_n}]\times\ldots\times[\frac{i_k-1}{m_n},\frac{i_k}{m_n}]} g(\boldsymbol{t}_P(\boldsymbol{u}))m_n^k d\boldsymbol{u}\right] \\
&= \frac{1}{m_n^k}\sum_{i_1=1}^{m_n}\sum_{i_2=1}^{m_n}\cdots\sum_{i_k=1}^{m_n}\left[g(\boldsymbol{t}_P(\boldsymbol{U}_{i_1,i_2,\ldots,i_k})) - Eg(\boldsymbol{t}_P(\boldsymbol{U}_{i_1,i_2,\ldots,i_k}))\right]. \qquad (1.62)
\end{aligned}
$$

Therefore

$$
E\Gamma_n(g) \overset{(1.62)}{=} \frac{1}{m_n^k}\sum_{i_1=1}^{m_n}\sum_{i_2=1}^{m_n}\cdots\sum_{i_k=1}^{m_n}\left[Eg(\boldsymbol{t}_P(\boldsymbol{U}_{i_1,i_2,\ldots,i_k})) - Eg(\boldsymbol{t}_P(\boldsymbol{U}_{i_1,i_2,\ldots,i_k}))\right] = 0
$$

i.e. $C_n(g)$ is an unbiased estimator of $E_P(g)$. We examine its variance and its asymptotic properties.

Theorem 1.7. Let $n \geq 2^k$ and let function g be l_1-K-Lipschitz. Moreover, 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 $x_1, x_2, \ldots, x_{\lambda-1} \in \mathrm{R}$ and each $1 \leq \lambda \leq k$ (symbol $F_1(\bullet)$ denotes the marginal d.f. of the P's first coordinate, symbol $F_\lambda(\bullet|x_1, \ldots, x_{\lambda-1})$ denotes the d.f. of the P's λ-th coordinate's conditional distribution given the first $\lambda - 1$ coordinates) and let there exist constants $C > 0$ and $0 \leq a \leq 1$ such that

$$
\frac{\partial}{\partial u}F_\lambda^{-1}(u|x_1, x_2, \ldots, x_{\lambda-1}) \leq Cu^{-a} \qquad (1.63)
$$

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

$$
\frac{\partial}{\partial u}F_\lambda^{-1}(u|x_1, x_2, \ldots, x_{\lambda-1}) \leq C(1-u)^{-a} \qquad (1.64)
$$

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

$$
\mathrm{var}(\Gamma_n(g)) \leq \begin{cases} \left(\frac{1}{3}\frac{(m_n-2)^k}{m_n^k} + \frac{1}{(2-a)(1-a)}\frac{m_n^k-(m_n-2)^k}{m_n^k}\right)m_n^{2a-2-k}K^2k^2C^2 & \text{if } a < 1 \\ \left(2 - \frac{5(m_n-2)^k}{3m_n^k}\right)m_n^{-k}K^2k^2C^2 & \text{if } a = 1. \end{cases} \qquad (1.65)
$$

Proof. *(i) Introduction.* Re-index random vectors $\boldsymbol{U}_{1,1,\ldots,1}$, $\boldsymbol{U}_{1,1,\ldots,2}$,…,$\boldsymbol{U}_{m_n,m_n,\ldots,m_n}$ as $\boldsymbol{U}_1$, $\boldsymbol{U}_2$, …, $\boldsymbol{U}_{m_n^k}$ and denote H_i the support of $\mathcal{L}(\boldsymbol{U}_i)$ for each $i = 1, 2, \ldots, m_n^k$. We can rewrite (1.62) as

$$\Gamma_n(g) = m_n^{-k} \sum_{i=1}^{m_n^k} Z_i \tag{1.66}$$

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

(ii) Ad. $0 \le a < 1$. We start with the second moments of $\boldsymbol{t}_P^\lambda(\boldsymbol{U}_i)$. 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 the condition $0 \le q < m_n/2$.

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

$$\begin{aligned}
\mathrm{var}(\boldsymbol{t}_P^\lambda(\boldsymbol{U}_i)) &= \int_{H_i^{\lambda-1}} \int_{q/m_n}^{(q+1)/m_n} [F_\lambda^{-1}(u_\lambda|\boldsymbol{t}_P^1(u_1), \ldots, \boldsymbol{t}_P^{\lambda-1}(u_1, \ldots, u_{\lambda-1})) - E(\boldsymbol{t}_P^\lambda(\boldsymbol{U}_i))]^2 \\
&\qquad\qquad m_n^\lambda du_\lambda du_{\lambda-1} \ldots du_1 \\
&\le \int_{H_i^{\lambda-1}} J(\boldsymbol{t}_P^1(u_1), \ldots, \boldsymbol{t}_P^{\lambda-1}(u_1, \ldots, u_{\lambda-1})) m_n^{\lambda-1} du_{\lambda-1} \ldots du_1
\end{aligned} \tag{1.67}$$

where

$$J(t_1, \ldots, t_{\lambda-1}) = \int_{q/m_n}^{(q+1)/m_n} [F_\lambda^{-1}(u|t_1, \ldots, t_{\lambda-1}) - F_\lambda^{-1}((q+1)/m_n|t_1, \ldots, t_{\lambda-1})]^2 m_n du. \tag{1.68}$$

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

If $q = 0$ then we may use inequality

$$F_\lambda^{-1}(v) - F_\lambda^{-1}(u) = \int_u^v (F_\lambda^{-1})'(x)dx \le \int_u^v Cx^{-a}dx = \frac{C}{1-a}(v^{1-a} - u^{1-a}) \tag{1.69}$$

to estimate

$$\begin{aligned}
J &= \int_0^{1/m_n} [F_\lambda^{-1}(1/m_n) - F_\lambda^{-1}(u)]^2 m_n du \\
&\le \left(\frac{C}{1-a}\right)^2 \int_0^{1/m_n} [(1/m_n)^{1-a} - u^{1-a}]^2 m_n du \\
&= \left(\frac{C}{1-a}\right)^2 \int_0^{1/m_n} [(1/m_n)^{2-2a} - 2(1/m_n)^{1-a}u^{1-a} + u^{2-2a}] m_n du \\
&= \left(\frac{C}{1-a}\right)^2 m_n \left[u(1/m_n)^{2-2a} - 2(1/m_n)^{1-a}\frac{u^{2-a}}{2-a} + \frac{u^{3-2a}}{3-2a}\right]_0^{1/m_n} \\
&= \left(\frac{C}{1-a}\right)^2 m_n \left[1 - \frac{2}{2-a} + \frac{1}{3-2a}\right](1/m_n)^{3-2a} \\
&= \left(\frac{C}{1-a}\right)^2 \left[1 - \frac{2}{2-a} + \frac{1}{3-2a}\right] m_n^{2a-2} \\
&= \left(\frac{C}{1-a}\right)^2 \left[\frac{-a}{2-a} + \frac{1}{3-2a}\right] m_n^{2a-2} \\
&\le \left(\frac{C}{1-a}\right)^2 \left[\frac{-a}{2-a} + \frac{1}{2-a}\right] m_n^{2a-2} \\
&= \left(\frac{C}{1-a}\right)^2 \left(\frac{1-a}{2-a}\right) m_n^{2a-2} = C^2 \frac{1}{(2-a)(1-a)} m_n^{2a-2}
\end{aligned} \tag{1.70}$$

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

$$\mathrm{var}(\boldsymbol{t}_P^\lambda(\boldsymbol{U}_i)) \overset{(1.67),(1.70)}{\leq} C^2 \frac{1}{(2-a)(1-a)} m_n^{2a-2} \int_{H_i^{\lambda-1}} m_n^{\lambda-1} du_{\lambda-1} \dots du_1 = \frac{C^2}{(2-a)(1-a)} m_n^{2a-2} \tag{1.71}$$

for $q = 0$.

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

$$F_\lambda^{-1}(v) - F_\lambda^{-1}(u) = (F_\lambda^{-1})'(b)(v-u) \tag{1.72}$$

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

$$F_\lambda^{-1}(v) - F_\lambda^{-1}(u) \leq C b^{-a}(v-u) \leq C u^{-a}(v-u). \tag{1.73}$$

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

$$\mathrm{var}(\boldsymbol{t}_P^\lambda(\boldsymbol{U}_i)) \overset{(1.67)}{\leq} \frac{C^2}{3} m_n^{2a-2} \tag{1.74}$$

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

Finally, if m_n is odd and $q = m_n/2 - 1/2$ then, similarly to (1.67),

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

Analogically to (1.74), 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 (1.74) holds also for m_n odd and $q = m_n/2 - 1/2$.

It can be easily proved using the symmetry that (1.74) holds for $m_n/2 \le q \le m_n - 2$ and that (1.71) holds for $q = m_n - 1$. Therefore and due to the previous calculations,

$$\operatorname{var}(\boldsymbol{t}_P^\lambda(\boldsymbol{U}_i)) \le \begin{cases} \frac{1}{(2-a)(1-a)} C^2 m_n^{2a-2} & \text{if } q = 0 \text{ or } q = m_n - 1 \\ \frac{1}{3} C^2 m_n^{2a-2} & \text{otherwise.} \end{cases} \tag{1.75}$$

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

$$\operatorname{var}(\boldsymbol{t}_P^\lambda(\boldsymbol{U}_i)) \le \frac{1}{3} C^2 m_n^{2a-2}$$

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

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

if H_i touches the boundary of the unit cube. Now we can use Lemma 1.6. to obtain

$$\begin{aligned} \operatorname{var}(Z_i) &\le K^2 \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \sqrt{\operatorname{var}(\boldsymbol{t}_P^{i_1}(\boldsymbol{U}_i)) \operatorname{var}(\boldsymbol{t}_P^{i_2}(\boldsymbol{U}_i))} \\ &\le \begin{cases} K^2 k^2 \frac{1}{(2-a)(1-a)} C^2 m_n^{2a-2} & \text{if } H_i \text{ touches the boundary of the unit cube} \\ K^2 k^2 \frac{1}{3} C^2 m_n^{2a-2} & \text{otherwise.} \end{cases} \end{aligned} \tag{1.76}$$

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

$$\begin{aligned} \sum_{i=1}^{m_n^k} \operatorname{var} Z_i &\le (m_n - 2)^k K^2 k^2 \frac{1}{3} C^2 m_n^{2a-2} + [m_n^k - (m_n - 2)^k] K^2 k^2 \frac{1}{(2-a)(1-a)} C^2 m_n^{2a-2} \\ &= \left(\frac{1}{3}(m_n - 2)^k + \frac{1}{(2-a)(1-a)}[m_n^k - (m_n - 2)^k]\right) m_n^{2a-2} K^2 k^2 C^2 \end{aligned} \tag{1.77}$$

so that

$$\begin{aligned} \operatorname{var}(\Gamma_n(g)) &= \operatorname{var}\left(m_n^{-k} \sum_{i=1}^{m_n^k} Z_i\right) = m_n^{-2k} \sum_{i=1}^{m_n^k} \operatorname{var}(Z_i) \\ &\overset{(1.77)}{\le} \left(\frac{1}{3}(m_n - 2)^k + \frac{1}{(2-a)(1-a)}[m_n^k - (m_n - 2)^k]\right) m_n^{2a-2-2k} K^2 k^2 C^2 \end{aligned}$$

which proves (1.65) for $a < 1$.

(iii) Ad. $a = 1$. We shall proceed analogously to *(ii)*: Let $1 \le q \le m_n/2 - 1$. It follows from the Mean Value Theorem and from (1.63) that for each $0 < u < v \le 1/2$ there exist $u \le b \le v$ such that

$$F_\lambda^{-1}(v) - F_\lambda^{-1}(u) = (F_\lambda^{-1})'(b)(v - u) \le C b^{-1}(v - u) \le C u^{-1}(v - u). \tag{1.78}$$

Therefore

$$J = \int_{q/m_n}^{(q+1)/m_n} [F_\lambda^{-1}((q+1)/m_n) - F_\lambda^{-1}(u)]^2 m_n du$$

$$
\begin{aligned}
&\overset{(1.78)}{\leq} && \int_{q/m_n}^{(q+1)/m_n} \left[C u^{-1} \left(u - \frac{q}{m_n} \right) \right]^2 m_n du \\
&\leq && \int_{q/m_n}^{(q+1)/m_n} \left[C \left(\frac{1}{m_n} \right)^{-1} \left(u - \frac{q}{m_n} \right) \right]^2 m_n du \\
&= && C^2 \left(\frac{1}{m_n} \right)^{-2} \int_0^{m_n^{-1}} v^2 m_n dv = C^2 m_n^2 m_n \left[v^3/3 \right]_0^{m_n^{-1}} = C^2/3
\end{aligned}
$$

(J is defined by (1.68).)

Let $q = 0$. Since it holds that

$$
F_\lambda^{-1}(v) - F_\lambda^{-1}(u) = \int_u^v (F_\lambda^{-1})'(x) dx \leq \int_u^v C x^{-1} dx = C(\ln v - \ln u) \tag{1.79}
$$

we have

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

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

$$
\operatorname{var}(\boldsymbol{t}_P^\lambda(\boldsymbol{U}_i) \overset{(1.67)}{\leq} \begin{cases} 2C^2 & \text{if } q = 0 \text{ or } q = m_n - 1 \\ \frac{1}{3} C^2 & \text{otherwise.} \end{cases}
$$

and, consequently,

$$
\operatorname{var}(Z_i) \leq \begin{cases} K^2 k^2 2C^2 & \text{if the cube } H_i \text{ touches the boundary of the unit cube} \\ K^2 k^2 \frac{1}{3} C^2 & \text{otherwise} \end{cases} \tag{1.80}
$$

so that

$$
\begin{aligned}
\operatorname{var}(\Gamma_n(g)) &= \operatorname{var}\left(m_n^{-k} \sum_{i=1}^{m_n^k} Z_i \right) = m_n^{-2k} \sum_{i=1}^{m_n^k} \operatorname{var}(Z_i) \\
&\overset{(1.80)}{\leq} m_n^{-2k} \left(\frac{1}{3} (m_n - 2)^k + 2[m_n^k - (m_n - 2)^k] \right) K^2 k^2 C^2 \\
&= m_n^{-2k} \left(2m_n^k - \frac{5}{3} (m_n - 2)^k \right) K^2 k^2 C^2 \\
&= m_n^{-k} \left(2 - \frac{5(m_n - 2)^k}{3 m_n^k} \right) K^2 k^2 C^2.
\end{aligned}
$$

which proves (1.65) for $a = 1$.
□

Corollary. Under the assumptions of the Theorem 1.7.,

$$\Gamma_n(g) = O_P(n^{-\frac{2-2a+k}{2k}}).$$

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

$$\begin{aligned} P(|n^{1/2}\Gamma_n(g)| \geq \epsilon) &\leq \frac{\mathrm{var}(n^{1/2}\Gamma_n(g))}{\epsilon^2} \overset{(1.65)}{\leq} \frac{nm_n^{-k}}{\epsilon^2}\left(2 - \frac{5(m_n-2)^k}{3m_n^k}\right)A \\ &\leq \frac{(m_n+1)^k}{\epsilon^2 m_n^k} 2A \overset{n\to\infty}{\longrightarrow} \epsilon^{-2}2A \end{aligned}$$

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

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

$$\begin{aligned} \mathrm{var}\left(n^{\frac{2-2a+k}{2k}}\Gamma_n(g)\right) &= n^{\frac{2-2a+k}{k}}\,\mathrm{var}\left(\Gamma_n(g)\right) \\ &\overset{(1.65)}{\leq} \frac{\left(\frac{1}{3}\frac{(m_n-2)^k}{m_n^k} + \frac{1}{(2-a)(1-a)}\frac{m_n^k-(m_n-2)^k}{m_n^k}\right)m_n^{2a-2-k}B}{(m_n+1)^{2a-2-k}} \\ &\overset{n\to\infty}{\longrightarrow} B/3 \end{aligned}$$

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

Remark 1.7. (distributions fulfilling the assumptions of Theorem 1.7.) The assumptions are trivially satisfied by uniform distribution (with $a = 0$).

They are also fulfilled by the triangular distribution (defined by density

$$f(x) = I_{(0,\alpha]}(x)\alpha^{-2}x + I_{(\alpha,2\alpha)}(x)\alpha^{-2}(2\alpha - x)$$

- its d.f. is $F(x) = I_{(0,\alpha]}(x)\alpha^{-2}x^2/2 + I_{(\alpha,2\alpha)}(x)(1-(2\alpha-x)^2/(2\alpha^2)$ so that the quantile function is $F^{-1}(u) = I_{(0,1/2]}(u)\alpha\sqrt{2u} + I_{(1/2,1)}(u)\alpha\left(2 - \sqrt{2(1-u)}\right)$ - we can thus put $a = 1/2$).

Exponential distribution trivially fulfils the assumptions of Theorem 1.7. with $a = 1$.

However, the constant $a = 1$ cannot be decreased in the case of exponential distribution: it has quantile function $-\alpha \ln\left(\frac{1-u}{\beta}\right)$ whose derivative $\gamma/(1-u)$ does not satisfy the assumptions for any $a < 1$ (α, β and γ are some constants).

Also the standard normal distribution fulfils them the assumptions of Theorem 1.7. with $a = 1$: The derivative of its quantile function is $(\Phi^{-1})' = 1/\varphi(\Phi^{-1})$ where Φ is the standard normal distribution function and φ is the standard normal density. We are asking whether there exists $C \in \mathrm{R}$ such that $1/\varphi(\Phi^{-1}(u)) \leq Cu^{-1}$. However, since

$$\lim_{u\to 0+}\frac{1/\varphi(\Phi^{-1}(u))}{u^{-1}} = \lim_{u\to 0+} u/\varphi(\Phi^{-1}(u)) = \lim_{x\to-\infty}\Phi(x)/\varphi(x) = \lim_{x\to-\infty}\frac{\varphi(x)}{\varphi'(x)} = \lim_{x\to-\infty}\frac{\varphi(x)}{xf(x)} = 0,$$

a constant C fulfilling (1.63) can be found (similarly with (1.64) and with non-standard normal distribution).

However, $a = 1$ is the best constant in the case of the normal distribution: the derivative of the standard normal distribution's quantile is $1/\varphi(\Phi^{-1}(u))$. The condition (1.63) can thus

be written as $1/\varphi(\Phi^{-1}(u)) \le Cu^{-a}$, $u \le 1/2$, which is equivalent to $1/\varphi(x) \le C\Phi^{-a}(x)$, $x \le 0$, which may be further reformulated as

$$\Phi^a(x)/\varphi(x) \le C, \tag{1.81}$$

$x \le 0$. For (1.81) to be fulfilled

$$\lim_{x\to-\infty} \Phi^a(x)/\varphi(x) \tag{1.82}$$

has to be finite. We show the converse.

Let $0 \le a < 1$. If $a = 0$ then (1.82) is trivially infinite. Let $a > 0$. If the limit (1.82) was finite then it would hold that

$$\lim_{x\to-\infty} \Phi^a(x)/\varphi(x) = \left(\lim_{x\to-\infty} \Phi(x)/\varphi^{1/a}(x)\right)^a < \infty.$$

However, using the l'Hospital rule twice, we get

$$\begin{aligned}
\lim_{x\to-\infty} \frac{\Phi(x)}{\varphi^{1/a}(x)} &\overset{l'H.R.}{=} \lim_{x\to-\infty} a\frac{\varphi(x)}{\varphi'(x)\varphi^{1/a-1}(x)} = \lim_{x\to-\infty} a\frac{\varphi(x)}{-x\varphi(x)\varphi^{1/a-1}(x)} \\
&= \lim_{x\to-\infty} a\frac{\varphi^{1-1/a}(x)}{-x} \overset{l'H.R.}{=} \lim_{x\to-\infty} -a(1-1/a)\varphi^{-1/a}(x)\varphi'(x) \\
&= \lim_{x\to-\infty} a(1/a-1)\varphi^{-1/a}(x)x\varphi(x) = \lim_{x\to-\infty} a(1/a-1)\varphi^{1-1/a}(x)x \\
&= \infty.
\end{aligned}$$

1.2.4 MSE Comparison

The question may arise, which form of approximation is "better": whether discretization, Monte Carlo or Quasi Monte Carlo. It may be difficult to answer the question since the discretization error is a constant while the (Quasi) Monte Carlo one is a random variable. However, we can use the measure applicable to both the deterministic and the random variables - the mean square error (MSE), i.e. to compare (upper bounds of) the MSE of the discretization error

$$\rho_n^D(\Pi_n, g) := (e(\Pi_n, g))^2$$

with the MSE of Monte Carlo error

$$\rho_n^M(g) := E\epsilon_n(g)^2 = E\mathcal{E}_n(g)^2 = \operatorname{var}(\mathcal{E}_n(g)) = \frac{1}{n}\operatorname{var}(g)$$

and with MSE of Quasi Monte Carlo

$$\rho_n^Q(g) := E\Gamma_n(g)^2 = \operatorname{var}(\Gamma_n(g)).$$

Assume that the variance of g is unknown to us but we know that g is l_1-K-Lipschitz.

If $n = m^k$ for some integer m, if marginal distribution functions of P are continuous and if Π_n is defined by the relations (1.8) and (1.9) with $m = n^{1/k}$ then it holds that

$$\rho_n^D(\Pi_n, g) \le \bar{\rho}_n^D, \qquad \bar{\rho}_n^D := \bar{e}_n$$

(see (1.10) for definition of $\bar{e}_n^2$).

If P has finite variance matrix then

$$\rho_n^M(g) \le \bar{\rho}_n^M, \qquad \bar{\rho}_n^M = \overline{V}^2/n$$

($\overline{V}$ is defined by (1.39)).

Finally, if the assumptions of Theorem 1.7. are fulfilled for some $a < 1$ then

$$\rho_n^Q(g) \le \bar{\rho}_n^Q,$$
$$\bar{\rho}_n^Q = \left(\frac{1}{3}\frac{(m-2)^k}{m^k} + \frac{1}{(2-a)(1-a)}\frac{m^k-(m-2)^k}{m^k}\right) m^{2a-2-k}K^2k^2C^2$$

(a and C are defined at Theorem 1.7.).

1.3 Numerical Illustration

Assume, during all subsection 1.3, that the distribution P is jointly uniform with zero mean, i. e. its density is of the form

$$f(\boldsymbol{x}) = \frac{1}{\prod_{i=1}^k (2a_i)} I_M(\boldsymbol{x}), \quad \text{where} \quad M = \prod_{i=1}^k [-a_i, a_i] \tag{1.83}$$

for some $a_1 > 0, a_2 > 0, \ldots, a_k > 0$.

1.3.1 Discretization

Still it could be rather complicated to evaluate $e_n(\Pi)$, defined by (1.2), for general discretization Π. However, it is easy to calculate it for a grid discretization (see Definition 1.1.).

For instance, if $n = m^k$ for some m, the error $e(\bar{\Pi}_n)$, where $\bar{\Pi}_n$ is defined in Theorem 1.1., is, according to Lemma 1.2.,

$$\begin{aligned} e(\bar{\Pi}_n) &= K\sum_{i=1}^k \sum_{j=1}^m \int_{-a_i+2a_ij/m}^{-a_i+2a_i(j-1)/m} \left|\xi - \left(\left[-a_i + \frac{2a_i(j-1)}{m}\right] + \frac{a_i}{m}\right)\right| \frac{1}{2a_i} d\xi \\ &= K\sum_{i=1}^k \left(\frac{m}{2a_i}\int_0^{2a_i/m} \left|u - \frac{a_i}{m}\right| du\right) = K\sum_{i=1}^k \left(\frac{m}{2a_i}\frac{a_i^2}{m^2}\right) = \frac{K}{2m}\sum_{i=1}^k a_i \end{aligned}$$

The bounds $\underline{e}_{n,\delta}$ and $\overline{e}_n$ (introduced by Theorem 1.2. and by Theorem 1.1.) can also be easily calculated: The marginal distribution functions of the jointly uniform distribution with the density (1.83) are

$$H_i(x) = \frac{1}{2} + \frac{1}{2a_i}x$$

for $x \in [-a_i, a_i]$ so that, in case that $n = m^k$ for some $m \in N$,

$$\underline{t}_i(m) = \bar{t}_i(m) = \int_{-a_i}^{-a_i+a_i/m} H_i(x)dx = \frac{a_i}{4m^2}$$

(see Theorem 1.1. for definition of $\underline{t}_i(m)$ and $\bar{t}_i(m)$). Hence, the upper bound introduced by Theorem 1.1. is

$$\begin{aligned}
\overline{e}_n &= K\sum_{i=1}^{k}\left[\frac{a_i}{4m^2}+\frac{1}{2m}\left(\left[a_i-\frac{a_i}{m}\right]-\left[-a_i+\frac{a_i}{m}\right]\right)+\frac{a_i}{4m^2}\right] \\
&= K\sum_{i=1}^{k}\left[\frac{a_i}{4m^2}+\frac{2a_i}{2m}-\frac{2a_i}{2m^2}+\frac{a_i}{4m^2}\right] \\
&= K\sum_{i=1}^{k}\left[\frac{a_i}{m}-\frac{a_i}{2m^2}\right] = K\left(\frac{1}{m}-\frac{1}{2m^2}\right)\sum_{i=1}^{k}a_i \qquad (1.84)
\end{aligned}$$

while the lower bound $\underline{e}_{n,\delta}$ from Theorem 1.2. comes out as

$$\begin{aligned}
\underline{e}_{n,\delta} &= \frac{K}{\prod_{i=1}^{k}(2a_i)}\max\left(\frac{1}{4m}-\frac{1}{4m^{2k^2-k+1}},\frac{1}{2m}-\frac{(m^{2k-1}+1)^{k+1}-1}{(k+1)2m^{2k^2+k}}\right)\sqrt[k]{k!\left[\prod_{i=1}^{k}(2a_i)\right]^{k+1}} \\
&= K\max\left(\frac{1}{4m}-\frac{1}{4m^{2k^2-k+1}},\frac{1}{2m}-\frac{(m^{2k-1}+1)^{k+1}-1}{(k+1)2m^{2k^2+k}}\right)\sqrt[k]{k!\prod_{i=1}^{k}(2a_i)}
\end{aligned}$$

(we have put $\delta = \frac{1}{\prod_{i=1}^{k}(2a_i)}$ in Theorem 1.2.).

The following table shows the values of the discretization error $e(\bar{\Pi}_n)$, its upper bound $\overline{e}_n$ and its lower bound $\underline{e}_{n,\delta}$ for various m and k, $K = 1$ and $a_1 = a_2 = \ldots = a_k = 1/2$:

$n = m^k$	$k=1$	2	3	4	5	6	7	8	9	10
$m=2$	2	4	8	16	32	64	128	256	512	1024
3	3	9	27	81	243	729	2187	6561	19683	59049
4	4	16	64	256	1024	4096	16384	65536	262144	1048576
5	5	25	125	625	3125	15625	78125	390625	1953125	9765625
$e(\bar{\Pi}_n)$										
$m=2$	0,125	0,25	0,375	0,5	0,625	0,75	0,875	1	1,125	1,25
3	0,0833	0,1667	0,25	0,3333	0,4167	0,5	0,5833	0,6667	0,75	0,8333
4	0,0625	0,125	0,1875	0,25	0,3125	0,375	0,4375	0,5	0,5625	0,625
5	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
$\overline{e}_n$										
$m=2$	0,1875	0,375	0,5625	0,75	0,9375	1,125	1,3125	1,5	1,6875	1,875
3	0,1389	0,2778	0,4167	0,5556	0,6944	0,8333	0,9722	1,1111	1,25	1,3889
4	0,1094	0,2188	0,3281	0,4375	0,5469	0,6563	0,7656	0,875	0,9844	1,0938
5	0,09	0,18	0,27	0,36	0,45	0,54	0,63	0,72	0,81	0,9
$\underline{e}_{n,\delta}$										
$m=2$	0,0625	0,186	0,3258	0,4383	0,5415	0,6412	0,7393	0,8365	0,9331	1,0293
3	0,0556	0,1481	0,2259	0,2949	0,3618	0,4277	0,4929	0,5577	0,6221	0,6862
4	0,0469	0,115	0,1701	0,2213	0,2714	0,3208	0,3697	0,4183	0,4666	0,5146
5	0,04	0,0931	0,1362	0,1771	0,2171	0,2566	0,2958	0,3346	0,3732	0,4117

1.3.2 Monte Carlo

The "large deviation" upper bound, introduced by Theorem 1.4. can also be easily (numerically) computed. It holds that

$$
\begin{aligned}
I(\theta) &:= \int_{-a_1}^{a_1}\int_{-a_2}^{a_2}\dots\int_{-a_k}^{a_k}\exp\left\{\theta K\sum_{i=1}^{k}|x_i|\right\}\frac{1}{\prod_{i=1}^{k}[2a_i]}dx_k dx_{k-1}\dots dx_1\\
&= \frac{1}{\prod_{i=1}^{k}[2a_i]}2^k\int_{0}^{a_1}\int_{0}^{a_2}\dots\int_{0}^{a_k}\exp\left\{\theta K\sum_{i=1}^{k}x_i\right\}dx_k dx_{k-1}\dots dx_1\\
&= \frac{1}{\prod_{i=1}^{k}a_i}\int_{0}^{a_1}e^{\theta K x_1}\int_{0}^{a_2}e^{\theta K x_2}\dots\int_{0}^{a_k}e^{\theta K x_k}dx_k dx_{k-1}\dots dx_1\\
&= \frac{1}{\prod_{i=1}^{k}a_i}\prod_{i=1}^{k}\left[\frac{1}{\theta K}\left(e^{\theta K a_i}-1\right)\right]=\prod_{i=1}^{k}\left[\frac{1}{\theta K a_i}\left(e^{\theta K a_i}-1\right)\right]
\end{aligned}
$$

so that

$$
\begin{aligned}
\hat{\epsilon}_n(\alpha) &= K\sum_{i=1}^{k}\int_{-a_i}^{a_i}|x|\frac{1}{2a_i}dx+\inf_{0<\theta<\theta_0}\left\{\frac{1}{\theta}\ln I(\theta)-\frac{1}{\theta n}\ln(\alpha)\right\}\\
&= K\sum_{i=1}^{k}\frac{a_i}{2}+\inf_{0<\theta<\theta_0}\left\{\frac{1}{\theta}\ln\prod_{i=1}^{k}\left[\frac{1}{\theta K a_i}\left(e^{\theta K a_i}-1\right)\right]-\frac{1}{\theta n}\ln(\alpha)\right\}.
\end{aligned}
$$

To compute the "CLT refinement" upper bound $\tilde{\epsilon}_n(\alpha)$ we need to know the second and the third moments of P's marginals: Since

$$
v_{i,j}=\begin{cases}\frac{a_i^2}{3} & \text{for } i=j,\\ 0 & \text{for } i\neq j,\end{cases}
$$

(see Lemma 1.6. for the definition of $v_{i,j}$) it holds that

$$
\overline{V}=K\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}\sqrt{\frac{a_i^2}{3}\frac{a_j^2}{3}}}=\frac{K}{\sqrt{3}}\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}a_i a_j} \tag{1.85}
$$

(see Lemma 1.6. for the definition of $\overline{V}$). As to the third moments,

$$
M_i^3=\frac{a_i^3}{4},\quad e_i=\frac{a_i^2}{2}
$$

(see Lemma 1.7. for the definition of M_i and e_i) so that

$$
\begin{aligned}
\overline{M} &= K\left[\sqrt[3]{\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\sum_{i_3=1}^{k}M_{i_1}M_{i_2}M_{i_3}}+\sum_{i=1}^{k}e_i\right]\\
&= K\left[\sqrt[3]{\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\sum_{i_3=1}^{k}\frac{a_{i_1}}{\sqrt[3]{4}}\frac{a_{i_2}}{\sqrt[3]{4}}\frac{a_{i_3}}{\sqrt[3]{4}}}+\sum_{i=1}^{k}\frac{a_i}{2}\right]\\
&= K\left[\frac{1}{\sqrt[3]{4}}\sqrt[3]{\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\sum_{i_3=1}^{k}a_{i_1}a_{i_2}a_{i_3}}+\frac{1}{2}\sum_{i=1}^{k}a_i\right]
\end{aligned}
$$

(see Lemma 1.7. for the definition of $\overline{M}$). To compute $\tilde{\epsilon}_n(\alpha)$, we solve (numerically) the problem (1.52).

In the end, we evaluate the "asymptotic upper bound" of $\epsilon_n(\alpha)$.

$$\dot{\epsilon}_n(\alpha) := n^{-1/2}\overline{\epsilon}(\alpha) = n^{-1/2}\overline{V}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) = n^{-1/2}\frac{1}{\sqrt{3}}K\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}a_i a_j}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$$

($\overline{\epsilon}(\alpha)$ is defined by (1.57)). Below, a table showing the values of the "standard deviation" bound $\hat{\epsilon}_n(\alpha)$, the "CLT refinement" bound $\tilde{\epsilon}_n(\alpha)$ and the approximate bound $\dot{\epsilon}_n(\alpha)$ of $\epsilon_n(\alpha)$ for various m and k, $K = 1$ and $a_1 = a_2 = \ldots = a_k = 1/2$ can be seen:

$n = m^k$	$k=1$	2	3	4	5	6	7	8	9	10
$m=2$	2	4	8	16	32	64	128	256	512	1024
3	3	9	27	81	243	729	2187	6561	19683	59049
4	4	16	64	256	1024	4096	16384	65536	262144	1048576
5	5	25	125	625	3125	15625	78125	390625	1953125	9765625
$\hat{\epsilon}_n(0.05)$										
$m=2$	0,721	1,24	1,714	2,176	2,639	3,108	3,583	4,062	4,547	5,035
3	0,682	1,164	1,617	2,078	2,551	3,032	3,52	4,012	4,508	5,005
4	0,663	1,124	1,576	2,044	2,525	3,014	3,507	4,004	4,502	5,001
5	0,648	1,099	1,555	2,028	2,514	3,007	3,503	4,002	4,501	5
$\tilde{\epsilon}_n(0.05)$										
$m=2$	34,798	43,938	41,644	35,087	27,725	21,038	15,527	11,231	8	5,632
3	26,61	25,679	18,614	12,004	7,265	4,226	2,393	1,33	0,729	0,395
4	21,984	17,544	10,519	5,615	2,816	1,359	0,639	0,296	0,135	0,061
5	18,962	13,057	6,759	3,118	1,352	0,565	0,231	0,093	0,037	0,015
$\dot{\epsilon}_n(0.05)$										
$m=2$	0,4001	0,5658	0,6001	0,5658	0,5001	0,4243	0,3501	0,2829	0,225	0,1768
3	0,3267	0,3772	0,3267	0,2515	0,1815	0,1257	0,0847	0,0559	0,0363	0,0233
4	0,2829	0,2829	0,2122	0,1414	0,0884	0,053	0,0309	0,0177	0,0099	0,0055
5	0,253	0,2263	0,1518	0,0905	0,0506	0,0272	0,0142	0,0072	0,0036	0,0018

1.3.3 Quasi Monte Carlo

Since we have no upper bound of $\Gamma_n(g)$ (defined by (1.61)) we exploit the fact that under the assumptions of Theorem 1.7.

$$n_\nu^{\frac{2-2a+k}{2k}} \operatorname{var}\Gamma_{n_\nu}{}^{-1/2}\Gamma_{n_\nu}(g) \xrightarrow{\mathcal{D}} N(0,1)$$

as $\nu \to \infty$ for some sequence n_ν,[4] and use an approximate $\Gamma_n(g)$'s α-level quantile bound

$$\bar{\Gamma}_n(\alpha) := n^{-\frac{2-2a+k}{2k}}\sigma_{n_\nu}\Phi^{-1}(1-\alpha/2),$$

where $\operatorname{var}\Gamma_{n_\nu} \le \sigma_{n_\nu}^2$ for each ν, as an illustration of QMC error's behavior.

In case of the uniform distribution with $a_i = 1/2$ it holds that $C = 1$ and $a = 0$ (see Theorem 1.7.) so that we may choose

$$\sigma_n = n^{-\frac{k+1}{k}}Kk\sqrt{\frac{1}{3}(n^{-k}-2)^k + \frac{1}{2}[n-(n^{-k}-2)^k]}$$

[4]We do not include the proof of the fact in the present work since it is very lengthy. However, it is not complicated: First we estimate the third moments of Z_i (see the proof of Theorem 1.7. for the definition of Z_i). Using the estimate, we show that $\lim_{n\to\infty}\sum_{i=1}^{m_n^k} E\left(m_n^{\frac{2-2a+k}{2}-k}Z_i\right)^3 = 0$. Finally, we apply the Central Limit Theorem for the Triangular Schemes to $m_n^{\frac{2-2a+k}{2}}\Gamma_n(g) = m_n^{\frac{2-2a+k}{2}-k}\sum_{i=1}^{m_n^k} Z_i$.

according to (1.65).

The following table shows values for $\bar{\Gamma}_n(\alpha)$ for $\alpha = 0.05$ and $K = 1$.

$n = m^k$	$k=1$	2	3	4	5	6	7	8	9	10
$m=2$	2	4	8	16	32	64	128	256	512	1024
3	3	9	27	81	243	729	2187	6561	19683	59049
4	4	16	64	256	1024	4096	16384	65536	262144	1048576
5	5	25	125	625	3125	15625	78125	390625	1953125	9765625
γ_n										
$m=2$	0.25	0.3536	0.375	0.3536	0.3125	0.2652	0.2188	0.1768	0.1406	0.1105
3	0.1283	0.1542	0.1352	0.1045	0.0755	0.0524	0.0353	0.0233	0.0151	0.0097
4	0.0807	0.0846	0.0649	0.0437	0.0275	0.0165	0.0097	0.0055	0.0031	0.0017
5	0.0566	0.0531	0.0366	0.0221	0.0125	0.0067	0.0035	0.0018	0.0009	0.0005

1.3.4 MSE Comparison

Let us compare the three techniques using the MSE now (see the subsection 1.2.4). Assume again, that $a_1 = a_2 = \ldots = a_k = 1/2$ and that $K = 1$. We shall compare the following values

$$\sqrt{\rho_n^D} = \left(\frac{1}{m} - \frac{1}{2m^2}\right)\frac{k}{2},$$

$$\sqrt{\rho_n^M} = n^{-1/2}\frac{k}{2\sqrt{3}}$$

and

$$\sqrt{\rho_n^Q} = km^{-1-k}\sqrt{\frac{1}{3}(m-2)^k + \frac{1}{2}[m^k - (m-2)^k]} = km^{-1-k}\sqrt{\frac{1}{2}m^k - \frac{1}{6}(m-2)^k}$$

where $m = \lfloor \sqrt[k]{n} \rfloor$.

The following table shows the results.

$n = m^k$	$k=1$	2	3	4	5	6	7	8	9	10
$m=2$	2	4	8	16	32	64	128	256	512	1024
3	3	9	27	81	243	729	2187	6561	19683	59049
4	4	16	64	256	1024	4096	16384	65536	262144	1048576
5	5	25	125	625	3125	15625	78125	390625	1953125	9765625
$\sqrt{\rho_n^D}$										
$m=2$	0.1875	0.375	0.5625	0.75	0.9375	1.125	1.3125	1.5	1.6875	1.875
3	0.1389	0.2778	0.4167	0.5556	0.6944	0.8333	0.9722	1.1111	1.25	1.3889
4	0.1094	0.2188	0.3281	0.4375	0.5469	0.6563	0.7656	0.875	0.9844	1.0938
5	0.09	0.18	0.27	0.36	0.45	0.54	0.63	0.72	0.81	0.9
$\sqrt{\rho_n^M}$										
$m=2$	0.2041	0.2887	0.3062	0.2887	0.2552	0.2165	0.1786	0.1443	0.1148	0.0902
3	0.1667	0.1925	0.1667	0.1283	0.0926	0.0642	0.0432	0.0285	0.0185	0.0119
4	0.1443	0.1443	0.1083	0.0722	0.0451	0.0271	0.0158	0.009	0.0051	0.0028
5	0.1291	0.1155	0.0775	0.0462	0.0258	0.0139	0.0072	0.0037	0.0019	0.0009
$\sqrt{\rho_n^Q}$										
$m=2$	0.25	0.3536	0.375	0.3536	0.3125	0.2652	0.2188	0.1768	0.1406	0.1105
3	0.1283	0.1542	0.1352	0.1045	0.0755	0.0524	0.0353	0.0233	0.0151	0.0097
4	0.0807	0.0846	0.0649	0.0437	0.0275	0.0165	0.0097	0.0055	0.0031	0.0017
5	0.0566	0.0531	0.0366	0.0221	0.0125	0.0067	0.0035	0.0018	0.0009	0.0005

Comment. We see that the Quasi Monte Carlo performs the best (we should keep in mind that the values for QMC are a very conservative upper bound). On the other hand, we should also notify that the QMC has "the best possible" conditions, i.e. $a = 0$ in the case of the uniform distribution.

1.4 Remarks and Open Problems

1.4.1 l_2-K-Lipschitz functions

An assertion, similar to Theorem 1.6., can be derived for l_2-K-Lipschitz functions:

Lemma 1.15. (Bounds of variance of l_2-K-Lipschitz functions) Let the random vector $\boldsymbol{X}$ have the distribution P with a finite variance matrix $(v_{i_1,i_2})_{i_1=1\dots k,i_2=1\dots k}$. Then it holds that

$$\underline{W}^2 \leq \sup_{g \text{ is } l_2\text{-}K\text{-Lipschitz}} \operatorname{var} g(\boldsymbol{X}) \leq \overline{W}^2 \tag{1.86}$$

where

$$\underline{W}^2 = K^2 \max_{i=1,2,\dots,k} v_{i,i} \quad \text{and} \quad \overline{W}^2 = K^2 \sum_{i=1}^{k} v_{i,i} \tag{1.87}$$

Proof. Ad the first "$\leq$": All the functions

$$g_i := K\boldsymbol{X}^i, \quad i = 1, 2, \dots, k,$$

whose variances are $K^2 v_{1,1}, K^2 v_{2,2}, \dots, K^2 v_{k,k}$, are l_2-K-Lipschitz so that

$$\sup_{g \text{ is } l_2\text{-}K\text{-Lipschitz}} \operatorname{var} g(\boldsymbol{X}) \geq K^2 v_{i,i}$$

for each $i = 1, 2, \dots, k$.
Ad the second "$\leq$":

$$\begin{aligned}
\operatorname{var}(g(\boldsymbol{X})) &= E[g(\boldsymbol{X}) - E(g(\boldsymbol{X}))]^2 \leq E[g(\boldsymbol{X}) - g(E\boldsymbol{X})]^2 \\
&\leq E\left(K\sqrt{\sum_{i=1}^{k}(\boldsymbol{X}^i - E\boldsymbol{X}^i)^2}\right)^2 = K^2 E\sum_{i=1}^{k}(\boldsymbol{X}^i - E\boldsymbol{X}^i)^2 \\
&= K^2 \sum_{i=1}^{k} \operatorname{var}(\boldsymbol{X}^i) = K^2 \sum_{i=1}^{k} v_{i,i}.
\end{aligned}$$

(We have used Lipschitz property at the last $\leq$.)
□

Theorem 1.8. Denote Φ^{-1} the inverse of the standard normal distribution's c.d.f. and assume that the distribution P has a finite variance matrix $(v_{i_1,i_2})_{i_1=1\dots k,i_2=1\dots k}$. Let g be l_2-K-Lipschitz function. It holds that

$$\lim_{n\to\infty} \sqrt{n}\, \epsilon_n(\alpha, g) \leq \bar{\epsilon}_2(\alpha) \quad \text{where} \quad \bar{\epsilon}_2(\alpha) := \overline{W}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Proof. The proof is analogical to the proof of Theorem 1.6.
□

1.4.2 Grid Discretizations and Scenario Trees

Definition 1.2. We say that the discrete p.d. $\Pi = \{\boldsymbol{x}_i, p_i\}_{i=1}^n$ where

$$\boldsymbol{x}_i = (\boldsymbol{x}_i^1, \boldsymbol{x}_i^2, \ldots, \boldsymbol{x}_i^m)' \in \Omega_1 \times \Omega_2 \times \ldots \times \Omega_\kappa$$

for some spaces $\Omega_1, \Omega_2, \ldots, \Omega_\kappa$ is a *scenario tree with stages at* $\Omega_1, \Omega_2, \ldots, \Omega_\kappa$ *of the dimension* $m_1, m_2, \ldots, m_\kappa$ if the marginal distribution of elements from Ω_1 has exactly m_1 atoms and the conditional distribution of elements from Ω_ν given values $(\boldsymbol{x}_i^1, \boldsymbol{x}_i^2, \ldots, \boldsymbol{x}_i^{\nu-1})'$ has exactly m_i atoms for each $i = 1, 2, \ldots, n$, $\nu = 2, 3, \ldots, \kappa$.[5]

When we imagine the atoms of the distribution as a graph of its possible realizations, the definition says that each node of stage ν has exactly $m_{\nu+1}$ successors.

It is obvious that each P's grid discretization of dimensions $m_1, m_2, \ldots, m_k$ is a tree of the dimensions $m_1, m_2, \ldots, m_k$. However, it is not true that each scenario tree is a grid discretization (any distribution with atoms $0, 0'$,$0, 1'$, $1, 0'$, $1, 2'$, for instance).

1.4.3 Open Problem - Is the Bound $\overline{V}$ optimal?

Lemma 1.6. states a lower bound $\underline{V}^2$ and an upper bound $\overline{V}^2$ of l_1-K-Lipschitz function's variance. It seems to the author that $\underline{V}^2$ is also the upper bound, however, he could not prove it.

1.4.4 Open Problem - The "true" Convergence Rate of Discretization

Many people say that the Monte Carlo estimation is better than the discretization if the dimension of the integral is large. However, it is a little tricky to compare the convergence rate $O(n^{-1/k+\delta})$ reached by the discretization with $O_P(n^{-1/2})$ of the Monte Carlo (see Theorem 1.3. and Theorem 1.6.). The problem is that while the rate $O_P(n^{-1/2})$ is reached by any MC estimate of $E_P g$, the rate $O(n^{-1/k+\delta})$ does not concern any actual function g but the upper bounds made separately for each n. More exactly, for each Π_n there exists a function g_n such that $e(\Pi_n, g_n) = O(n^{-1/k+\delta})$. The question arises: what is the worst convergence rate of $e(\Pi_n, g)$ if g is a fixed Lipschitz function?

1.4.5 Open Problem - Better Convergence Rate of QMC?

We have shown that the Quasi Monte Carlo estimate of $E_P g$ is $O_P\left(n^{-\frac{2a-2-k}{2k}}\right)$ where a depends on the distribution (see Theorem 1.7.). Could not the rate be made better by unequal partitioning of the unit cube. e.g. smaller cubes at the tails and greater ones in the meddle?

[5] Our definition of scenario tree is different from those from literature, however, it is consistent with them. See [10], II.5.1 for more about scenario trees.

Chapter 2

Approximation of One Stage Stochastic Programming Problem

2.1 Stochastic Programming Problem

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\mathcal{X}$ be an abstract set. Let $g : \mathcal{X} \times \Omega \to \mathrm{R}$ be a function such that $g(x, \bullet)$ is measurable and $E_{P(\omega)}|g(x, \omega)| < \infty$ for each $x \in \mathcal{X}$.

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

$$E_P g(\hat{x}) = \min_{x \in \mathcal{X}} E_P g(x) \tag{2.1}$$

and suppose that it cannot be solved exactly for some reason.

There are many possible ways how to approximate (2.1). We restrict ourselves to the approach consisting in looking for (one of the solutions) $\dot{x}$ such that

$$G(\dot{x}) = \min_{x \in \mathcal{X}} G(x) \tag{2.2}$$

where G is some approximation of $E_P g$ (for an alternative method of SPP approximation - gradient estimation - see [13], [32] or [42]).

The question arises which type of the approximate problem (2.2) should be chosen to compute (2.1). There are more ways how to approximate $E_P g$, we shall deal with the techniques consisting in replacing $F(x)$ by a discrete distribution, as discussed in the previous Chapter - i.e. using the discretization, the Monte Carlo and the Quasi Monte Carlo.

2.1.1 Optimality Gap

We are using the *optimality gap*

$$\eta(\dot{x}) := W(\dot{x}) - W(\hat{x}), \qquad W(x) = E_{P(\omega)} g(x, \omega) = \int_\Omega g(x, \omega) dP(\omega) \tag{2.3}$$

to evaluate the error of approximations. There are more ways of measuring the quality of the approximation (2.2) (e. g. determining the distance between the "exact" and the "approximate" optimal values), the optimality gap, however, seems the most natural because it measures the error in the "units" of the original optimization problem: If costs are minimized, for instance,

then $\eta(\dot{x})$ can be interpreted as the expected price for the approximation.[1] Moreover, $\eta(\dot{x})$ may be combined with $EVPI$[2]: we can define the *expected value of perfect information and exact computation* as

$$EVPIC := EVPI + \eta(\dot{x}) = E_P g(\dot{x}) - E_P \min_{x \in \mathcal{X}} g(x).$$

evaluating the cost of not having the perfect information and not being able to compute (2.1) precisely.

The idea of using the optimality gap as the measure of approximation error is not new and some work concerning the optimality gap was done. The paper [31] suggests a technique of the optimality gap's estimation using the Monte Carlo methods and the article [36] is devoted to the estimation of the optimality gap by means of the Wasserstein distance, for instance.

Before we deal with particular types of approximation, we shall recall a useful and well known result:

Lemma 2.1. For each optimal solution $\dot{x}$ of (2.2) and each optimal solution $\hat{x}$ of (2.1) the following three inequalities hold:

$$\begin{aligned}
(i) \qquad & \eta(\dot{x}) \leq \sup_{x \in \mathcal{X}} [E_P g(x) - G(x)] + [G(\hat{x}) - E_P g(\hat{x})], \\
(ii) \qquad & \eta(\dot{x}) \leq \sup_{x \in \mathcal{X}} [E_P g(x) - G(x)] + \sup_{x \in \mathcal{X}} [G(x) - E_P g(x)], \\
(iii) \qquad & \eta(\dot{x}) \leq 2 \sup_{x \in \mathcal{X}} |G(x) - E_P g(x)|.
\end{aligned}$$

Proof.

$$\begin{aligned}
\eta(\dot{x}) &= E_P g(\dot{x}) - G(\dot{x}) + G(\dot{x}) - E_P g(\hat{x}) \leq E_P g(\dot{x}) - G(\dot{x}) + G(\hat{x}) - E_P g(\hat{x}) \\
&\leq \sup_{x \in \mathcal{X}} [E_P g(x) - G(x)] + G(\hat{x}) - E_P g(\hat{x}) \\
&\leq \sup_{x \in \mathcal{X}} [E_P g(x) - G(x)] + \sup_{x \in \mathcal{X}} [G(x) - E_P g(x)] \\
&\leq 2 \sup_{x \in \mathcal{X}} |G(x) - E_P g(x)|.
\end{aligned}$$

□

2.2 Approximation of SPP

2.2.1 Discretization

The discretization of SPP consists in looking for (one of the solutions) $\dot{x}$ such that

$$E_\Pi g(\dot{x}) = \min_{x \in \mathcal{X}} E_\Pi g(x) \tag{2.4}$$

where Π is a suitable discrete distribution.

The idea of using the discretization to approximate SPP is not new. In [15], the discretization is used to solve the two stage linear SPPs. In [21], the result is generalized to the case

[1] This fact is obvious if $\dot{x}$ is deterministic. Later on, we show that this definition makes sense also in case of random $\dot{x}$.

[2] $EVPI$ - expected value of perfect information - is defined as $EVPI := E_P g(\hat{x}) - E_P \min_{x \in \mathcal{X}} g(x)$, see [2], chp. 4.1. for more about $EVPI$.

of non-linear SPP. In [16], a sequential discretization method is suggested to approximate the two stage SPP. In [11], Theorem 3, a bound of Hausdorff distance between ϵ-optimal solutions sets of the exact and the approximated problem is stated for the case when P is a discrete probability measure with a large number of atoms. In [36], the author suggests an algorithm finding the minimum of the Wasserstein distance of the approximative and the exact measure.

In the present work, even possibly non-convex SPP problems with the probability measure having a possibly unbounded support are handled. As in [36], we use the Wasserstein distance to estimate the approximation error. However, since the computation of the distance could be very complicated and since its minimization can have local minima, we do not try to find an optimal discretization. Instead, we use a sub-optimal discretization for which the distance can be efficiently computed.

We shall assume until the end of the section 2.2.1 that

(a) $\Omega = \mathrm{R}^k$, $\mathcal{A}$ is Borel σ-algebra and all the marginal distributions of P have finite first absolute moments

(b) $\min_{x\in\mathcal{X}} E_P g(x)$ exists,

(c) $\min_{x\in\mathcal{X}} \sum_{i=1}^n g(x_i,\omega_i)p_i$ exists for each $\omega_1,\omega_2,\ldots,\omega_n \in \Omega$ and $p_1,p_2,\ldots,p_n \geq 0$ and

(d) $g(x,\bullet)$ is uniformly l_1-K-Lipschitz on $\mathcal{X}$, i.e. there exists $K > 0$ such that $|g(x,\omega_1) - g(x,\omega_2)| \leq K\|\omega_1 - \omega_2\|_1$ for each $\omega_1,\omega_2 \in \Omega$ and $x \in \mathcal{X}$.

The following Theorem gives an upper bound of the optimality gap caused by the approximation of P by a grid discretization. Moreover, it determines the convergence rate of the optimality gap.

Theorem 2.1. (i) If $\dot{x}$ is an optimal solution of (2.4) and if the approximative distribution Π is a grid discretization of P then

$$\eta(\dot{x}) \leq 2K \sum_{i=1}^{k} d_W(P_i,\Pi_i)$$

where P_i and Π_i, $i = 1,2,\ldots,k$ are the marginal distributions of P, Π respectively (see the section 1.2.1 for the definition of grid discretization).

(ii) If there exists $a > 1$ such that $F_i(x) = O(|x|^{-a})$ as $x \to -\infty$ and $(1 - F_i(x)) = O(|x|^{-a})$ as $x \to \infty$ for each marginal distribution function F_i of P, $i = 1,2,\ldots,k$, then a sequence $\{\dot{\Pi}^n\}_{n=1}^{\infty}$, of at most n-atom discrete probability distributions can be constructed such that

$$\eta(\dot{x}_n) = O(n^{1/k(-1+1/a)})$$

where $\dot{x}_n$ is an (arbitrary) optimal solution of the approximative problem with the distribution $\dot{\Pi}^n$ for each n.

(iii) Let P have marginal densities $f_i(x) = e^{-\phi_i(x)}$, $i = 1,2,\ldots,k$, such that, for some constants $C > 0$, $D > 0$ and for each $|x| \geq D$ it holds that $\phi_i(x) \geq C|x|$. Then there exists a sequence $\{\dot{\Pi}^n\}_{n=1}^{\infty}$ of at most n-atom discrete probability distributions such that, for each $\delta > 0$,

$$\eta(\dot{x}_n) = o(n^{-\frac{1}{k}+\delta}) \quad \text{as } n \to \infty$$

where $\dot{x}_n$ is an (arbitrary) optimal solution of the approximative problem with the distribution $\dot{\Pi}^n$ for each n.

Proof. *Ad (i).* According to Lemma 1.2. it holds that

$$\sup_{h \text{ is } l_1\text{-}K\text{-Lipschitz}} |E_{\Pi(\omega)}h(\omega) - E_{P(\omega)}h(\omega)| = K\sum_{i=1}^{k} d_W(P_i, \Pi_i). \tag{2.5}$$

By the combination of the formula with (iii) of Lemma 2.1. we get

$$\begin{aligned}\eta(\dot{x}) &\overset{\text{Lemma 2.1.,(iii)}}{\leq} 2\sup_{x\in\mathcal{X}} |E_{\Pi(\omega)}g(x,\omega) - E_{P(\omega)}g(x,\omega)| \leq 2\sup_{h \text{ is } l_1\text{-}K\text{-Lipschitz}} |E_{\Pi(\omega)}h(\omega) - E_{P(\omega)}h(\omega)| \\ &\overset{(2.5)}{=} 2K\sum_{i=1}^{k} d_W(P_i, \Pi_i) \end{aligned} \tag{2.6}$$

which proves (i) of the present Theorem.

Ad (ii). Theorem 1.3. guarantees that under the assumptions of (ii) of the present Theorem there exists a sequence of at most n-atom discrete probability distributions $\dot{\Pi}^n$ such that $\sup_{h \text{ is } l_1\text{-}K\text{-Lipschitz}} |E_{\dot{\Pi}^n(\omega)}h(\omega) - E_{P(\omega)}h(\omega)| = O(n^{1/k(-1+1/a)})$. Using it and the first two "$\leq$"'s of (2.6) we get (ii) of the present Theorem.

The same way we prove (iii) of the present Theorem using the Corollary of Theorem 1.3.
□

2.2.2 Monte Carlo

Monte Carlo (MC) approximation of SPP consists in looking for (one of the solutions) $\tilde{x}_n$ such that

$$G_n(\tilde{x}_n) = \min_{x\in\mathcal{X}} G_n(x).$$

where $G_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(x, \zeta_i)$ and $\zeta_1, \zeta_2, \ldots, \zeta_n$ is a sequence of independent random elements with the distribution P.

There was a lot of work done in the area of the MC approximation of the stochastic programming problems. In the early paper [20] the convergence of the optimal value of the approximate problem to one of the exact problem is proved and the rate of the convergence is shown. Similar results appeared independently in [48]. In [23], the empirical estimates based on dependent samples are studied. In [12], the asymptotic normality of the optimal values' distance is proved. This result is further generalized in [41].

Assume until the end of the section 2.2.2 that

(a) $\mathcal{X}$ is compact subset of R^r where $r > 0$ is some integer,

(b) $E(g(\boldsymbol{x}))^2$ is finite at each $\boldsymbol{x} \in \mathcal{X}$ and

(c) there exists a random variable $b : \Omega \to \mathrm{R}$ such that Eb^2 is finite and that $|g(\boldsymbol{x}, \omega) - g(\boldsymbol{y}, \omega)| \leq b(\omega)\|\boldsymbol{x} - \boldsymbol{y}\|_2$ for each $\omega \in \Omega$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$.

It was proven by Shapiro in [41] that under these assumptions and provided that the problem (2.1) has a unique solution, $\sqrt{n}\,[\min_{x\in\mathcal{X}} Eg(\boldsymbol{x}) - \min_{\boldsymbol{x}\in\mathcal{X}} G_n(\boldsymbol{x})]$ converges to the normal distribution. Our result concerns the optimality gap and covers also the situation in which the exact problem has multiple solutions.

Let $\xi \sim P$ be a random vector. If $\tilde{x}_n(\zeta_1, \ldots, \zeta_n)$ is measurable then we define the *optimality gap of Monte Carlo approximation of SPP* as

$$\tilde{\eta}(\tilde{x}_n) = \tilde{\eta}(\tilde{x}_n(\zeta_1, \ldots, \zeta_n)) := E(g(\tilde{x}_n(\zeta_1, \ldots, \zeta_n), \xi)|\zeta_1, \ldots, \zeta_n) - Eg(\hat{x}, \xi) \tag{2.7}$$

($\hat{x}$ is an optimal solution of (2.1)).

Remark 2.1. Thanks to the assumption (c) of the present subsection, there exists a measurable version of $\tilde{x}_n$, according to [12], Proposition 3.2.

Assume until the end of the subsection (2.2.2) that ξ is independent of $\zeta_1, \ldots, \zeta_n$.[3] Under this assumption, P may serve as a conditional distribution of ξ given $\zeta_1, \ldots, \zeta_n$ so that

$$E(g(\tilde{x}_n, \xi)|\zeta_1, \ldots, \zeta_n) = \int g(\tilde{x}_n(\zeta_1, \ldots, \zeta_n), \omega) dP(\omega) = E_{P(\omega)} g(\tilde{x}_n(\zeta_1, \ldots, \zeta_n), \omega)$$

hence

$$\tilde{\eta}(\tilde{x}_n) = \eta(\tilde{x}_n), \tag{2.8}$$

($\eta(\bullet)$ is defined by (2.3)).

Therefore, we may use (i) of Lemma 2.1. to get that

$$\tilde{\eta}(\tilde{x}_n) \quad \leq \quad \bar{\eta}_n \tag{2.9}$$

where

$$\bar{\eta}_n = \max_{x \in \mathcal{X}} [Eg(x, \xi) - G_n(x)] + G_n(\hat{x}) - Eg(\hat{x}, \xi).$$

The following Theorem estimates the asymptotic behavior of the bound $\bar{\eta}_n$.

Theorem 2.2. It holds that

$$n^{1/2} \bar{\eta}_n \xrightarrow{\mathcal{D}} \max_{\boldsymbol{x} \in \mathcal{X}} Y(\boldsymbol{x}) \tag{2.10}$$

where $Y(\bullet)$ is such random element of $C(\mathcal{X})$ that the distribution of each vector $(Y(\boldsymbol{x}_1), Y(\boldsymbol{x}_2), \ldots, Y(\boldsymbol{x}_p))'$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathcal{X}$, $p \in \mathrm{N}$, is normal with $EY(\boldsymbol{x}_1) = EY(\boldsymbol{x}_2) = \ldots = EY(\boldsymbol{x}_p) = 0$ and

$$\mathrm{cov}(Y(\boldsymbol{x}_i), Y(\boldsymbol{x}_j)) = \mathrm{var}(g(\hat{\boldsymbol{x}})) - \mathrm{cov}(g(\boldsymbol{x}_i), g(\hat{\boldsymbol{x}})) - \mathrm{cov}(g(\boldsymbol{x}_j), g(\hat{\boldsymbol{x}})) + \mathrm{cov}(g(\boldsymbol{x}_i), g(\boldsymbol{x}_j)) \tag{2.11}$$

for each $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, p$.

Proof. By rewriting the definition of $\bar{\eta}_n$ we get

$$\bar{\eta}_n = \max_{\boldsymbol{x} \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^{n} h(\boldsymbol{x}, \zeta_i) \quad \text{where} \quad h(\boldsymbol{x}, \zeta_i) := E_P g(\boldsymbol{x}) - g(\boldsymbol{x}, \zeta_i) + g(\hat{\boldsymbol{x}}, \zeta_i) - E_P g(\hat{\boldsymbol{x}}). \tag{2.12}$$

It holds that

$$\begin{aligned} \mathrm{cov}(h(\boldsymbol{x}), h(\boldsymbol{y})) &= \mathrm{cov}(g(\hat{\boldsymbol{x}}) - g(\boldsymbol{x}), g(\hat{\boldsymbol{x}}) - g(\boldsymbol{y})) \\ &= \mathrm{var}(g(\hat{\boldsymbol{x}})) - \mathrm{cov}(g(\boldsymbol{x}), g(\hat{\boldsymbol{x}})) - \mathrm{cov}(g(\boldsymbol{y}), g(\hat{\boldsymbol{x}})) \\ &\quad + \mathrm{cov}(g(\boldsymbol{x}), g(\boldsymbol{y})) \end{aligned} \tag{2.13}$$

for each $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$. Moreover, the following three statements are true:

[3] It is natural to assume that the future realization of the chance is independent on our Monte Carlo sample.

(A) h can be considered as $C(\mathcal{X})$-valued random element. (Proof: Since $h(\bullet,\omega) \in C(\mathcal{X})$ for each $\omega \in \Omega$ due to the assumption (c) and since $h(\boldsymbol{x},\bullet)$ is measurable for each $\boldsymbol{x} \in \mathcal{X}$, h is random element with values from $C(\mathcal{X})$ according to [28], Appendix, Proposition A4.)

(B) $h(\bullet,\omega)$ is Lipschitz with the constant $(b(\omega)+Eb)$. (Proof: It holds that

$$\begin{aligned} |h(\boldsymbol{x},\omega)-h(\boldsymbol{y},\omega)| &= |(g(\boldsymbol{y},\omega)-g(\boldsymbol{x},\omega))-(Eg(\boldsymbol{y})-Eg(\boldsymbol{x}))| \\ &\leq |g(\boldsymbol{y},\omega)-g(\boldsymbol{x},\omega)|+E|g(\boldsymbol{y})-g(\boldsymbol{x})| \\ &\leq b(\omega)\|\boldsymbol{x}-\boldsymbol{y}\|_2+(Eb)\|\boldsymbol{x}-\boldsymbol{y}\|_2=(b(\omega)+Eb)\|\boldsymbol{x}-\boldsymbol{y}\|_2 \quad (2.14) \end{aligned}$$

for each $\boldsymbol{x},\boldsymbol{y} \in \mathcal{X}$.) Moreover, it follows from the assumptions of the present Theorem that $E(b(\omega)+Eb)^2 < \infty$.

(C) $Eh(\boldsymbol{x})^2 < \infty$. (Proof: It holds that $Eh(\boldsymbol{x})^2 \stackrel{Eh(\boldsymbol{x})=0}{=} \mathrm{var}(h(\boldsymbol{x})) = \mathrm{var}(g(\hat{\boldsymbol{x}})-g(\boldsymbol{x})) \leq \mathrm{var}(g(\hat{\boldsymbol{x}})) + 2\sqrt{\mathrm{var}(g(\hat{\boldsymbol{x}}))\,\mathrm{var}(g(\boldsymbol{x}))} + \mathrm{var}(g(\boldsymbol{x})) < \infty$ by the Schwarz inequality and by the assumptions of the present section.)

Since (A) (B) and (C) hold true, it follows Theorem A.3. that there exists random element Y taking values in $C(\mathcal{X})$ with zero mean and covariances (2.13) such that

$$n^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n} h(\bullet)\right) \xrightarrow{\mathcal{D}} Y(\bullet)$$

as random elements in $C(\mathcal{X})$. Therefore and since the function max is continuous on $\mathcal{C}(\mathcal{X})$ we may use Theorem A.2. to get that

$$n^{1/2}\left(\max_{\boldsymbol{x}\in\mathcal{X}} \frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{x})\right) \xrightarrow{\mathcal{D}} \max_{\boldsymbol{x}\in\mathcal{X}} Y(\boldsymbol{x}).$$

Hence, the Theorem is proven.
□

Corollary. Under the assumptions of Theorem 2.2., $\tilde{\eta}_n(\tilde{x}_n) = O_P(n^{-1/2})$.

Proof. The Corollary is a direct consequence of (2.9) and (2.10).
□

2.2.3 Quasi Monte Carlo

Assume until the end of the subsection 2.2.3 that

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

(b) $\Omega = \mathrm{R}^k$ and $\mathcal{A}$ is Borel σ-algebra,

(c) $g(x,\bullet)$ is uniformly l_1-Lipschitz on $\mathcal{X}$ with a constant L (i.e. there exists a constant L such that $|g(\boldsymbol{x},\omega_1)-g(\boldsymbol{x},\omega_2)| \leq L\|\omega_1-\omega_2\|_1$ for each $\omega_1,\omega_2 \in \Omega$ and $\boldsymbol{x} \in \mathcal{X}$),

(d) $g(\bullet,\omega)$ is uniformly l_2-Lipschitz on Ω with a constant K (i.e. there exists a constant K such that $|g(\boldsymbol{x},\omega)-g(\boldsymbol{y},\omega)| \leq K\|\boldsymbol{x}-\boldsymbol{y}\|_2$ for each $\boldsymbol{x},\boldsymbol{y} \in \mathcal{X}$ and $\omega \in \Omega$) and

(e) $|g(\boldsymbol{x},\omega)-E_{P(\omega')}g(\boldsymbol{x},\omega')| < H$ for each $\boldsymbol{x} \in \mathcal{X}$, $\omega \in \Omega$ and for some $H > 0$ independent of $\boldsymbol{x}$ and ω.

The Quasi Monte Carlo (QMC) approximation of (2.1) consists in finding $\check{\boldsymbol{x}}_{\boldsymbol{n}}$ such that

$$C_n(\check{\boldsymbol{x}}_{\boldsymbol{n}}) = \min_{x\in\mathcal{X}} C_n(\boldsymbol{x}), \qquad C_n(\boldsymbol{x}) := \frac{1}{m_n^k}\sum_{i_1=1}^{m_n}\sum_{i_2=1}^{m_n}\dots\sum_{i_k=1}^{m_n} g(\boldsymbol{x}, \boldsymbol{t}_P(\boldsymbol{U}_{n,i_1,i_2,\dots,i_k})) \tag{2.15}$$

where $m_n = \lfloor \sqrt[k]{n} \rfloor$, the mapping $\boldsymbol{t}_P$ is defined at Lemma 1.12. and $\boldsymbol{U}_{n,i_1,i_2,\dots,i_k}$, $i_1 = 1, 2, \dots, m_n$, $i_2 = 1, 2, \dots, m_n$, …, $i_k = 1, 2, \dots, m_n$, are mutually independent random vectors each having uniform distribution on $\left[\frac{i_1-1}{m_n}, \frac{i_1}{m_n}\right] \times \left[\frac{i_2-1}{m_n}, \frac{i_2}{m_n}\right] \times \dots \times \left[\frac{i_k-1}{m_n}, \frac{i_k}{m_n}\right]$.

If we assume that $\check{\boldsymbol{x}}_{\boldsymbol{n}}$ is measurable and that $\boldsymbol{U}_{n,i_1,i_2,\dots,i_k}$ is independent of ξ (we recall that $\xi \sim P$) then we may adapt the definition of the optimality gap from the subsection 2.2.2 with $\boldsymbol{t}_P(\boldsymbol{U}_{n,i_1,i_2,\dots,i_k})$ instead of ζ_i and define the *optimality gap of QMC approximation of SPP* as

$$\check{\eta}(\check{\boldsymbol{x}}_{\boldsymbol{n}}) = E(g(\check{x}_n, \xi)|\boldsymbol{t}_P(\boldsymbol{U}_{n,1,1,\dots,1}), \dots, \boldsymbol{t}_P(\boldsymbol{U}_{n,m_n,m_n,\dots,m_n})) - Eg(\hat{x}, \xi)$$

and show that

$$\check{\eta}(\check{\boldsymbol{x}}_{\boldsymbol{n}}) = \eta(\check{\boldsymbol{x}}_{\boldsymbol{n}})$$

($\eta(\bullet)$ is defined by (2.3)). Moreover, the assumption (c) of the present section guarantees the existence of a measurable version of $\check{\boldsymbol{x}}_{\boldsymbol{n}}$ (see Remark 2.1.).

The following Theorem estimates the convergence rate of the optimality gap.

Theorem 2.3. Let $\frac{\partial}{\partial u}F_\lambda^{-1}(u|z_1, z_2, \dots, z_{\lambda-1})$ exist for each $u \in (0,1)$, each $z_1, z_2, \dots, z_{\lambda-1} \in \mathrm{R}$ and each $1 \le \lambda \le k$ (F_1 denotes the first marginal d. f. of P and, for each $\lambda = 2, 3, \dots, k$, $F_\lambda(\bullet|\boldsymbol{z})$ denotes the d. f. of the conditional distribution of the P's λ-th component given that the first $\lambda - 1$ components equal to $\boldsymbol{z}$) and let there exist constants $C > 0$ and $0 \le a \le 1$ such that

$$\frac{\partial}{\partial u}F_\lambda^{-1}(u|z_1, z_2, \dots, z_{\lambda-1}) \le Cu^{-a}$$

for each $0 < u \le 1/2$, $z_1, z_2, \dots, z_{\lambda-1} \in \mathrm{R}$, $1 \le \lambda \le k$, and

$$\frac{\partial}{\partial u}F_\lambda^{-1}(u|z_1, z_2, \dots, z_{\lambda-1}) \le C(1-u)^{-a}$$

for each $1/2 \le u < 1$, $z_1, z_2, \dots, z_{\lambda-1} \in \mathrm{R}$, $1 \le \lambda \le k$. Let $\delta > 0$. If $a \ge 1 - k/2$ then

$$\check{\eta}(\check{\boldsymbol{x}}_n) = o_P\left(n^{-\frac{2-2a+k}{2k}+\delta}\right)$$

where $\check{\boldsymbol{x}}_n$ is an arbitrary solution of (2.15) for each $n \in \mathrm{N}$.

Before we prove the Theorem, we formulate a useful Lemma.

Lemma 2.2. Let F_λ fulfill the assumptions of the Theorem 2.3. Denote $\Gamma_n(\boldsymbol{x}) := C_n(\boldsymbol{x}) - E_P g(\boldsymbol{x})$. If $a \ge 1 - k/2$ then for each $\eta > 0$ there exists a constant $C_\eta > 0$, independent of $\boldsymbol{x}$ and n, such that

$$P\left[n^{\frac{2-2a+k}{2k}-\delta}|\Gamma_n(\boldsymbol{x})| \ge \eta\right] \le 2\exp\left\{-C_\eta n^\delta\right\}$$

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

Proof. Denote $\kappa := \frac{1-a}{k}$ and $N = \lfloor \sqrt[k]{n} \rfloor^k$. We recall that

$$\Gamma_n(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N} Z_{n,i}, \qquad Z_{n,i} = g(\boldsymbol{x}, \boldsymbol{t}_P(\boldsymbol{U}_{n,i})) - Eg(\boldsymbol{x}, \boldsymbol{t}_P(\boldsymbol{U}_{n,i}))$$

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

$$E Z_{n,i} = 0, \qquad \operatorname{var} Z_{n,i} \le LBN^{-2\kappa}. \tag{2.16}$$

for some constant B independent of i and $\boldsymbol{x}$ (see (1.66), (1.80) and (1.76), we recall that L is the Lipschitz constant introduced by the assumption (c) of the present section). By putting $\bar{X} = \Gamma_n(\boldsymbol{x})$, $\lambda = \eta n^{-\kappa+\delta}$ and $b = H$ into Theorem A.4. we get

$$P\left(n^{\kappa+1/2-\delta}\Gamma_n(\boldsymbol{x}) \ge \eta\right) \le \exp\left\{-\frac{n^{-2\kappa+2\delta}\eta^2}{2\sigma^2}\psi\left(\frac{n^{-\kappa+\delta}\eta H}{\sigma^2\sqrt{n}}\right)\right\}$$

where $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} \operatorname{var} Z_{n,i}$ and ψ is defined at Theorem A.4. Since

$$\sigma^2 \overset{(2.16)}{\le} N^{-2\kappa}V \le (n^{1/k}-1)^{-2k\kappa}V,$$

where $V = LB$, and since $x\psi(x)$ is nondecreasing function according to (b) of Theorem A.4., we have

$$\begin{aligned}
P\left(n^{\kappa+1/2-\delta}\Gamma_n(\boldsymbol{x}) \ge \eta\right) &\le \exp\left\{-\frac{n^{-2\kappa+2\delta}\eta^2}{2[n^{1/k}-1]^{-2k\kappa}V}\psi\left(\frac{n^{-\kappa+\delta}\eta H}{(n-1)^{-2\kappa}V\sqrt{n}}\right)\right\} \\
&\overset{\text{(a) of Th. A.4.}}{\le} \exp\left\{-\frac{n^{-2\kappa+2\delta}\eta^2}{2[n^{1/k}-1]^{-2k\kappa}V}\psi\left(\frac{n^{-\kappa+\delta}\eta H}{n^{-2\kappa}V\sqrt{n}}\right)\right\} \\
&= \exp\left\{-\left(\frac{n^{1/k}-1}{n^{1/k}}\right)^{2k\kappa} n^{2\delta}\frac{\eta^2}{2V}\psi\left(n^{\kappa+\delta-1/2}\frac{\eta H}{V}\right)\right\} \\
&\overset{\text{(c) of Th. A.4.}}{\le} \exp\left\{-\left(1-\frac{1}{n^{1/k}}\right)^{2k\kappa} n^{2\delta}\frac{\eta^2}{2V\left[1+n^{\kappa+\delta-1/2}\frac{\eta H}{3V}\right]}\right\} \\
&= \exp\left\{-\left(1-\frac{1}{n^{1/k}}\right)^{2k\kappa} n^{\delta}\frac{\eta^2}{2V\left[n^{-\delta}+n^{\kappa-1/2}\frac{\eta H}{3V}\right]}\right\} \\
&\le \exp\left\{-\left(1-\frac{1}{n^{1/k}}\right)^{2k\kappa} n^{\delta}\frac{\eta^2}{2V\left[1+\frac{\eta H}{3V}\right]}\right\}
\end{aligned}$$

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

$$\begin{aligned}
P\left(n^{\kappa+1/2-\delta}|\Gamma_n(\boldsymbol{x})| \ge \eta\right) &\le 2\exp\left\{-\left(1-\frac{1}{n^{1/k}}\right)^{2k\kappa} n^{\delta}\frac{\eta^2}{2\left[V+\frac{\eta H}{3}\right]}\right\} \\
&\le 2\exp\left\{-\left(1-\frac{1}{\sqrt[k]{2}}\right)^{2k\kappa} n^{\delta}\frac{\eta^2}{2\left[V+\frac{\eta H}{3}\right]}\right\} \\
&\le 2\exp\left\{-n^{\delta}C_\eta\right\}
\end{aligned}$$

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

Proof of the Theorem 2.3. Let $\eta > 0$. We shall proceed similarly to [27]. Denote $A = \frac{2-2a+k}{2k}$. Since $g(\bullet, \omega)$ is Lipschitz with the constant K, the function $\Gamma_n(\bullet)$ is Lipschitz with the constant $2K$ so that

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

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

Using a procedure, identical to [20][4], proof of Theorem 2, we may show that for each $n \in \mathrm{N}$ there exists $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N(\eta,n)}$ with $N(\eta, n) \leq (D_\eta 2Kn^{A-\delta} + 1)^k$ for some constant D_η, independent of n, such that

$$P\left[\max_{\boldsymbol{x}\in\mathcal{X}} |\gamma_n(\boldsymbol{x})| \geq \eta\right] \leq P\left[\max_{i=1,2,\ldots,N(\eta,n)} |\gamma_n(\boldsymbol{x}_i)| \geq \frac{\eta}{3}\right].$$

Further, we may estimate

$$\begin{aligned} P\left[\max_{i=1,2,\ldots,N(\eta,n)} |\gamma_n(\boldsymbol{x}_i)| \geq \frac{\eta}{3}\right] &\leq \sum_{i=1}^{N(\eta,n)} P\left[|\gamma_n(\boldsymbol{x}_i)| \geq \frac{\eta}{3}\right] \\ &\overset{\text{Lemma 2.2.}}{\leq} \sum_{i=1}^{N(\eta,n)} 2\exp\left\{-n^\delta C_{\frac{\eta}{3}}\right\} = N(\eta,n)2\exp\left\{-n^\delta C_{\frac{\eta}{3}}\right\} \\ &\leq (D_\eta 2Kn^{A-\delta}+1)^k 2\exp\left\{-n^\delta C_{\frac{\eta}{3}}\right\} \overset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

Since

$$0 \leq n^{A-\delta}\check{\eta}(\check{\boldsymbol{x}}_n) \overset{\text{Lemma 2.1.}}{\leq} n^{A-\delta} 2 \max_{\boldsymbol{x}\in\mathcal{X}} |\Gamma_n(\boldsymbol{x})| = 2 \max_{\boldsymbol{x}\in\mathcal{X}} |\gamma_n(\boldsymbol{x})|$$

we have that

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

□

2.3 Illustration

To illustrate the results of the previous section, we consider a simple SPP:

$$\min_{x\in C} E_{P(\omega)}(f(\omega) - x)^2 \tag{2.17}$$

where P is some distribution defined on R^k, $C = [-c, c]$ for some $c \in \mathrm{R}$ and $f : \mathrm{R}^k \to \mathrm{R}$ is a Lipschitz function fulfilling $-c \leq E_P f \leq c$ and $\mathrm{var}_P\, f > 0$. Since $E(X - x)^2$ is minimized at EX for each random variable X, the exact solution of (2.17) is $\hat{x} := E_P f$.

The approximation of (2.17) by the problem with a discrete p.d. $\Pi^n = \{\boldsymbol{z}_i, p_i\}_{i=1}^n$ leads to the problem

$$\min_{x\in C} E_{\Pi^n}(f - x)^2. \tag{2.18}$$

Evidently, the solution of the approximate problem is $\dot{x}_n := E_{\Pi^n} f$.

Since

$$E_P(f - \dot{x}_n)^2 = E_P(f - E_P f)^2 + (E_P f - \dot{x}_n)^2$$

the optimality gap of the solution $\dot{x}_n$ is

$$\begin{aligned} \eta(\dot{x}_n) &= E_P(f - \dot{x}_n)^2 - E_P(f - \hat{x})^2 = E_P(f - E_{\Pi^n} f)^2 - E_P(f - E_P f)^2 = (E_P f - E_{\Pi^n} f)^2 \\ &= (e(\Pi^n, f))^2 \end{aligned}$$

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

(see (1.1) for definition of $e(\Pi^n, f)$). Hence, in case of the approximation of P by a grid discretization having $n = m^k$ atoms,

$$\eta(\dot{x}_n) \leq B \left(\sum_{i=1}^{m} d_W(P_i, \Pi_i^n) \right)^2 = O\left(n^{-\frac{2}{k}(-1+1/a)}\right)$$

for each a fulfilling the assumptions of Theorem 1.3. and for some $B > 0$ (since f is bounded Lipschitz function, the function $(f(\bullet) - x)^2$ is Lipschitz and we may use Theorems 1.1. and 1.3. to estimate the value of $e(\Pi^n, g)$, its rate convergence rate respectively).

In case of the MC estimate based on a sample $\zeta_1, \zeta_2, \ldots, \zeta_n$ from P, the solution of (2.18) is

$$\tilde{x}_n = \overline{f}_n = \frac{1}{n} \sum_{i=1}^{n} f(\zeta_n^i)$$

so that the expected optimality gap is

$$E\tilde{\eta}(\tilde{x}_n) = E(E_P f - \overline{f}_n)^2 = \mathrm{var}(\overline{f}_n) = \frac{1}{n} \sum_{i=1}^{k} \mathrm{var}(f) = O(n^{-1}).$$

Similarly we get an estimate of the convergence rate of the QMC approximation error: The approximate solution is

$$\check{x}_n = \acute{f}_n = \frac{1}{m_n^k} \sum_{i_1=1}^{m_n} \sum_{i_2=1}^{m_n} \cdots \sum_{i_k=1}^{m_n} f(\boldsymbol{t}_P(\boldsymbol{U}_{i_1,i_2,\ldots,i_k}))$$

so that

$$E\check{\eta}(\check{x}_n) = E(E_P f - \acute{f}_n)^2 = \mathrm{var}(\acute{f}_n) = \mathrm{var}(\Gamma_n(f))$$

(Γ_n is defined by (1.61)). Hence, according to Theorem 1.7.,

$$E\eta(\dot{x}_n) = O\left(n^{-\frac{2-2a+k}{k}}\right)$$

where a is some constant fulfilling the assumptions of Theorem 1.7.

To summarize, the test problem outperformed the convergence rates estimates derived in the previous section twice in all the three cases (discretization, MC and QMC). In case of the discretization and of the QMC, the actual convergence rate may be even better than the one shown here (we still have only estimates of their convergence rates). On the other hand, the convergence rate cannot be better in the case of MC.

Chapter 3

Approximation of Multistage SP Problem

3.1 Multistage Stochastic Programming

Let $(\Omega, \mathcal{A}, P)$ be a probability space such that $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_M$ for some subspaces Ω_i, $i = 1, 2, \ldots, M$, $M \in \mathrm{N}$, and that $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \ldots \otimes \mathcal{A}_M$ where $\mathcal{A}_i$ is a σ-algebra defined on Ω_i for each $i = 1, 2, \ldots, M$. Let $S_1, S_2, \ldots, S_M$ be some sets. We define the (M-stage) *multistage stochastic programming problem* (MSSPP) as

$$\min_{x_1 \in \mathcal{X}_1} E_{P_1(\omega_1)} g_1(x_1, \omega_1) \tag{3.1}$$

where P_1 is the marginal distribution (with respect to P) of elements of Ω_1, $\mathcal{X}_1 = S_1$,

$$g_M : S_1 \times S_2 \times \ldots \times S_M \times \Omega_1 \times \Omega_2 \times \ldots \times \Omega_M \longrightarrow \mathrm{R}$$

is a function and

$$g_i(\bar{x}_i, \bar{\omega}_i) = \min_{x_{i+1} \in \mathcal{X}_{i+1}(\bar{x}_i, \bar{\omega}_i)} E_{P_{i+1}(\omega_{i+1}|\bar{\omega}_i)} g_{i+1}(\bar{x}_{i+1}, \bar{\omega}_{i+1}) \tag{3.2}$$

for $1 \leq i < M$, where

$$\mathcal{X}_i : S_1 \times S_2 \times \ldots \times S_{i-1} \times \Omega_1 \times \Omega_2 \times \ldots \times \Omega_{i-1} \longrightarrow 2^{S_i},$$

$i = 2, 3, \ldots, M$, are some point-to-set mappings and $P_i(\bullet|\bar{\omega}_{i-1})$ is a conditional distribution (w.r.t. P) of elements of Ω_i given $\bar{\omega}_{i-1} \in \Omega_1 \times \ldots \times \Omega_{i-1}$ ($\bar{x}_i$ stands for $(x_1, x_2, \ldots, x_i)$, $\bar{\omega}_i$ stands for $(\omega_1, \omega_2, \ldots, \omega_i)$). Assume that $g_i(\bar{x}_i, \bullet)$ is measurable for each $i = 1, 2, \ldots, M$.

An overview of the theory of multistage stochastic programming can be found in [8] or in [10] chp. II.2. For some examples of (economic) applications of MSSPP, see [10] chp. II.4. For a non-economic application, namely hydro-energetic one, see [34].

In most applications, we are unable to evaluate exactly the (conditional) expectations in (3.1) and (3.2) so that we have to make an approximation. More ways how to approximate MSSPP exist. The one which is used most, however, is the replacement of P by a scenario tree.

Let π be a scenario tree with stages at $\Omega_1, \Omega_2, \ldots, \Omega_i$ and with dimensions $n_1, n_2, \ldots, n_M$ (see the section 1.4.2 for the definition of scenario tree). If we denote by

$$\pi_1 = \{\varsigma_{j_1}, p_{j_1}\}_{j_1 = 1, 2, \ldots, n_1} \tag{3.3}$$

its marginal distribution of the elements of Ω_1 and by

$$\pi_i(\bullet|\bar{\varsigma}_{j_1,j_2,\ldots,j_{i-1}}) = \{\varsigma_{j_1,j_2,\ldots,j_i}, p_{j_1,j_2,\ldots,j_i}\}_{j_i=1,2,\ldots,n_i} \tag{3.4}$$

its conditional distribution of the elements of Ω_i given $\bar{\varsigma}_{j_1,j_2,\ldots,j_{i-1}} := (\varsigma_{j_1}, \varsigma_{j_1,j_2}, \ldots, \varsigma_{j_1,j_2,\ldots,j_{i-1}})$ for each $i = 2, 3, \ldots, M$, then the problem originated by the replacement of P by π in (3.1) and (3.2) can be written as

$$\min_{x_1 \in \mathcal{X}_1} E_{\pi_1(\varsigma)} \tilde{g}_1(x_1, \varsigma) = \min_{x_1 \in \mathcal{X}_1} \sum_{j_1=1}^{n_1} p_{j_1} \tilde{g}_1(x_1, \varsigma_{j_1}) \tag{3.5}$$

where

$$\tilde{g}_M(\bar{x}_M, \bar{\varsigma}_{j_1,j_2,\ldots,j_M}) = g_M(\bar{x}_M, \bar{\varsigma}_{j_1,j_2,\ldots,j_M}) \tag{3.6}$$

and

$$\begin{aligned} \tilde{g}_i(\bar{x}_i, \bar{\varsigma}_{j_1,j_2,\ldots,j_i}) &= \min_{x_{i+1} \in \mathcal{X}_{i+1}(\bar{x}_i, \bar{\varsigma}_{j_1,j_2,\ldots,j_i})} E_{\pi_{i+1}(\varsigma_{i+1}|\bar{\varsigma}_{j_1,j_2,\ldots,j_i})} \tilde{g}_{i+1}(\bar{x}_i, x_{i+1}, \bar{\varsigma}_{j_1,j_2,\ldots,j_i}, \varsigma_{i+1}) \\ &= \min_{x_{i+1} \in \mathcal{X}_{i+1}(\bar{x}_i, \bar{\varsigma}_{j_1,j_2,\ldots,j_i})} \sum_{j_{i+1}=1}^{n_{i+1}} p_{j_1,j_2,\ldots,j_{i+1}} \tilde{g}_{i+1}(\bar{x}_i, x_{i+1}, \bar{\varsigma}_{j_1,j_2,\ldots,j_i}, \varsigma_{j_1,j_2,\ldots,j_{i+1}}) \end{aligned} \tag{3.7}$$

for each $1 \leq i < M$.

Assume, until the end of the Chapter 3, that

(a) $E_{P_1(\omega_1)} g_1(x_1, \omega_1)$ exists and is finite for each $x_1 \in \mathcal{X}_1$ and $E_{P_i(\omega_i|\bar{\omega}_{i-1})} g_i(\bar{x}_i, \bar{\omega}_i)$ exists and is finite for each possible $\bar{\omega}_{i-1}$ (i.e. $\bar{\omega}_{i-1} \in \Omega_1 \times \ldots \times \Omega_{i-1}$), each $\bar{x}_{i-1}$ feasible given $\bar{\omega}_{i-1}$ (i.e. such that $x_1 \in \mathcal{X}_1$ $x_2 \in \mathcal{X}_2(x_1, \omega_1)$, …, $x_i \in \mathcal{X}_i(\bar{x}_{i-1}, \bar{\omega}_{i-1})$) and each $i = 1, 2, \ldots, M$,

(b) the exact problem is well defined, i.e. the minimum of (3.1) exists and the minimum of (3.2) exists for each possible $\bar{\omega}_{i-1}$, each $\bar{x}_i$ feasible given $\bar{\omega}_{i-1}$ and each $i = 1, 2, \ldots, M-1$,

(c) the approximate problem is well defined for each discrete distribution π defined on Ω,

(d) Ω is a complete separable metric space,

Remark 3.1. The assumption (d) guarantees the existence of the conditional distributions with respect to P (see [45], VI.1.21).

3.1.1 Problem of Interior Problems' Solution

Before we turn our attention to the definition of the optimality gap originated by the approximate solution of MSSPP, we have to consider the following problem: after the first stage decision is taken and the first stage "chance" realizes itself, one has to solve the second stage interior problem (defined by (3.2) with $i = 2$). However, the approximate model provides its solution only for those cases when the "chance" realized itself as one of the atoms of the approximating distribution π_1. If the "chance" fell outside the set of the atoms of π_1, then the solution of the approximate model may be useless since it may not be feasible. The situation at stages $3, 4, \ldots, M$ is similar.[1]

[1]The problem, described here, is reported in [6], for instance.

Hence, for a successful application of the MSSPP, it does not suffice to solve (3.5) but, in addition, one has to be able to find

$$\min_{x_{i+1}\in\mathcal{X}_i(\bar{x}_i,\bar{\omega}_i)} E_{P_{i+1}(\omega_{i+1}|\bar{\omega}_i)} g_{i+1}(\bar{x}_i, x_{i+1}, \bar{\omega}_i, \omega_{i+1}), \tag{3.8}$$

where $\bar{x}_i := (x_1, x_2, \ldots, x_i)$ are decisions taken in the first i stages, for each $\bar{\omega}_i \in \Omega_1 \times \Omega_2 \times \ldots \times \Omega_i$ and for each $i = 1, 2, \ldots, M-1$.

There could be more ways how to solve (3.8), e.g. taking the "nearest" feasible solution of (3.5). We suppose, however, that the most natural approach to the interior problems is to generate a new scenario tree for each problem (3.8) (see the sections 3.2.1, 3.2.2 and 3.2.3 for details).

3.2 Approximation of MSSPP

We will discuss three ways how to generate the approximating distribution in (3.5) - the discretizazion of P, sampling from P (Monte Carlo) and creating an approximating distribution using the Quasi Monte Carlo.

3.2.1 Discretization

Similarly to the previous text, the word "discretization" means the replacement of the "exact" distribution by a discrete one.

There exist many results concerning the accuracy of the MSSPP's discretization (see [11] or [25], for instance). All these works use the distance of the optimal values as a measure of the accuracy. We provide some results concerning the optimality gap.

First let us discuss the problem mentioned in the section 3.1.1. As we have already said, we suggest to generate a new scenario tree to solve each problem (3.8). i.e. to construct mappings

$$\Pi_i : \Omega_1 \times \Omega_2 \times \ldots \times \Omega_{i-1} \to \mathcal{S}_{\Omega_i, n_i}, \tag{3.9}$$

$i = 2, 3, \ldots, M$, where $\mathcal{S}_{\Omega_i,n_i}$ is the space of n_i-atom discrete distributions defined on Ω_i, and to approximate each problem (3.8) by

$$\min_{x_{i+1}\in\mathcal{X}_i(\bar{x}_i,\bar{\omega}_i)} E_{(\Pi_i(\bar{\omega}_i))(\varsigma_{i+1})} \dot{g}_{i+1}(\bar{x}_i, x_{i+1}, \bar{\omega}_i, \varsigma_{i+1}) \tag{3.10}$$

where $\dot{g}_M = g_M$ and

$$\dot{g}_\nu(\bar{x}_\nu, \bar{\omega}_\nu) \;=\; \min_{x_{\nu+1}\in\mathcal{X}_{\nu+1}(\bar{x}_\nu,\bar{\omega}_\nu)} E_{(\Pi_{\nu+1}(\bar{\omega}_\nu))(\varsigma_{\nu+1})} \dot{g}_{\nu+1}(\bar{x}_\nu, x_{\nu+1}, \bar{\omega}_\nu, \varsigma_{\nu+1})$$

for each $\nu = i+1, i+2, \ldots, M-1$.

In the sequel, the ordered set $(\Pi_1, \Pi_2, \ldots, \Pi_M)$, where Π_1 is some n_1-atom discrete probability distribution defined on Ω_1 and where Π_i are some mappings of the type (3.9), will be called *approximation scheme of dimensions* $n_1, n_2, \ldots, n_M$ (we shall write $\Pi_i(\varsigma_i|\bar{\varsigma})$ instead of $(\Pi_i(\bar{\varsigma}))(\varsigma_i)$ for the sake of readability). Under the *number of atoms of the approximation scheme* Π we shall understand the quantity $\prod_{i=1}^{k} n_i$.

Now let us proceed with the definition of the optimality gap.

Denote $\hat{x}_1$ (an arbitrary) solution of (3.1), $\dot{x}_1$ (an arbitrary) solution of (3.5), $\hat{x}_i(\bar{x}_{i-1}) = \hat{x}_i(\bar{x}_{i-1}, \bar{\omega}_{i-1})$ (an arbitrary) solution of (3.8), and $\dot{x}_i(\bar{x}_{i-1}) = \dot{x}_i(\bar{x}_{i-1}, \bar{\omega}_{i-1})$ (an arbitrary) solution of (3.10). Let $\Pi = (\Pi_1, \Pi_2, \ldots, \Pi_M)$ be an approximating scheme. We define the *optimality gap of MSSPP approximation by scheme* Π as

$$\begin{aligned} \eta_\Pi &:= E_{P(\bar{\omega}_M)} g_M(\dot{x}_1, \dot{x}_2(\dot{x}_1, \omega_1), \ldots, \dot{x}_M(\dot{x}_1, \ldots, \dot{x}_{M-1}, \bar{\omega}_{M-1}), \bar{\omega}_M) \\ &\quad - E_{P(\bar{\omega}_M)} g_M(\hat{x}_1, \hat{x}_2(\hat{x}_1, \omega_1), \ldots, \hat{x}_M(\hat{x}_1, \ldots, \hat{x}_{M-1}, \bar{\omega}_{M-1}), \bar{\omega}_M). \end{aligned} \tag{3.11}$$

An interpretation of the optimality gap of MSSPP is analogous to the interpretation of the one-stage programming optimality gap: Suppose that the "chance" (gradually) realized itself as ω_1,ω_2,..., ω_{M-1}. If we were able to solve the MSSPP and its interior problems exactly, then the sequence of our decisions would have been

$$\hat{x}_1, \hat{x}_2(\hat{x}_1, \omega_1), \ldots, \hat{x}_M(\hat{x}_1, \ldots, \hat{x}_{M-1}, \bar{\omega}_{M-1}).$$

However, since we are using the approximative models, our decisions are

$$\dot{x}_1, \dot{x}_2(\dot{x}_1, \omega_1), \ldots, \dot{x}_M(\dot{x}_1, \ldots, \dot{x}_{M-1}, \bar{\omega}_{M-1}).$$

Hence, η_Π measures the expected loss caused by the approximation.[2]

The following Lemma shows a way how the optimality gap may be decomposed.

Lemma 3.1. Denote

$$\eta_1(\omega_1) = g_1(\dot{x}_1, \omega_1) - g_1(\hat{x}_1, \omega_1) \tag{3.12}$$

and

$$\begin{aligned} \eta_i(\bar{\omega}_i) &:= g_i(\dot{x}_1, \dot{x}_2(\dot{x}_1), \ldots, \dot{x}_{i-1}(\dot{x}_1, \ldots, \dot{x}_{i-2}), \dot{x}_i(\dot{x}_1, \ldots, \dot{x}_{i-1}), \bar{\omega}_i) \\ &\quad - g_i(\dot{x}_1, \dot{x}_2(\dot{x}_1), \ldots, , \dot{x}_{i-1}(\dot{x}_1, \ldots, \dot{x}_{i-2}), \hat{x}_i(\dot{x}_1, \ldots, \dot{x}_{i-1}), \bar{\omega}_i) \end{aligned} \tag{3.13}$$

for $2 \leq i \leq M$. Then

$$\eta_\Pi = \sum_{i=1}^{M} E_P \eta_i. \tag{3.14}$$

Remark 3.2. The quantity η_i may be interpreted as the optimality gap of the i-th interior problem provided that we are able to solve the problems at stages $i+1, i+2, \ldots, M$ exactly.

Proof of Lemma 3.1. We proceed by induction according to M. If $M = 1$ then the assertion coincides with the definition of η_Π. If $M > 1$ and (3.14) holds for $M - 1$ then

$$\begin{aligned} \eta_\Pi &\overset{(3.11),(3.2)}{=} E_P g_M(\dot{x}_1, \dot{x}_2(\dot{x}_1), \ldots, \dot{x}_M(\dot{x}_1, \ldots, \dot{x}_{M-1})) - E_{P_1} g_1(\hat{x}_1) \\ &= [E_P g_M(\dot{x}_1, \dot{x}_2(\dot{x}_1), \ldots, \dot{x}_M(\dot{x}_1, \ldots, \dot{x}_{M-1})) - E_{P_1} g_1(\dot{x}_1)] \\ &\quad + E_{P_1}[g_1(\dot{x}_1) - g_1(\hat{x}_1)] \\ &\overset{(3.2),(3.12)}{=} E_{P_1(\omega_1)} A(\dot{x}_1, \omega_1) + E_{P_1(\omega_1)} \eta_1(\omega_1) \end{aligned} \tag{3.15}$$

where

$$\begin{aligned} A(\dot{x}_1, \omega_1) &= E_{P(\omega_2, \ldots, \omega_M | \omega_1)} g_M(\dot{x}_1, \dot{x}_2(\dot{x}_1), \ldots, \dot{x}_M(\dot{x}_1, \ldots, \dot{x}_{M-1}), \bar{\omega}_M) \\ &\quad - E_{P(\omega_2, \ldots, \omega_M | \omega_1)} g_M(\dot{x}_1, \hat{x}_2(\dot{x}_1), \ldots, \hat{x}_M(\dot{x}_1, \hat{x}_2, \ldots, \hat{x}_{M-1}), \bar{\omega}_M), \end{aligned}$$

[2] Obviously, the optimality gap depends on the choice of the approximate solution. However, our further results hold independently of the choice.

$P(\bullet|\omega_1)$ denotes the conditional distribution of the elements from $\Omega_2 \times \ldots \times \Omega_M$ given ω_1. However, when we fix ω_1 we see that $A(\dot{x}_1, \omega_1)$ is nothing else but the optimality gap of the $M-1$ stage problem

$$\min_{x_2 \in \mathcal{X}_2(\dot{x}_1, \omega_1)} E_{P_2(\omega_2|\omega_1)} g_2(\dot{x}_1, x_2, \omega_1, \omega_2).$$

Therefore, according to the induction assumption,

$$A(\dot{x}_1, \omega_1) = \sum_{i=1}^{M-1} E_{P(\omega_2,\ldots,\omega_M|\omega_1)} \eta'_{i,\omega_1}(\omega_2, \ldots, \omega_{i+1}) \tag{3.16}$$

where

$$\begin{aligned} \eta'_{i,\omega_1}(\omega_2, \ldots, \omega_{i+1}) &= g_{i+1}(\dot{x}_1, \dot{x}_2(\dot{x}_1), \ldots, \dot{x}_{i+1}(\dot{x}_1, \dot{x}_2(\dot{x}_1), \ldots), \bar{\omega}_{i+1}) \\ &\quad - g_{i+1}(\dot{x}_1, \dot{x}_2(\dot{x}_1), \ldots, \hat{x}_{i+1}(\dot{x}_1, \dot{x}_2(\dot{x}_1), \ldots), \bar{\omega}_{i+1}) \end{aligned}$$

for $1 \leq i \leq M-1$. Since $\eta'_{i,\omega_1}(\omega_2, \ldots, \omega_{i+1}) = \eta_{i+1}(\omega_1, \omega_2, \ldots, \omega_{i+1})$ for each $1 \leq i \leq M-1$ it suffices to impose (3.16) into (3.15) to prove (3.14).
□

It is clear that the evaluation of η_Π requires the ability to compute the expectation with respect to P. In case we are unable to compute such expectations, η_Π has to be estimated. The following Theorem provides one of possible upper bounds of η_Π.

Theorem 3.1. Let $\Pi = (\Pi_1, \Pi_2, \ldots, \Pi_M)$ be an approximation scheme and let there exist constants $d_1, d_2, \ldots, d_M$ such that

$$\left| E_{P_1(\omega_1)} g_1(x_1, \omega_1) - E_{\Pi_1(\varsigma_1)} g_1(x_1, \varsigma_1) \right| \leq d_1$$

for each $x_1 \in \mathcal{X}_1$ and

$$\left| E_{P_i(\omega_i|\bar{\omega}_{i-1})} g_i(\bar{x}_i, \bar{\omega}_{i-1}, \omega_i) - E_{\Pi_i(\varsigma_i|\bar{\omega}_{i-1})} g_i(\bar{x}_i, \bar{\omega}_{i-1}, \varsigma_i) \right| \leq d_i$$

for each $i = 2, 3, \ldots, M$, $\bar{\omega}_{i-1} \in \Omega_1 \times \Omega_2 \times \ldots \times \Omega_{i-1}$, $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2(x_1, \omega_1)$, …, $x_i \in \mathcal{X}_i(\bar{x}_{i-1}, \omega_{i-1})$. Then

$$\eta_\Pi \leq 2 \sum_{i=1}^{M} i d_i.$$

Proof. First we show that

$$\sup_{x_1 \in \mathcal{X}_1} |E_{P_1} g_1(x_1) - E_{\Pi_1} \tilde{g}_1(x_1)| \leq \sum_{\nu=1}^{M} d_\nu \tag{3.17}$$

by induction according to M: If $M = 1$ then $\tilde{g}_1$ coincides with g_1 so that (3.17) follows from the definition of d_1. If $M > 1$ and (3.17) holds for $M-1$ then

$\sup_{x_1 \in \mathcal{X}_1} |E_{P_1} g_1(x_1) - E_{\Pi_1} \tilde{g}_1(x_1)|$

$$\begin{aligned} &\leq \sup_{x_1 \in \mathcal{X}_1} |E_{P_1} g_1(x_1) - E_{\Pi_1} g_1(x_1)| + \sup_{x_1 \in \mathcal{X}_1} |E_{\Pi_1} g_1(x_1) - E_{\Pi_1} \tilde{g}_1(x_1)| \\ &\leq d_1 + E_{\Pi_1} \sup_{x_1 \in \mathcal{X}_1} |g_1(x_1) - \tilde{g}_1(x_1)| \end{aligned}$$

$$\begin{aligned}
&\overset{(3.2)}{=} d_1 + E_{\Pi_1(\varsigma_1)} \sup_{x_1\in\mathcal{X}_1} \Big| \min_{x_2\in\mathcal{X}_2(x_1,\varsigma_1)} E_{P_2(\omega_2|\varsigma_1)} g_2(x_1,x_2,\varsigma_1,\omega_2) \\
&\qquad\qquad - \min_{x_2\in\mathcal{X}_2(x_1,\varsigma_1)} E_{\Pi_2(\varsigma_2|\varsigma_1)} \tilde{g}_2(x_1,x_2,\varsigma_1,\varsigma_2) \Big| \\
&\leq d_1 + E_{\Pi_1(\varsigma_1)} \sup_{x_1\in\mathcal{X}_1} \left(\sup_{x_2\in\mathcal{X}_2(x_1,\varsigma_1)} \left[E_{P_2(\omega_2|\varsigma_1)} g_2(x_1,x_2,\varsigma_1,\omega_2) - E_{\Pi_2(\varsigma_2|\varsigma_1)} \tilde{g}_2(x_1,x_2,\varsigma_1,\varsigma_2) \right] \right) \\
&\leq d_1 + \sum_{\nu=2}^{M} d_\nu \qquad (3.18)
\end{aligned}$$

according to the induction assumption being applied to the first stage interior problem

$$\min_{x_2\in\mathcal{X}_2(x_1,\omega_1)} E_{P_2(\omega_2|\omega_1)} g_2(x_1,x_2,\omega_1,\omega_2).$$

Hence, (3.17) is proven[3] and we may use a well known procedure to get

$$\begin{aligned}
\eta_1 &= E_{P_1} g_1(\dot{x}_1) - E_{P_1} g_1(\hat{x}_1) \\
&= E_{P_1} g_1(\dot{x}_1) - E_{\Pi_1} \tilde{g}_1(\dot{x}_1) + E_{\Pi_1} \tilde{g}_1(\dot{x}_1) - E_{P_1} g(\hat{x}_1) \\
&\leq E_{P_1} g_1(\dot{x}_1) - E_{\Pi_1} \tilde{g}_1(\dot{x}_1) + E_{\Pi_1} \tilde{g}_1(\hat{x}_1) - E_{P_1} g(\hat{x}_1) \\
&\leq 2 \sup_{x_1\in\mathcal{X}_1} |E_{P_1} g_1(x_1) - E_{\Pi_1} \tilde{g}_1(x_1)| \overset{(3.17)}{\leq} 2 \sum_{\nu=1}^{M} d_\nu \qquad (3.19)
\end{aligned}$$

(η_1 is defined by (3.12)).

Let $2 \leq i \leq M$ and $\bar{\omega}_{i-1} \in \Omega_1 \times \ldots \times \Omega_{i-1}$. When we apply (3.19) to the interior problem

$$\min_{x_i\in\mathcal{X}_i(\dot{x}_1,\dot{x}_2(\dot{x}_1),\ldots,\dot{x}_{i-1}(\dot{x}_1,\ldots),\bar{\omega}_{i-1})} E_{P_i(\omega_i|\bar{\omega}_{i-1})} g_i(\dot{x}_1,\dot{x}_2(\dot{x}_1),\ldots,\dot{x}_{i-1}(\dot{x}_1,\ldots),x_i,\bar{\omega}_{i-1},\omega_i)$$

we get that

$$\eta_i(\bar{\omega}_{i-1}) \leq 2 \sum_{\nu=i}^{M} d_\nu \qquad (3.20)$$

so that

$$\eta_\Pi \overset{\text{Lemma 3.1.}}{=} \sum_{i=1}^{M} E_P \eta_i \overset{(3.19),(3.20)}{\leq} 2 \sum_{i=1}^{M} \sum_{\nu=i}^{M} d_\nu = 2 \sum_{i=1}^{M} i d_\nu.$$

□

To assure the assumptions of the Theorem 3.1., we shall suppose, until the end of the subsection 3.2.1, that

(1.a) $\Omega_i = \mathrm{R}^{k_i}$ for some k_i, $i = 1, 2, \ldots, M$ and $\mathcal{A}_i$ is the Borel sigma algebra for each $i = 1, 2, \ldots, M$.

(1.b) each function $g_i(\bar{x}_i, \bar{\omega}_{i-1}, \bullet)$, $1 \leq i \leq M$, is l_1-Lipschitz (i.e. there exists a constant K_i such that $|g_i(\bar{x}_i, \omega_1, \ldots, \omega_{i-1}, \omega) - g_i(\bar{x}_i, \omega_1, \ldots, \omega_{i-1}, \omega')| \leq K_i \|\omega - \omega'\|_1$, for each possible $\bar{\omega}_{i-1}$, each $\bar{x}_i$ feasible given $\bar{\omega}_{i-1}$ and each $\omega, \omega' \in \Omega_i$),

[3] To derive (3.18) we have exploited idea similar to Kaňková in [25], pp. 137-138 as well as the well known fact that $\min f(x) - \min g(x) = f(\arg\min f(x)) - g(\arg\min g(x)) \leq f(\arg\min g(x)) - g(\arg\min g(x)) \leq \sup_x [f(x) - g(x)]$.

Remark 3.3. Sufficient conditions for the assumption (1.b) are stated in [24], for instance.

There are many ways how to choose a discrete distribution approximating the "exact" distribution at each stage (see [10] chp. II. 5). We shall use grid discretizations (see the section 1.2.1, Definition 1.1.). The following Theorem provides an upper bound of the optimality gap.

Theorem 3.2. Let Π be an approximating scheme of dimensions $n_1, n_2, \ldots, n_M$ such that $\Pi_i(\bullet|\bar{\omega}_{i-1})$ is a grid discretization of $P_i(\bullet|\bar{\omega}_{i-1})$ for each $\bar{\omega}_{i-1} \in \mathrm{R}^{k_1} \times \mathrm{R}^{k_2} \times \ldots \times \mathrm{R}^{k_{i-1}}$, $i = 1, 2, \ldots, M$. Moreover, let there exist constants $d_{i,\nu}$, $i = 1, 2, \ldots, M$, $\nu = 1, 2, \ldots, k_i$ such that

$$d_W(P_{i,\nu}(\bullet|\bar{\omega}_{i-1}), \Pi_{i,\nu}(\bullet|\bar{\omega}_{i-1})) \leq d_{i,\nu} \tag{3.21}$$

for each $i = 1, 2, \ldots, M$ and $\bar{\omega}_{i-1} \in \mathrm{R}^{k_1} \times \mathrm{R}^{k_2} \times \ldots \times \mathrm{R}^{k_{i-1}}$ ($P_{i,\nu}(\bullet|\bar{\omega}_{i-1})$ denotes the ν-th marginal distribution of $P_i(\bullet|\bar{\omega}_{i-1})$, $\Pi_{i,\nu}(\bullet|\bar{\omega}_{i-1})$ denotes the ν-th marginal distribution of p.d. $\Pi_i(\bullet|\bar{\omega}_{i-1})$, we put $P_1(\bullet|\bar{\omega}_0) := P_1(\bullet)$ and $\Pi_1(\bullet|\bar{\omega}_0) := \Pi_1(\bullet)$). Then

$$\eta_\Pi \leq 2 \sum_{i=1}^{M} \left(i \cdot K_i \cdot \sum_{\nu=1}^{k_i} d_{i,\nu} \right)$$

for each choice of the approximating problem's solutions.

Proof. Since

$$\left| E_{P_i(\omega_i|\bar{\omega}_{i-1})} g_i(\bar{x}_i, \bar{\omega}_{i-1}, \omega_i) - E_{\Pi_i(\varsigma_i|\bar{\omega}_{i-1})} g_i(\bar{x}_i, \bar{\omega}_{i-1}, \varsigma_i) \right|$$

$$\overset{\text{Lemma 1.2.}}{\leq} K_i \sum_{\nu=1}^{k_i} d_W(P_{i,\nu}(\bullet|\bar{\omega}_{i-1}), \Pi_{i,\nu}(\bullet|\bar{\omega}_{i-1})) \overset{(3.21)}{\leq} K_i \sum_{\nu=1}^{k_i} d_{i,\nu}$$

for each $1 \leq i \leq M$, each $\bar{\omega}_{i-1} \in \Omega_1 \times \Omega_2 \times \ldots \times \Omega_{i-1}$ and each $\bar{x}_i$ feasible given $\bar{\omega}_{i-1}$, the assertion of the present Theorem is a direct consequence of Theorem 3.1.

□

The next Theorem evaluates the convergence rate of the optimality gap. We examine the convergence rate with respect to the number of atoms of the approximating scheme. The reason is that the size (number of constrains) of the approximating problem depends just on the number of the atoms of the approximating distribution.

Theorem 3.3. Let $P_i(\bullet|\bar{\omega}_{i-1})$ have finite first absolute moment for each $1 \leq i \leq M$ and each $\bar{\omega}_{i-1} \in \Omega_1 \times \ldots \times \Omega_{i-1}$. Assume that there exists a constant D and a function $h(x) = O(x^{-a})$ as $x \to \infty$ such that

$$\max(F_{i,\nu}^{\bar{\omega}_{i-1}}(-x), 1 - F_{i,\nu}^{\bar{\omega}_{i-1}}(x)) \leq h(x) \tag{3.22}$$

for each $i = 1, 2, \ldots, M$, $\nu = 1, 2, \ldots, k_i$, $\bar{\omega}_{i-1} \in \mathrm{R}^{k_1} \times \mathrm{R}^{k_2} \times \ldots \times \mathrm{R}^{k_{i-1}}$ and $x \geq D$ (the symbol $F_{i,\nu}^{\bar{\omega}_{i-1}}$ denotes the distribution function of $P_{i,\nu}(\bullet|\bar{\omega}_{i-1})$). Then there exists a sequence $\{\Pi^n\}_{n=1}^{\infty}$ of at most n-atom approximation schemes such that

$$\eta_{\Pi_n} = O\left(n^{\frac{1}{k}(-1+1/a)}\right), \qquad k = \sum_{i=1}^{M} k_i$$

independently of the choice of the approximating problem's solutions (see Theorem 3.1. for the notation).

Proof. Denote $\mu_n = \lfloor n^{1/k} \rfloor$ and assume that $\mu_n^{1/a} \geq D$. When we put $C = \mu_n^{1/a}$ in Lemma 1.5. we get that for each $1 \leq i \leq M$, $1 \leq \nu \leq k_i$ and $\bar{\omega}_{i-1} \in \Omega_1 \times \ldots \times \Omega_{i-1}$ there exists a grid discretization $\Pi^{\mu_n}_{i,\nu,\bar{\omega}_{i-1}}$ of $P_{i,\nu}(\bullet|\bar{\omega}_{i-1})$ with the dimension no greater than μ_n such that

$$d_W(\Pi^{\mu_n}_{i,\nu,\bar{\omega}_{i-1}}, P_{i,\nu}(\bullet|\bar{\omega}_{i-1})) \overset{\text{Lemma 1.5.}}{\leq} \int_{-\infty}^{-\mu_n^{1/a}} F_{i,\nu}^{\bar{\omega}_{i-1}}(\xi)d\xi + \frac{2\mu_n^{1/a}}{\mu_n - 2} + \int_{\mu_n^{1/a}}^{\infty}[1 - F_{i,\nu}^{\bar{\omega}_{i-1}}(\xi)]d\xi \overset{(3.22)}{\leq} \bar{\delta}_{\mu_n} \tag{3.23}$$

where

$$\bar{\delta}_{\mu_n} = 2\int_{\mu_n^{1/a}}^{\infty} h(\xi)d\xi + \frac{2\mu_n^{1/a}}{\mu_n - 2}$$

and it could be shown analogously to proof of Theorem 1.3. that

$$\bar{\delta}_{\mu_n} = \bar{\delta}_{\lfloor n^{1/k} \rfloor} = O(n^{1/k(-1+1/a)}). \tag{3.24}$$

Let $\Pi_1^{\mu_n}$ be the grid discretization defined by $\Pi^{\mu_n}_{1,1}, \Pi^{\mu_n}_{1,2}, \ldots, \Pi^{\mu_n}_{1,k_1}$ and let $\Pi^{\mu_n}_{i,\bar{\omega}_{i-1}}$ be the grid discretization defined by $\Pi^{\mu_n}_{i,1,\bar{\omega}_{i-1}}, \Pi^{\mu_n}_{i,2,\bar{\omega}_{i-1}}, \ldots, \Pi^{\mu_n}_{i,k_i,\bar{\omega}_{i-1}}$ for each $i = 2, 3, \ldots, M$, $\bar{\omega}_{i-1} \in \Omega_1 \times \Omega_2 \times \ldots \times \Omega_{i-1}$ (see Remark 1.2., (iii) for explanation how $\Pi^{\mu_n}_{i,\bar{\omega}_{i-1}}$ is defined using its marginal distributions) and let Π^n be the approximation scheme defined by $\Pi_1^{\mu_n}, \Pi^{\mu_n}_{2,\bullet}, \ldots, \Pi^{\mu_n}_{M,\bullet}$. Clearly, Π^n has

$$\mu_n^{k_1} \cdot \mu_n^{k_2} \cdot \ldots \cdot \mu_n^{k_M} = \mu_n^k \leq n$$

atoms and

$$\eta_{\Pi_n} \overset{\text{Theorem 3.2.}}{\leq} \bar{\delta}_{\mu_n} \sum_{i=1}^{M} ik_iK_i \overset{(3.24)}{=} O(n^{1/k(-1+1/a)}).$$

□

3.2.2 Monte Carlo

The Monte Carlo approximation of MSSPP may be viewed as the replacement of the original distribution by a discrete distribution with random atoms. The procedure of the approximation is as follows:

At first, a random sample

$$\zeta_1^1, \zeta_2^1, \ldots, \zeta_{n_1}^1, \quad \zeta_{j_1}^1 \sim P_1, \quad j_1 = 1, 2, \ldots, n_1 \tag{3.25}$$

is generated. Further, we proceed by induction: Once $\zeta^1_{j_1,j_2,\ldots,j_i}$ is generated for some $1 \leq j_1 \leq n_1, 1 \leq j_2 \leq n_2, \ldots, 1 \leq j_i \leq n_i$, $i < M$, we generate a (conditionally independent) sample from $P_{i+1}(\bullet|\zeta^1_{j_1}, \zeta^1_{j_1,j_2}, \ldots, \zeta^1_{j_1,j_2,\ldots,j_i})$. Mathematically speaking, we define

$$\zeta^1_{j_1,j_2,\ldots,j_i,1}, \zeta^1_{j_1,j_2,\ldots,j_i,2}, \ldots, \zeta^1_{j_1,j_2,\ldots,j_i,n_{i+1}},$$

$$\mathcal{L}(\zeta^1_{j_1,j_2,\ldots,j_i,j_{i+1}}|\zeta^1_{j_1}, \zeta^1_{j_1,j_2}, \ldots, \zeta^1_{j_1,j_2,\ldots,j_i}) = P_{i+1}(\bullet|\zeta^1_{j_1}, \zeta^1_{j_1,j_2}, \ldots, \zeta^1_{j_1,j_2,\ldots,j_i}), \quad j_{i+1} = 1, 2, \ldots, n_{i+1} \tag{3.26}$$

such that $\zeta^1_{j_1,j_2,\ldots,j_i,j_{i+1}}$, $j_{i+1} = 1, 2, \ldots, n_{i+1}$, are mutually conditionally independent given $\zeta^1_{j_1}, \zeta^1_{j_1,j_2}, \ldots, \zeta^1_{j_1,j_2,\ldots,j_i}$.[4] We repeat the induction step until $\zeta^1_{j_1,j_2,\ldots,j_M}$ is generated for each $j_1 \leq n_1, j_2 \leq n_2, \ldots, j_M \leq n_M$.

[4]We say that variables A and B are conditionally independent given variable C if there exists $\mathcal{L}(A, B|C)$, $\mathcal{L}(A|C)$ and $\mathcal{L}(B|C)$ such that $\mathcal{L}(A, B|c) = \mathcal{L}(A|c) \otimes \mathcal{L}(B|c)$ for each possible value c of C.

Then, P_1 is replaced by the discrete distribution $\left\{\zeta_{j_1}^1, 1/n_1\right\}_{j_1=1}^{n_1}$, each $P_i(\bullet|\zeta_{j_1}^1,\ldots,\zeta_{j_1,\ldots,j_{i-1}}^1)$, $1 \le j_1 \le n_1$, ..., $1 \le j_i \le n_i$, $i = 2,3,\ldots,M$, is replaced by the discrete distribution $\left\{\zeta_{j_1,j_2,\ldots,j_i}^1, 1/n_i\right\}_{j_i=1}^{n_i}$ and the approximate problem is solved.

In the sequel, we shall write $\bar{\zeta}_{j_1,j_2,\ldots,j_i}^1$ instead of $\zeta_{j_1}^1, \zeta_{j_1,j_2}^1, \ldots, \zeta_{j_1,j_2,\ldots,j_i}^1$.

The idea of sampling from conditional distributions to approximate MSSPP, described above, is not new. In [43], for instance, the consistency of the approximate problem's optimal value as an estimator of the exact problem's optimal value is shown. In the present work the optimality gap of the MC approximation of MSSPP is defined, its decomposition and its upper bound are provided and the convergence rate is shown.

Remark 3.4.

$$\bar{\zeta}_{j_1,j_2,\ldots,j_M}^1 \sim P \tag{3.27}$$

for each $1 \le j_1 \le n_1, 1 \le j_2 \le n_2, \ldots, 1 \le j_M \le n_M$.

Proof. We proceed by induction:

From (3.25), we have that $\mathcal{L}(\zeta_{j_1}^1) = P_1$ for each $1 \le j_1 \le n_1$.

Induction step. Let $1 < i \le M$, put $Q_i(\bullet|\bullet) := \mathcal{L}(\zeta_{j_1,j_2,\ldots,j_i}^1|\bar{\zeta}_{j_1,j_2,\ldots,j_{i-1}}^1)$ and $\bar{Q}_i := \mathcal{L}(\bar{\zeta}_{j_1,j_2,\ldots,j_i}^1)$, denote $\bar{P}_i$ the marginal distribution of elements from $\Omega_1 \times \Omega_2 \times \ldots \times \Omega_i$ with respect to P and assume that

$$\bar{Q}_{i-1} = \bar{P}_{i-1}. \tag{3.28}$$

Let $A \subseteq \Omega_1 \times \ldots \times \Omega_{i-1}$ be measurable set (w.r.t. $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_{i-1}$) and let $B \subseteq \Omega_i$ be measurable set (w.r.t. $\mathcal{A}_i$). Then, according to the definition of conditional probability,

$$\begin{aligned}
\bar{Q}_i[A \times B] &\overset{\text{def. of c. p.}}{=} \int_A Q_i(B|z_1, z_2, \ldots, z_{i-1}) d\bar{Q}_{i-1}(z_1, z_2, \ldots, z_{i-1}) \\
&\overset{(3.28)}{=} \int_A Q_i(B|z_1, z_2, \ldots, z_{i-1}) d\bar{P}_{i-1}(z_1, z_2, \ldots, z_{i-1}) \\
&\overset{(3.26)}{=} \int_A P_i(B|z_1, z_2, \ldots, z_{i-1}) d\bar{P}_{i-1}(z_1, z_2, \ldots, z_{i-1}) \\
&\overset{\text{def. of c. p.}}{=} \bar{P}_i[A \times B].
\end{aligned}$$

By putting $i = M$ we get the assertion of the Remark.

□

Let $\bar{\xi}_M = (\xi_1, \xi_2, \ldots, \xi_M) \sim P$. Similarly to the start of the subsection 3.2.1, we assume that once the "chance" has realized itself as $\bar{\xi}_i := (\xi_1, \xi_2, \ldots, \xi_i)$ at stage $1 \le i < M$, the user of the model solves (3.8) the way analogous to the first stage: First, the distribution $P_{i+1}(\bullet|\bar{\xi}_i)$ is replaced by the distribution with probabilities $1/n_{i+1}$ and with atoms

$$\zeta_1^{i+1}, \zeta_2^{i+1}, \ldots, \zeta_{n_{i+1}}^{i+1}, \tag{3.29}$$

$$\mathcal{L}(\zeta_{j_{i+1}}^{i+1}|\bar{\xi}_i) = P_{i+1}(\bullet|\bar{\xi}_i), \quad j_{i+1} = 1, 2, \ldots, n_{i+1}, \tag{3.30}$$

such that $\zeta_1^{i+1}, \zeta_2^{i+1}, \ldots, \zeta_{n_{i+1}}^{i+1}$ are conditionally independent given $\bar{\xi}_i$. Then, each distribution $P_\nu(\bullet|\bar{\xi}_i, \zeta_{j_{i+1}}^{i+1}, \zeta_{j_{i+2}}^{i+1}, \ldots, \zeta_{j_{\nu-1}}^{i+1})$ is replaced by the distribution with probabilities $1/n_\nu$ and with atoms

$$\zeta_{j_{i+1},j_{i+2},\ldots,1}^{i+1}, \zeta_{j_{i+1},j_{i+2},\ldots,2}^{i+1}, \ldots, \zeta_{j_{i+1},j_{i+2},\ldots,n_\nu}^{i+1} \tag{3.31}$$

$$\mathcal{L}(\zeta^{i+1}_{j_{i+1},j_{i+2},\ldots,j_\nu}|\bar{\xi}_i,\bar{\zeta}^{i+1}_{j_{i+1},j_{i+2},\ldots,j_{\nu-1}}) = P_\nu(\bullet|\bar{\xi}_i,\bar{\zeta}^{i+1}_{j_{i+1},j_{i+2},\ldots,j_{\nu-1}}),\quad j_\nu = 1,2,\ldots,n_\nu, \tag{3.32}$$

such that

$$\zeta^{i+1}_{j_{i+1},j_{i+2},\ldots,1},\zeta^{i+1}_{j_{i+1},j_{i+2},\ldots,2},\ldots,\zeta^{i+1}_{j_{i+1},j_{i+2},\ldots,n_\nu}$$

are conditionally independent given $(\bar{\xi}_i,\bar{\zeta}^{i+1}_{j_{i+1},j_{i+2},\ldots,j_{\nu-1}})$ (in the sequel, we shall denote

$$\bar{\zeta}^{i+1}_{j_{i+1},j_{i+2},\ldots,j_\nu} = \zeta^{i+1}_{j_{i+1}},\zeta^{i+1}_{j_{i+1}j_{i+2}},\ldots,\zeta^{i+1}_{j_{i+1},\ldots,j_\nu}$$

for readability).

Remark 3.5.

$$(\bar{\xi}_i,\bar{\zeta}^{i+1}_{j_{i+1},j_{i+2},\ldots,j_M}) \sim P \tag{3.33}$$

for each $j_{i+1} = 1,2,\ldots,n_{i+1}, j_{i+2} = 1,2,\ldots,n_{i+2},\ldots,j_M = 1,2,\ldots,n_M$.

Proof. It could be proved similarly to Remark 3.4. that

$$\mathcal{L}(\bar{\zeta}^{i+1}_{j_{i+1},j_{i+2},\ldots,j_M}|\bar{\xi}_i) = \mathcal{L}(\xi_{i+1},\xi_{i+2},\ldots,\xi_M|\bar{\xi}_i)$$

which implies the assertion of the present Remark.

□

Denote

$$\Xi_i = (\bar{\zeta}^i_{1,1,\ldots,1},\bar{\zeta}^i_{1,1,\ldots,2},\ldots,\bar{\zeta}^i_{n_i,n_{i+1},\ldots,n_M})$$

the vector of all atoms of approximating distributions at the i-th stage and assume that

(2.a) Ξ_1 is independent of $\bar{\xi}_M$ and

(2.b) Ξ_i is conditionally independent of $\bar{\xi}_M$ given $\bar{\xi}_{i-1}$ for each $1 < i \leq M$.

It would seem straightforward to define the optimality gap as a conditional expectation given all the approximating distributions. However, if we did so, then we would get into trouble with the dependence of the atoms $\zeta^i_\bullet$, $1 < i \leq M$, on ξ: First, it would be difficult to work with such a quantity. Second, it would be strange to think of "expected costs given the future Monte Carlo sampling". Moreover, we may encounter problems with the measurability of the approximate solution. Therefore, we proceed different way.

The actual loss caused by the approximation is

$$g_M(\tilde{x}_1,\tilde{x}_2(\tilde{x}_1),\ldots,\tilde{x}_M(\tilde{x}_1,\ldots,\tilde{x}_{M-1}),\bar{\xi}_M) - g_M(\hat{x}_1,\hat{x}_2(\hat{x}_1),\ldots,\hat{x}_M(\hat{x}_1,\ldots,\hat{x}_{M-1}),\bar{\xi}_M)$$

where $\hat{x}_1$ (an arbitrary) solution of (3.1), $\tilde{x}_1 = \tilde{x}_1(\Xi_1)$ (an arbitrary solution) of the first stage approximating problem, $\hat{x}_i(\bar{x}_{i-1}) = \hat{x}_i(\bar{x}_{i-1},\bar{\xi}_{i-1})$ (an arbitrary) solution of (3.8) with $\bar{\omega}_{i-1} = \bar{\xi}_{i-1}$ and $\tilde{x}_i(\bar{x}_{i-1}) = \tilde{x}_i(\bar{x}_{i-1},\bar{\xi}_{i-1},\Xi_i)$ (an arbitrary) solution of the i-th stage approximating problem given that decision(s) $\bar{x}_{i-1}$ were taken in the first $i-1$ stages. Of course, we may write

$$g_M(\tilde{x}_1,\tilde{x}_2,\ldots,\tilde{x}_M,\bar{\xi}_M) - g_M(\hat{x}_1,\hat{x}_2,\ldots,\hat{x}_M,\bar{\xi}_M) = \sum_{i=1}^M h_i$$

where

$$\begin{aligned} h_i &= h_i(\tilde{\xi}_M,\Xi_1,\ldots,\Xi_i) \\ &= g_M(\tilde{x}_1,\ldots,\tilde{x}_i,\hat{x}_{i+1},\ldots,\hat{x}_M,\bar{\xi}_M) - g_M(\tilde{x}_1,\ldots,\tilde{x}_{i-1},\hat{x}_i,\ldots,\hat{x}_M,\bar{\xi}_M). \end{aligned}$$

If all $\tilde{x}_1, \tilde{x}_2(\tilde{x}_1), \ldots, \tilde{x}_M(\tilde{x}_1, \ldots, \tilde{x}_{M-1})$ are measurable then we could define (but we will not) the optimality gap as

$$\tilde{\eta}'_{n_1,n_2,\ldots,n_M} = \sum_{i=1}^{M} E(h_i|\bar{\xi}_{i-1}, \Xi_1, \ldots, \Xi_i).$$

It is well known that if X, Y and Z are random variables such that X is conditionally independent of Y given Z then

$$\mathcal{L}(X|Y,Z) = \mathcal{L}(X|Z). \tag{3.34}$$

Using it we get

$$\begin{aligned}
&E(h_i|\bar{\xi}_{i-1}, \Xi_1, \ldots, \Xi_i) \\
&= \int_{\Omega_i \times \ldots \times \Omega_M} h_i(\bar{\xi}_{i-1}, \omega_i, \ldots, \omega_M, \Xi_1, \ldots, \Xi_i) \\
&\qquad d\mathcal{L}(\xi_i, \ldots, \xi_M|\bar{\xi}_{i-1}, \Xi_1, \ldots, \Xi_i)(\omega_i, \ldots, \omega_M) \\
&\overset{(2.b),(3.34)}{=} \int_{\Omega_i \times \ldots \times \Omega_M} h_i(\bar{\xi}_{i-1}, \omega_i, \ldots, \omega_M, \Xi_1, \ldots, \Xi_i) \\
&\qquad d\mathcal{L}(\xi_i, \ldots, \xi_M|\bar{\xi}_{i-1})(\omega_i, \ldots, \omega_M) \\
&= E_{P_i(\omega_i|\bar{\xi}_{i-1})} g_i(\tilde{x}_1, \ldots, \tilde{x}_{i-1}, \tilde{x}_i, \bar{\xi}_{i-1}, \omega_i) \\
&\qquad - E_{P_i(\omega_i|\bar{\xi}_{i-1})} g_i(\tilde{x}_1, \ldots, \tilde{x}_{i-1}, \hat{x}_i, \bar{\xi}_{i-1}, \omega_i).
\end{aligned} \tag{3.35}$$

(we got the last "=" by a successive application of the fact, following from (3.2), that

$$g_\nu(\bar{x}_\nu, \bar{\omega}_\nu) = E_{P_{\nu+1}(\omega_{\nu+1}|\bar{\omega}_\nu)} g_{\nu+1}(\bar{x}_\nu, \hat{x}_{\nu+1}, \bar{\omega}_{\nu+1})$$

for each possible $\bar{\omega}_\nu$, each $\bar{x}_\nu$ feasible given $\bar{\omega}_{\nu-1}$ and each $\nu = i, \ldots, M-1$).

Now we may formulate an analogy of Lemma 3.1.

Lemma 3.2. If all $\tilde{x}_1, \tilde{x}_2(\tilde{x}_1), \ldots, \tilde{x}_M(\tilde{x}_1, \ldots, \tilde{x}_{M-1})$ are measurable and if the assumptions (2.a) and (2.b) are fulfilled then it holds that

$$\tilde{\eta}'_{n_1,n_2,\ldots,n_M} = \sum_{i=1}^{M} \tilde{h}^i_{n_i,\ldots,n_M}$$

where

$$\tilde{h}^i_{n_i,\ldots,n_M} = E_{P_i(\omega_i|\bar{\xi}_{i-1})} g_i(\tilde{x}_1, \ldots, \tilde{x}_{i-1}, \tilde{x}_i, \bar{\xi}_{i-1}, \omega_i) - E_{P_i(\omega_i|\bar{\xi}_{i-1})} g_i(\tilde{x}_1, \ldots, \tilde{x}_{i-1}, \hat{x}_i, \bar{\xi}_{i-1}, \omega_i).$$

Proof. The assertion follows from the previous text.
□

Unfortunately, no sufficient condition assuring the measurability of the approximate solutions, applicable to the definition (3.1) of MSSPP, is known to the author. However, we may define the *optimality gap of the Monte Carlo approximation of MSSPP* even for non-measurable approximate solutions, using the reformulation given by Lemma 3.2., as

$$\begin{aligned}
&\tilde{\eta}_{n_1,n_2,\ldots,n_M} := \textstyle\sum_{i=1}^{M} \tilde{h}^i_{n_i,\ldots,n_M} \\
&= \sum_{i=1}^{M} [E_{P_i(\omega_i|\bar{\xi}_{i-1})} g_i(\tilde{x}_1, \ldots, \tilde{x}_{i-1}, \tilde{x}_i, \bar{\xi}_{i-1}, \omega_i) - E_{P_i(\omega_i|\bar{\xi}_{i-1})} g_i(\tilde{x}_1, \ldots, \tilde{x}_{i-1}, \hat{x}_i, \bar{\xi}_{i-1}, \omega_i)].
\end{aligned}$$

It should be stressed that $\tilde{\eta}_{n_1,n_2,\ldots,n_M}$ may not be measurable, however, we introduce its measurable bounds in the present work.
The following assertion is an analogy of Theorem 3.1.

Theorem 3.4. There exists a random variable G such that

$$\eta_{n_1,n_2,\ldots,n_M} \le G \tag{3.36}$$

for each choice of approximate problem's solution and

$$EG \le \sum_{i=1}^{M} i \cdot [ED_i + ED_i'] \tag{3.37}$$

where

$$\begin{aligned} D_i &= D_i(\bar{\xi}_{i-1}, \zeta_\bullet^i) = D_i(\bar{\xi}_{i-1}, \zeta_1^i, \ldots, \zeta_{n_i}^i) \\ &= \sup_{x_1 \in \mathcal{X}_1, \bar{x}_2 \in \mathcal{X}_2(x_1,\xi_1), \ldots, \bar{x}_i \in \mathcal{X}_i(\bar{x}_{i-1}, \bar{\xi}_{i-1})} \left[E_{P_i(\omega_i|\bar{\xi}_{i-1})} g_i(\bar{x}_i, \bar{\xi}_{i-1}, \omega_i) - \frac{1}{n_i} \sum_{j_i=1}^{n_i} g_i(\bar{x}_i, \bar{\xi}_{i-1}, \zeta_{j_i}^i) \right] \end{aligned}$$

and

$$\begin{aligned} D_i' &= D_i'(\bar{\xi}_{i-1}, \zeta_\bullet^i) = D_i'(\bar{\xi}_{i-1}, \zeta_1^i, \ldots, \zeta_{n_i}^i) \\ &= \sup_{x_1 \in \mathcal{X}_1, \bar{x}_2 \in \mathcal{X}_2(x_1,\xi_1), \ldots, \bar{x}_i \in \mathcal{X}_i(\bar{x}_{i-1}, \bar{\xi}_{i-1})} \left[\frac{1}{n_i} \sum_{j_i=1}^{n_i} g_i(\bar{x}_i, \bar{\xi}_{i-1}, \zeta_{j_i}^i) - E_{P_i(\omega_i|\bar{\xi}_{i-1})} g_i(\bar{x}_i, \bar{\xi}_{i-1}, \omega_i) \right] \end{aligned}$$

(both sides of (3.37) may take infinite values).

Proof. Similarly to the proof of Theorem 3.1. we get that

$$\tilde{h}^1_{n_1,n_2,\ldots,n_M} \le G_1 + G_1' \tag{3.38}$$

where

$$G_1 := \sup_{x_1 \in \mathcal{X}_1} \left[Eg_1(x_1, \xi_1) - \frac{1}{n_1} \sum_{j_1=1}^{n_1} \tilde{g}_1(x_1, \zeta_{j_1}^1) \right]$$

and

$$G_1' := \sup_{x_1 \in \mathcal{X}_1} \left[\frac{1}{n_1} \sum_{j_1=1}^{n_1} \tilde{g}_1(x_1, \zeta_{j_1}^1) - Eg_1(x_1, \xi_1) \right].$$

($\tilde{g}_1$ is defined by (3.7), we have put $\varsigma_{j_1,\ldots,j_\nu} = \zeta^1_{j_1,\ldots,j_\nu}$ and $p_{j_1,\ldots,j_\nu} = 1/n_\nu$ for each ν in the definition).

Further, it could be shown analogously to the proof of Theorem 3.1. that

$$G_1 \le D_1(\zeta_\bullet^1) + \sum_{\nu=1}^{M-1} \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \cdots \sum_{j_\nu=1}^{n_\nu} \frac{1}{\prod_{i=1}^{\nu} n_i} D_{\nu+1}(\bar{\zeta}^1_{j_1,j_2,\ldots,j_\nu}, \zeta^1_{j_1,j_2,\ldots,j_\nu,\bullet}) \tag{3.39}$$

and, by taking expectation on the both sides of (3.39) and by using Remark 3.4. and Remark 3.5., we get that

$$EG_1 \le E \sum_{\nu=1}^{M} D_\nu(\xi_{\nu-1}, \zeta_\bullet^\nu). \tag{3.40}$$

Analogously we obtain that

$$EG'_1 \leq E\sum_{\nu=1}^{M} D'_\nu(\xi_{\nu-1}, \zeta_\bullet^\nu)..\tag{3.41}$$

When we apply the same proof to the i-the stage problem, we get that, for each $i = 2, 3, \ldots, M$, there exist G_i and G'_i such that

$$\tilde{h}^i_{n_i,\ldots,n_M} \leq G_i + G'_i \tag{3.42}$$

and

$$E(G_i + G'_i) \leq E\sum_{\nu=i}^{M}[D_\nu + D'_\nu]. \tag{3.43}$$

By the combination of (3.38), (3.40), (3.41), (3.42) and (3.43) we get that

$$G := \sum_{i=1}^{n}[G_i + G'_i]$$

fulfils (3.36) and (3.37). Hence, the Theorem is proven.
□

Before we state a bound of the MSSPP approximation convergence rate, assume that

(2.c) $\mathcal{X}_1 \subset \mathrm{R}^{r_1}$ and $\mathcal{X}_i(\bar{\boldsymbol{x}}_{i-1}, \bar{\omega}_{i-1}) \subset \mathrm{R}^{r_i}$ for each possible $\bar{\omega}_{i-1}$, each $\bar{\boldsymbol{x}}_{i-1}$ feasible given $\bar{\omega}_{i-2}$, and $1 < i \leq M$,

(2.d) each function $g_i(\bullet, \bar{\omega}_i)$, $1 \leq i \leq M$, is uniformly l_2-Lipschitz (i.e. there exists $K_i \in \mathrm{R}$ such that $|g_i(\bar{\boldsymbol{x}}_i, \bar{\omega}_i) - g_i(\bar{\boldsymbol{y}}_i, \bar{\omega}_i)| \leq K_i\|\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{y}}_i\|_2$, for each possible $\bar{\omega}_i$ and each $\bar{\boldsymbol{x}}_i$,$\bar{\boldsymbol{y}}_i$ feasible given $\bar{\omega}_{i-1}$) and

(2.e) there exist $H_1, H_2, \ldots, H_M \in \mathrm{R}$ such that

$$\left|E\left(g_i(\bar{\boldsymbol{x}}_i, \bar{\xi}_i)\middle|\, \bar{\xi}_{i-1}\right) - g_i(\bar{\boldsymbol{x}}_i, \bar{\xi}_i)\right| \leq H_i$$

for each possible $\bar{\xi}_i$, $i = 1, 2, \ldots, M$.

Theorem 3.5. Denote

$$\hat{\mathcal{X}}_1 = \mathcal{X}_1$$

and

$$\hat{\mathcal{X}}_i = \left\{\bar{\boldsymbol{x}}_i : \bar{\boldsymbol{x}}_{i-1} \in \hat{\mathcal{X}}_{i-1}, \boldsymbol{x}_i \in \mathcal{X}_i(\bar{\boldsymbol{x}}_{i-1}, \bar{\omega}_{i-1}) \text{ for some } \bar{\omega}_{i-1} \in \Omega_1 \times \ldots \times \Omega_{i-1}\right\}$$

and assume that $\hat{\mathcal{X}}_i$ are compact for each $i = 1, 2, \ldots, M$. If $n_i = \lfloor n^{1/M} \rfloor$ for each $1 \leq i \leq M$ then there exist random variables $T_1, T_2, \ldots$ such that

$$\tilde{\eta}_n := \tilde{\eta}_{n_1,n_2,\ldots,n_M} \leq T_n$$

and that

$$T_n = o_P\left(n^{-\frac{1}{2M}+\delta}\right)$$

as $n \to \infty$ for each arbitrarily small $\delta > 0$.

Proof. We do an estimate similar to the one from the proof of Theorem 3.4.:

$$\begin{aligned}
T_{1,n} &:= \max_{\boldsymbol{x}_1\in\mathcal{X}_1}\left[\frac{1}{n_1}\sum_{j_1=1}^{n_1}\Big(Eg_1(\boldsymbol{x}_1,\xi_1)-\tilde{g}_1(\boldsymbol{x}_1,\zeta_{j_1}^1)\Big)\right]\\
&\le \max_{\boldsymbol{x}_1\in\mathcal{X}_1}\left[\frac{1}{n_1}\sum_{j_1=1}^{n_1}\Big(Eg_1(\boldsymbol{x}_1,\xi_1)-g_1(\boldsymbol{x}_1,\zeta_{j_1}^1)\Big)\right]\\
&\qquad+\max_{\boldsymbol{x}_1\in\mathcal{X}_1}\left[\frac{1}{n_1}\sum_{j_1=1}^{n_1}\Big(g_1(\boldsymbol{x}_1,\zeta_{j_1}^1)-\tilde{g}_1(\boldsymbol{x}_1,\zeta_{j_1}^1)\Big)\right]\\
&= \max_{\boldsymbol{x}_1\in\mathcal{X}_1}\left[\frac{1}{n_1}\sum_{j_1=1}^{n_1}\Big(Eg_1(\boldsymbol{x}_1,\xi_1)-g_1(\boldsymbol{x}_1,\zeta_{j_1}^1)\Big)\right]\\
&\qquad+\max_{\boldsymbol{x}_1\in\mathcal{X}_1}\left[\frac{1}{n_1n_2}\sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2}\Big(E_{P_2(\omega_2|\zeta_{j_1}^1)}g_2(\boldsymbol{x}_1,\hat{\boldsymbol{x}}_2,\zeta_{j_1}^1,\omega_2)-\tilde{g}_2(\boldsymbol{x}_1,\tilde{\boldsymbol{x}}_2,\bar{\zeta}_{j_1,j_2}^1)\Big)\right]\\
&\le \max_{\boldsymbol{x}_1\in\mathcal{X}_1}\left[\frac{1}{n_1}\sum_{j_1=1}^{n_1}\Big(Eg_1(\boldsymbol{x}_1,\xi_1)-g_1(\boldsymbol{x}_1,\zeta_{j_1}^1)\Big)\right]\\
&\qquad+\max_{\bar{\boldsymbol{x}}_2\in\hat{\mathcal{X}}_2}\left[\frac{1}{n_1n_2}\sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2}\Big(E_{P_2(\omega_2|\zeta_{j_1}^1)}g_2(\bar{\boldsymbol{x}}_2,\zeta_{j_1}^1,\omega_2)-\tilde{g}_2(\bar{\boldsymbol{x}}_2,\bar{\zeta}_{j_1,j_2}^1)\Big)\right]\\
&\le \cdots\\
&\le \sum_{\nu=1}^{M}\max_{\bar{\boldsymbol{x}}_i\in\hat{\mathcal{X}}_i}B_i(\bar{\boldsymbol{x}}_i)
\end{aligned}$$

where

$$B_i(\bar{\boldsymbol{x}}_i)=\frac{1}{\prod_{j=1}^{i}n_j}\sum_{j_1=1}^{n_1}\cdots\sum_{j_i=1}^{n_i}G^i_{j_1,\ldots,j_i},$$

$$G^i_{j_1,\ldots,j_i}=G^i(\bar{\boldsymbol{x}}_i,\bar{\zeta}^1_{j_1,\ldots,j_i})=\left[E\left(g_i(\bar{\boldsymbol{x}}_i,\bar{\zeta}^1_{j_1,\ldots,j_{i-1},1})\middle|\,\bar{\zeta}^1_{j_1,\ldots,j_{i-1}}\right)-g_i(\bar{\boldsymbol{x}}_i,\bar{\zeta}^1_{j_1,\ldots,j_i})\right].$$

Similarly we could get that

$$T'_{1,n}:=\max_{x_1\in\mathcal{X}_1}\left[\frac{1}{n_1}\sum_{j_1=1}^{n_1}\Big(\tilde{g}_1(x_1,\zeta_{j_1}^1)-Eg_1(x_1,\xi_1)\Big)\right]\le\sum_{\nu=1}^{M}\max_{\bar{\boldsymbol{x}}_i\in\hat{\mathcal{X}}_i}B'_i(\bar{\boldsymbol{x}}_i)$$

where

$$B'_i(\bar{\boldsymbol{x}}_i)=\frac{1}{\prod_{j=1}^{i}n_j}\sum_{j_1=1}^{n_1}\cdots\sum_{j_i=1}^{n_i}G'^i_{j_1,\ldots,j_i},$$

$$G'^i_{j_1,\ldots,j_i}=G'^i(\bar{\boldsymbol{x}}_i,\bar{\zeta}^1_{j_1,\ldots,j_i})=\left[g_i(\bar{\boldsymbol{x}}_i,\bar{\zeta}^1_{j_1,\ldots,j_i})-E\left(g_i(\bar{\boldsymbol{x}}_i,\bar{\zeta}^1_{j_1,\ldots,j_{i-1},1})\middle|\,\bar{\zeta}^1_{j_1,\ldots,j_{i-1}}\right)\right].$$

The rest of the proof is very similar to proof one of Theorem 2.3.: Denote $d=M\delta$. Since

$$E(G^i_{j_1,j_2,\ldots,j_i}|\bar{\zeta}^1_{j_1,\ldots,j_{i-1}})=0,$$

$$\begin{aligned}
\operatorname{var}(G^i_{j_1,j_2,\ldots,j_i}|\bar{\zeta}^1_{j_1,\ldots,j_{i-1}}) &= E\left(\left[E\left(g_i(\bar{\boldsymbol{x}}_i,)|\,\bar{\zeta}^1_{j_1,\ldots,j_{i-1}}\right)-g_i(\bar{\boldsymbol{x}}_i,\bar{\zeta}^1_{j_1,\ldots,j_i})\right]^2\middle|\,\bar{\zeta}^1_{j_1,\ldots,j_{i-1}}\right)\\
&\overset{\text{Remark 3.4., (2.e)}}{\le} H_i^2 \qquad (3.44)
\end{aligned}$$

and since $G^i_{j_1,j_2,\ldots,1}, G^i_{j_1,j_2,\ldots,2}, \ldots, G^i_{j_1,j_2,\ldots,n_1}$ are conditionally independent given $\bar{\zeta}^1_{j_1,\ldots,j_i}$, we may apply Theorem A.4. to get

$$
\begin{aligned}
&P\left(n_i^{1/2-d}\tfrac{1}{n_i}\textstyle\sum_{j_i=1}^{n_i} G^i_{j_1,j_2,\ldots,j_i} \geq \eta \middle| \zeta^1_{j_1,j_2,\ldots,j_{i-1}}\right) \\
&\quad = \quad P\left(n_i^{1/2}\frac{1}{n_i}\sum_{j_i=1}^{n_i} G^i_{j_1,j_2,\ldots,j_i} \geq n_i^d\eta \middle| \zeta^1_{j_1,j_2,\ldots,j_{i-1}}\right) \\
&\quad \leq \quad \exp\left\{-\frac{n_i^{2d}\eta^2}{2\operatorname{var}(G^i_{j_1,\ldots,j_i}|\zeta^1_{j_1,j_2,\ldots,j_{i-1}})}\psi\left(\frac{n_i^d\eta H_i}{\operatorname{var}(G^i_{j_1,\ldots,j_i}|\zeta^1_{j_1,j_2,\ldots,j_{i-1}})\sqrt{n}}\right)\right\} \\
&\quad \overset{\text{(b) of Th. A.4.}}{\leq} \quad \exp\left\{-\frac{n_i^{2d}\eta^2}{2H_i^2}\psi\left(\frac{n_i^d\eta H_i}{H_i^2\sqrt{n}}\right)\right\} \overset{\text{(c) of Th. A.4.}}{\leq} \exp\left\{-\frac{n_i^{2d}\eta^2}{2H_i^2}\frac{1}{1+\frac{n_i^d\eta}{3H_i\sqrt{n}}}\right\} \\
&\quad \leq \quad \exp\left\{-n_i^d\frac{\eta^2}{2H^2n_i^{-d}+\frac{2}{3}H_i\eta n_i^{-1/2}}\right\} \leq \exp\left\{-n_i^dC_\eta^i\right\}
\end{aligned}
$$

where $C^i_\eta = \frac{\eta^2}{2H_i^2+\frac{2}{3}\eta H_i}$. Since the upper bound does not depend on the condition, we also have that

$$
P\left(n_i^{1/2-d}\frac{1}{n_i}\sum_{j_i=1}^{n_i} G^i_{j_1,j_2,\ldots,j_i} \geq \eta\right) \leq \exp\left\{-n_i^dC_\eta^i\right\}.
$$

Further, since $P(\bar{X}_n \geq a) \leq \sum_{i=1}^n P(X_i \geq a)$ for each sequence of random variables $X_1, \ldots, X_n$, we get

$$
P\left(n_i^{1/2-d}B_i(\bar{\boldsymbol{x}}_i) \geq \eta\right) \leq \prod_{\nu=1}^{i-1} n_\nu \exp\left\{-n_i^dC_\eta^i\right\}
$$

(we take $\prod_{\nu=1}^0 n_\nu = 1$). Finally, we get, analogously to the proof of Theorem 2.3., that there exist $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{N(n_i,\eta)} \in \hat{\mathcal{X}}_i$,

$$
N(n_i,\eta) \leq \left(D^i_\eta n_i^{1/2-d}+1\right)^{\sum_{\nu=1}^i k_\nu}
$$

for some constant D^i_η, independent of n_i, such that

$$
\begin{aligned}
P\left(n_i^{1/2-d}\max_{\bar{\boldsymbol{x}}_i\in\hat{\mathcal{X}}_i} B_i(\bar{\boldsymbol{x}}_i) \geq \eta\right) \quad &\leq \quad \sum_{\nu=1}^{N(n_i,\eta)} P\left(n_i^{1/2-d}B_i(\boldsymbol{y}_\nu) \geq \eta/3\right) \\
&\overset{(3.51)}{\leq} \quad \left(D^i_\eta n_i^{1/2-d}+1\right)^{\sum_{\nu=1}^i k_\nu} \prod_{\nu=1}^{i-1} n_\nu \exp\left\{-n_i^dC_\eta^i/3\right\} \\
&= \quad \left(D^i_\eta \lfloor n^{1/k}\rfloor^{1/2-d}+1\right)^{\sum_{\nu=1}^i k_\nu} \lfloor n^{1/k}\rfloor^{i-1} \exp\left\{-\lfloor n^{1/k}\rfloor^d C_\eta^i/3\right\} \\
&\overset{n\to\infty}{\longrightarrow} \quad 0.
\end{aligned}
$$

so that

$$
\max_{\bar{\boldsymbol{x}}_i\in\hat{\mathcal{X}}_i} B_i(\bar{\boldsymbol{x}}_i) = o_P\left(n_i^{-1/2+d}\right) = o_P\left(n^{-\frac{1}{2M}+\delta}\right)
$$

Similarly we get that

$$
\max_{\bar{\boldsymbol{x}}_i\in\hat{\mathcal{X}}_i} B'_i(\bar{\boldsymbol{x}}_i) = o_P\left(n^{-\frac{1}{2M}+\delta}\right).
$$

so that

$$\tilde{h}^1_{n_1,n_2,\dots,n_M} \leq T_{1,n} + T'_{1,n} \leq \max_{\bar{\boldsymbol{x}}_i \in \hat{\mathcal{X}}_i} B_i(\bar{\boldsymbol{x}}_i) + \max_{\bar{\boldsymbol{x}}_i \in \hat{\mathcal{X}}_i} B'_i(\bar{\boldsymbol{x}}_i) = o_P\left(n^{-\frac{1}{2M}+\delta}\right). \tag{3.45}$$

Analogously we could get that there exist $T_{\nu,n}$, $T'_{\nu,n}$ such that

$$\tilde{h}^\nu_{n_\nu,\dots,n_M} \leq T_{\nu,n} + T'_{\nu,n} = o_P\left(n^{-\frac{1}{2M}+\delta}\right) \tag{3.46}$$

for each $1 < \nu \leq M$ (the only difference would be that we would have $(\bar{\xi}_{i-1}, \bar{\zeta^i}_{j_i,\dots,j_i})$ instead of $\zeta^1_{j_1,\dots,j_i}$ and that we would have less sums in the estimate of $T_{\nu,n}$ than in case of $T_{1,n}$).

Using the definition of the optimality gap and the relations (3.45) and (3.46) we would then get that

$$T_n := \sum_{\nu=1}^{M} (T_{\nu,n} + T'_{\nu,n})$$

fulfils the assertion of the Theorem.

□

One may ask whether the convergence rate $o_P(n^{-1/M+\delta})$ is not too conservative. We show that it is not: we construct a problem fulfilling the formal definition of MSSPP with the convergence rate of the expected optimality gap being $O(n^{-1/M})$ where n is the number of atoms of the approximating distribution.

Remark 3.6. For each $M \geq 1$ there exists an M-stage optimization problem such that

$$E\tilde{\eta}_n = E\tilde{\eta}_{n_1,\dots,n_M} = O\left(n^{-\frac{1}{M}}\right)$$

where $n_i = \lfloor \sqrt[M]{n} \rfloor$ for each $i = 1, 2, \dots, M$.

Proof. Consider MSSPP

$$\min_{x_1 \in C} E g_1(x_1, \xi_1) \tag{3.47}$$

where

$$g_M(\bar{x}_i, \bar{\xi}_i) = \sum_{i=1}^{M} (f_i(\xi_i) - x_i)^2 \tag{3.48}$$

and

$$g_i(\bar{x}_i, \bar{\xi}_i) = \min_{x_{i+1} \in C} E(g_{i+1}(\bar{x}_{i+1}, \bar{\xi}_{i+1}) | \bar{\xi}_i) \tag{3.49}$$

for each $i = 1, 2, \dots, M-1$, $C = [-c, c]$ is some interval, $f_i : \mathrm{R} \to \mathrm{R}$ are some Lipschitz functions such that $-c \leq f_i(\bullet) \leq c$ and $\mathrm{var}(f_i(\xi_i)) > 0$ for each i. Moreover, let ξ_i be independent of ξ_j for each $i \neq j$.

Since $E(X-x)^2$ is minimized by $x = EX$ for each random variable X, it is easy to determine the exact solution of the problem defined by (3.48) and (3.49) - it is

$$\hat{x}_1 = E f_1(\xi_1), \quad \hat{x}_2 = E f_2(\xi_2), \quad \dots, \quad \hat{x}_M = E f_M(\xi_M).$$

Similarly to the exact problem, the solution of the approximating problem defined by (3.48) and

$$\dot{g}_i(\bar{x}_i, \bar{\xi}_i) = \min_{x_{i+1} \in C} \frac{1}{n_{i+1}} \sum_{j_{i+1}=1}^{n_i} \dot{g}_{i+1}(\bar{x}_{i+1}, \bar{\xi}_i, \zeta^i_{j_i})$$

is

$$\tilde{x}_1 = \frac{1}{n_1} \sum_{j_1}^{n_1} f_1(\zeta^1_{j_1}), \quad \tilde{x}_2 = \frac{1}{n_2} \sum_{j_2=1}^{n_2} f_2(\zeta^2_{j_2}), \quad \dots, \quad \tilde{x}_M = \frac{1}{n_M} \sum_{j_M=1}^{n_M} f_M(\zeta^M_{j_M}).$$

We have that

$$\begin{aligned} E\tilde{h}^i_{n_i,\dots,n_M} &= E\left[E[(f_i(\xi_i)-\tilde{x}_i)^2|\Xi_i]-E(f_i(\xi_i)-\hat{x}_i)^2\right] \\ &= E[(f_i(\xi_i)-\tilde{x}_i)^2]-E(f_i(\xi_i)-\hat{x}_i)^2 \\ &= E(f_i(\xi_i)-Ef_i(\xi_i))^2+E(\tilde{x}_i-Ef_i(\xi_i))^2-E(f_i(\xi_i)-Ef_i(\xi_i))^2 \\ &= \operatorname{var}\tilde{x}_i=\frac{1}{n_i}\operatorname{var}f_i(\xi_i) \end{aligned}$$

(we have used the fact that $\zeta^i_j \sim \mathcal{L}(\xi_i)$ for each $1 \le j \le n_i$ following from (3.33)). Hence

$$E\tilde{\eta}_n = \frac{1}{\lfloor n^{1/M}\rfloor}\sum_{i=1}^{M}\operatorname{var}f_i(\xi_i) = O(n^{-1/M}).$$

□

Remark 3.7. The problem, constructed in the proof of Remark 3.6., fulfils the assumptions of Theorem 3.5.

Proof. It is obvious that the assumption (2.c) and the compactness of the sets $\hat{\mathcal{X}}$ are fulfilled. Since

$$g_i(\bar{x}_i,\bar{\omega}_i)=\sum_{j=1}^{i}E_{P_j}(f_j-x_j)^2+\sum_{j=i+1}^{m}E_{P_j}(f_j-\hat{x}_j)^2$$

is bounded, we immediately see that the assumption (2.e) is fulfilled. We prove the assumption (2.d): Since

$$|E_{P_i}(f_i-x)^2-E_{P_i}(f_i-y)^2| = |E_{P_i}(x^2-y^2+2(y-x)f_i)| \le |x+y||x-y|+2|x-y||E_Pf_i| \le 4c|x-y|,$$

the Lipschitz property of the function g_i follows easily.
□

Remark 3.8. The problem, constructed in the proof of Remark 3.6., fulfils the assumption (1.b) of the subsection 3.2.1.

Proof. f_i are bounded and Lipschitz, so that

$$|(f_i(\omega_1)-x_i)^2-(f_i(\omega_2)-x_i)^2| = |f_i(\omega_1)+f_i(\omega_2)-2x_i||f_i(\omega_1)-f_i(\omega_2)| \le C\|\omega_1-\omega_2\|$$

for some $C \in \mathrm{R}$.
□

3.2.3 Quasi Monte Carlo

Assume that

(3.a) $\Omega_i = \mathrm{R}^\kappa$, $i = 1,2,\dots,M$, for some $\kappa \in \mathrm{N}$ and $\mathcal{A}_i$ is Borel sigma algebra for each $i = 1,2,\dots,M$.

Further, suppose that $n_i = m_i^\kappa$ for some integers m_i, $i = 1,2,\dots,M$.

The Quasi Monte Carlo approximation is done a similar way as the Monte Carlo one, hence we keep the notation of the section 3.2.2 in the present section. The "only" difference from

the Monte Carlo is the procedure of generating the approximating distributions: For each $1 \leq \nu \leq M$, $\nu \leq i \leq M$, $1 \leq j_\nu \leq n_\nu, 1 \leq j_{\nu+1} \leq n_{\nu+1}, \ldots, 1 \leq j_i \leq n_i$, we put

$$\zeta^{\nu}_{j_\nu, j_{\nu+1}, \ldots, j_i} = \zeta^{\nu, n_i}_{j_\nu, j_{\nu+1}, \ldots, j_i} = \boldsymbol{t}_{P_i(\bullet|\bar{\xi}_{\nu-1}, \bar{\zeta}^{\nu, n_{i-1}}_{j_\nu, j_{\nu+1}, \ldots, j_{i-1}})}(\boldsymbol{U}^{\nu}_{n_i, j_\nu, j_2, \ldots, j_i}), \tag{3.50}$$

where $\boldsymbol{U}^{\nu}_{n_i, j_\nu, j_{\nu+1}, \ldots, j_i}$ is a random vector with the uniform distribution on a κ-dimensional cube with the edge $1/m_i$ such that

$$\boldsymbol{U}^{\nu}_{n_i, j_\nu, j_{\nu+1}, \ldots, 1}, \boldsymbol{U}^{\nu}_{n_i, j_\nu, j_{\nu+1}, \ldots, 2}, \ldots, \boldsymbol{U}^{\nu}_{n_i, j_\nu, j_{\nu+1}, \ldots, n_i}$$

are i.i.d. and their supports cover the unit cube (we take $P_1(\bullet|\bullet) = P_1(\bullet)$ and $(\bar{\xi}_0, \bar{\zeta}^1_{j_1, j_2, \ldots, j_{i-1}}) = \bar{\zeta}^1_{j_1, j_2, \ldots, j_{i-1}}$ in (3.50)).

Let us describe the procedure in words:

1. We split the unit cube into n_1 cubes, we generate exactly one random "observation" from the uniform distribution on each cube and we transform each "observation" by $\boldsymbol{t}_{P_1}$. The resulting set serve us as atoms of the first stage approximating distribution.

2. To get atoms of the distribution, approximating $P_i(\bullet|\bar{\zeta}^1_{j_1, \ldots, j_{i-1}})$, we split the unit cube into n_i cubes, generate an "observation" from each cube, and transform each "observation" by $\boldsymbol{t}_{P_i(\bullet|\bar{\zeta}^1_{j_1, \ldots, j_{i-1}})}$. We repeat this step until the approximating scenario tree is completely generated.

3. After the "chance" realized itself as $\bar{\xi}_{\nu-1}$, we approximate the ν-th interior problem according to the steps 1. and 2.

Let $\bar{\xi}_M \sim P$. In the present situation, we may adopt the definition of the optimality gap from the subsection 3.2.2, i.e. we define the *optimality gap of QMC approximation of MSSPP* as

$$\check{\eta}_{n_1, n_2, \ldots, n_M} := \sum_{i=1}^{M} \check{h}^i_{n_i, \ldots, n_M},$$

$$\check{h}^i_{n_i, \ldots, n_M} \;=\; E_{P_i(\omega_i|\bar{\xi}_{i-1})} g_i(\check{x}_1, \ldots, \check{x}_{i-1}, \check{x}_i, \bar{\xi}_{i-1}, \omega_i) - E_{P_i(\omega_i|\bar{\xi}_{i-1})} g_i(\check{x}_1, \ldots, \check{x}_{i-1}, \hat{x}_i, \bar{\xi}_{i-1}, \omega_i),$$

where $\check{x}_i$, $i = 1, 2, \ldots, M$ are the solutions of QMC approximation of MSSPP. If we further assume that

(3.b) Ξ_1 is independent of $\bar{\xi}_M$ and

(3.c) Ξ_i is conditionally independent of $\bar{\xi}_M$ given $\bar{\xi}_{i-1}$ for each $1 < i \leq M$

then we may derive a result analogous to Lemma 3.2.

Lemma 3.3. Let $\check{x}_i$ be measurable for each $i = 1, 2, \ldots, M$. Then, under the assumptions (3.a), (3.b) and (3.c) it holds that

$$\check{\eta}_{n_1, n_2, \ldots, n_M} = \sum_{i=1}^{M} E(\check{\mu}^i_{n_i, \ldots, n_M} | \bar{\xi}_{i-1}, \Xi_1, \ldots, \Xi_i)$$

where

$$\breve{\mu}^{i}_{n_i,\ldots,n_M} = g_M(\breve{x}_1,\ldots,\breve{x}_i,\hat{x}_{i+1},\ldots,\hat{x}_M,\bar{\xi}_M) - g_M(\breve{x}_1,\ldots,\breve{x}_{i-1},\hat{x}_i,\ldots,\hat{x}_M,\bar{\xi}_M).$$

Proof. The proof is the same as the proof of Lemma 3.2.
□

Assume, until the end of the subsection 3.2.3 that

(3.d) $\mathcal{X}_1 \subset \mathrm{R}^{r_1}$ and $\mathcal{X}_i(\bar{\boldsymbol{x}}_{i-1},\bar{\omega}_{i-1}) \subset \mathrm{R}^{r_i}$ for each possible $\bar{\omega}_{i-1}$, each $\bar{\boldsymbol{x}}_{i-1}$ feasible given $\bar{\omega}_{i-2}$, and $1 < i \leq M$,

(3.e) each function $g_i(\bullet,\bar{\omega}_i)$, $1 \leq i \leq M$, is uniformly l_2-Lipschitz (i.e. there exists $K_i \in \mathrm{R}$ such that $|g_i(\bar{\boldsymbol{x}}_i,\bar{\omega}_i) - g_i(\bar{\boldsymbol{y}}_i,\bar{\omega}_i)| \leq K_i\|\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{y}}_i\|_2$, for each possible $\bar{\omega}_i$ and each $\bar{\boldsymbol{x}}_i$,$\bar{\boldsymbol{y}}_i$ feasible given $\bar{\omega}_{i-1}$),

(3.f) each function $g_i(\bar{x}_i,\bar{\omega}_{i-1},\bullet)$, $1 \leq i \leq M$, is uniformly l_1-Lipschitz (i.e. there exists a constant L_i such that $|g_i(\bar{x}_i,\omega_1,\ldots,\omega_{i-1},\omega) - g_i(\bar{x}_i,\omega_1,\ldots,\omega_{i-1},\omega')| \leq L_i\|\omega - \omega'\|_1$, for each possible $\bar{\omega}_{i-1}$, each feasible $\bar{x}_i$ given $\bar{\omega}_{i-1}$ and each $\omega,\omega' \in \Omega_i$) and

(3.g) there exist $H_1, H_2, \ldots, H_M \in \mathrm{R}$ such that

$$\left|E\left(g_i(\bar{\boldsymbol{x}}_i,\bar{\xi}_i)\middle|\,\bar{\xi}_{i-1}\right) - g_i(\bar{\boldsymbol{x}}_i,\bar{\xi}_i)\right| \leq H_i$$

for each possible $\bar{\xi}_i$, $i = 1,2,\ldots,M$.

Under these assumptions, we may formulate the following Theorem.

Theorem 3.6. Let the sets $\hat{\mathcal{X}}_1,\ldots,\hat{\mathcal{X}}_M$, defined at Theorem 3.5., be compact. Further, let $\frac{\partial}{\partial u}F^{-1}_{\lambda}(u|z_1,z_2,\ldots,z_{\lambda-1})$ exist for each $u \in (0,1)$, each $z_1,z_2,\ldots,z_{\lambda-1} \in \mathrm{R}$, and each $1 \leq \lambda \leq k$, where $k = M\kappa$ (the symbol F_1 denotes the first marginal d. f. of P and, for each $\lambda = 2,3,\ldots,k$, $F_\lambda(\bullet|\boldsymbol{z})$ denotes the d. f. of the conditional distribution of the P's λ-th component given that the first $\lambda-1$ components equal to $\boldsymbol{z}$) and let there exist constants $C > 0$ and $0 \leq a \leq 1$ such that

$$\frac{\partial}{\partial u}F^{-1}_{\lambda}(u|z_1,z_2,\ldots,z_{\lambda-1}) \leq Cu^{-a}$$

for each $0 < u \leq 1/2$, $z_1,z_2,\ldots,z_{\lambda-1} \in \mathrm{R}$, $1 \leq \lambda \leq k$, and

$$\frac{\partial}{\partial u}F^{-1}_{\lambda}(u|z_1,z_2,\ldots,z_{\lambda-1}) \leq C(1-u)^{-a}$$

for each $1/2 \leq u < 1$, $z_1,z_2,\ldots,z_{\lambda-1} \in \mathrm{R}$, $1 \leq \lambda \leq k$. If $a \geq 1 - \kappa/2$ and if $n_i = \lfloor N_i^{1/\kappa}\rfloor^{\kappa}$, $N_i = n^{1/M}$, for each $1 \leq i \leq M$ then there exist random variables $T_1, T_2, \ldots$ such that

$$\breve{\eta}_n := \breve{\eta}_{n_1,n_2,\ldots,n_M} \leq T_n$$

and

$$T_n = o_P\left(n^{-\frac{2-2a+\kappa}{2M\kappa}+\delta}\right)$$

as $n \to \infty$ for each arbitrarily small $\delta > 0$.

Proof. Since the proof is nearly identical to the proof of Theorem 3.4., we shall keep its notation and we shall stress only the differences from the proof of the Theorem 3.4.

Fix $1 \leq i \leq M$, let $d = \delta M$ and define

$$\Gamma^i_n(\bar{\boldsymbol{x}}_i) := \frac{1}{n_i} \sum_{j_i=1}^{n_i} G^i_{j_1,j_2,\ldots,j_i}(\bar{\boldsymbol{x}}_i).$$

Since g_i is bounded according to (3.g) and Lipschitz in ω_i according to (3.f), since the d.f. of the first marginal of $P_i(\bullet|\bar{\zeta}^1_{j_1,\ldots,j_{i-1}})$ is $F_{\bar{k}_i+1}(\bullet|\bar{\zeta}^1_{j_1,\ldots,j_{i-1}})$, where $\bar{k}_i = (i-1)\kappa$, and since the j-th conditional d.f. of $P_i(\bullet|\bar{\zeta}^1_{j_1,\ldots,j_{i-1}})$ given $z_1,\ldots,z_{j-1}$ is $F_{\bar{k}_i+j}(\bullet|\bar{\zeta}^1_{j_1,\ldots,j_{i-1}}, z_1,\ldots,z_{j-1})$, we may apply Lemma 2.2. to the conditional distribution $P_i(\bullet|\bar{\zeta}^1_{j_1,\ldots,j_{i-1}})$ and get that for each $\eta > 0$ there exists a constant C^i_η independent of $\bar{\boldsymbol{x}}_i$ and n_i such that

$$P\left[n_i^{\frac{2-2a+\kappa}{2\kappa}-d} |\Gamma^i_n(\bar{\boldsymbol{x}}_i)| \geq \eta \middle| \bar{\zeta}^1_{j_1,\ldots,j_{i-1}} \right] \leq 2\exp\left\{-C^i_\eta n_i^d\right\}$$

for each $n > 1$ and $\bar{\boldsymbol{x}}_i \in \hat{\mathcal{X}}_i$, hence

$$P\left[n_i^{\frac{2-2a+\kappa}{2\kappa}-d} |\Gamma^i_n(\bar{\boldsymbol{x}}_i)| \geq \eta \right] \leq 2\exp\left\{-C^i_\eta n_i^d\right\}$$

for each $n > 1$ and $\bar{\boldsymbol{x}}_i \in \hat{\mathcal{X}}_i$. Similarly to the proof of Theorem 3.4., we get that

$$P\left(n_i^{\frac{2-2a+\kappa}{2\kappa}-d} B_i(\bar{\boldsymbol{x}}_i) \geq \eta \right) \leq \prod_{\nu=1}^{i-1} n_\nu \exp\left\{-n_i^d C^i_\eta\right\} \tag{3.51}$$

for each $n > 1$ and $\bar{\boldsymbol{x}}_i \in \hat{\mathcal{X}}_i$ (B_i is defined in the proof of Theorem 3.4.) and, thanks to (3.e),

$$P\left(n_i^{\frac{2-2a+\kappa}{2\kappa}-d} \max_{\bar{\boldsymbol{x}}_i \in \hat{\mathcal{X}}_i} B_i(\bar{\boldsymbol{x}}_i) \geq \eta \right) \leq \left(D^i_\eta n_i^{\frac{2-2a+\kappa}{2\kappa}-d} + 1 \right)^{i\kappa} \prod_{\nu=1}^{i-1} n_\nu \exp\left\{-n_i^d C^i_\eta/3\right\}$$

for some D^i_η independent of n_i, so that

$$P\left(n_i^{\frac{2-2a+\kappa}{2\kappa}-d} \max_{\bar{\boldsymbol{x}}_i \in \hat{\mathcal{X}}_i} B_i(\bar{\boldsymbol{x}}_i) \geq \eta \right) \overset{n_i \to \infty}{\longrightarrow} 0.$$

The same way as in the proof of Theorem 3.4., we may show that

$$\check{h}^1_{n_1,\ldots,n_M} \leq T_{1,n} + T'_{1,n} = o_P\left(n^{\frac{2-2a+\kappa}{2M\kappa}-\delta} \right)$$

and

$$\check{h}^i_{n_i,\ldots,n_M} \leq T_{i,n} + T'_{i,n} = o_P\left(n^{\frac{2-2a+\kappa}{2M\kappa}-\delta} \right).$$

for each $i = 2,\ldots,M$ ($T_{i,n}, T'_{i,n}$ are defined in the proof of Theorem 3.4.). By putting

$$T_n = \sum_{i=1}^{M} [T_{i,n} + T'_{i,n}]$$

we prove the present Theorem.

□

3.3 Numerical Illustration

3.3.1 Test Problem Definition

Consider the problem of buying the unit amount of a divisible commodity under the following conditions

1. It is possible to buy the commodity at instants $1, 2, \ldots, M+1$ for (random) nonnegative prices $p_1, p_2, \ldots, p_{M+1}$.
2. At each instant, it is possible to buy at most amount δ of the commodity.
3. Aim of the buyer is to minimize the total cost.

The mathematical definition of the problem is as follows:

$$\min_{0 \le x_1 \le \delta} E\left[p_1 x_1 + \phi_2(1 - x_1)\right] \tag{3.52}$$

where

$$\phi_i(x) = \min_{0 \le x_i \le \delta} E_i\left[p_i x_i + \phi_{i+1}(x - x_i)\right]$$

for $i = 2, 3, \ldots, M$ and

$$\phi_{M+1}(x) = \begin{cases} E_{M+1} p_{M+1} x & \text{if } 0 \le x \le \delta, \\ \infty & \text{otherwise} \end{cases}$$

(E_i stands for $E_{P_i(p_i|\bar{p}_{i-1})}$ and we adopt convention that $\min_{z \in \emptyset} g(z) = \infty$).

Evidently, (3.52) is M-stage MSSPP with

$$g_M(\bar{x}_M, \bar{p}_M) \;=\; \sum_{i=1}^{M} p_i x_i + E_{M+1} p_{M+1} \left(1 - \sum_{i=1}^{M} x_i\right)$$

and

$$g_i(\bar{x}_i, \bar{p}_i) \;=\; \min_{0 \le x_{i+1} \le \delta, \sum_{i=1}^{i+1} x_i \le 1} E_{i+1} g_{i+1}(\bar{x}_i, \bar{p}_{i+1})$$

$i = 1, 2, \ldots, M-1$.

3.3.2 Test Problem Solution

If we assume that δ is such that there exists $\kappa \le M+1$ fulfilling the condition $\kappa\delta = 1$ then (3.52) is equivalent to

$$\min_{x_1 \in \{0, \delta\}} E\left[p_1 x_1 + \phi'_2(1 - x_1)\right] \tag{3.53}$$

where

$$\phi'_i(x) = \min_{x_i \in \{0, \delta\}} E_i\left[p_i x_i + \phi'_{i+1}(x - x_i)\right]$$

for $i = 2, 3, \ldots, M$ and

$$\phi'_{M+1}(x) = \begin{cases} E_{M+1} p_{M+1} x & \text{if } x = 0 \text{ or } x = \delta \\ \infty & \text{otherwise} \end{cases}.$$

We prove the fact by induction: From the definition of ϕ_{M+1} we see that $\phi_{M+1}(\bullet)$ is piece-wise linear convex mapping to $\mathrm{R}\cup\{\infty\}$ with breaks at the multiples of δ (we regard function $g \equiv \infty$ as linear).

Let $1 \leq i \leq M$ and assume that $\phi_{M-i+1}(\bullet)$ is piece-wise linear convex with (possible) breaks at multiples of δ. If we denote

$$f(z) := E_{M-i}[-p_{M-i}z + \phi_{M-i+1}(z)].$$

then we may write

$$\begin{aligned}
\phi_{M-i}(x) &= \min_{0\leq y\leq\delta} E_{M-i}[p_{M-i}y + \phi_{M-i+1}(x-y)] \\
&= \min_{0\leq y\leq\delta} E_{M-i}[p_{M-i}y - p_{M-i}x + p_{M-i}x + \phi_{M-i+1}(x-y)] \\
&= E_{M-i}p_{M-i}x + \min_{0\leq y\leq\delta} f(x-y) = E_{M-i}p_{M-i}x + \min_{0\leq x-z\leq\delta} f(z) \\
&= E_{M-i}p_{M-i}x + \min_{x-\delta\leq z\leq x} f(z). \qquad (3.54)
\end{aligned}$$

It follows from the induction assumption that f is piece-wise linear convex with (possible) breaks at the multiples of δ. Denote $\hat{\zeta}(x)$ (one of the) points in which $f(\bullet)$ attains its minimum. From the piece-wise linearity of f it follows that we may choose $\hat{\zeta}(x)$ so that $\hat{\zeta}(x) = n\delta$ for some integer n. Since f is convex, one of the optimal solution(s) of the problem $\min_{x-\delta\leq z\leq x} f(z)$ happens either in $\hat{\zeta}(x)$ (if $\hat{\zeta}(x)$ lies in the feasibility set), on the left boundary of the feasibility set (if $\hat{\zeta}(x)$ lies left to the feasibility set) or on the right boundary of the feasibility set (if $\hat{\zeta}(x)$ lies on the right of the set), i.e.

$$\hat{z}(x) := \min(\max(x-\delta, \hat{\zeta}(x)), x) = \min(\max(x-\delta, n\delta), x) \qquad (3.55)$$

is an optimal solution of (3.54). It is clear that $\hat{z}(\bullet)$ is piecewise linear with breaks at multiples of δ. Moreover, since $\hat{z}(\bullet)$ is constant on $[n\delta, (n+1)\delta])$ and it is either constant or linear with the unit slope outside this interval, it has to be nondecreasing and, moreover, $\hat{z}(x_i)$ equals to a multiple of δ for each break x_i of $\hat{z}$. Therefore, $f(\hat{z}(x))$ is piecewise linear convex with breaks at multiples of δ and, consequently,

$$\phi_{M-i}(x) = E_{M-1}p_{M-i}x + f(\hat{z}(x)) \qquad (3.56)$$

is piece-wise linear convex with breaks at multiples of δ.

By repeating the induction step we get that ϕ_2 is piece-wise linear with breaks at multiples of δ. Since, according to our requirements, 1 is a multiple of δ, ϕ_2 is linear on $[1-\delta, 1]$ so that (one of the) optimal solutions of $\min_{0\leq y\leq\delta} E[p_1 y + \phi_2(1-y)]$ happens either at $x_1 = 0$ or at $x_1 = \delta$. Therefore, (one of the) solution(s) of the second stage problem is $x_2 = 0$ or $x_2 = \delta$ (it is because ϕ_3 is linear on $[1-2\delta, 1-\delta]$). By repeating the step we get that each interior problem has (one of the) solution(s) at 0 or at δ. Hence, the problem (3.52) is equivalent to the problem (3.53).

3.3.3 The Illustration

In our illustration, we put

$$\delta = \begin{cases} 1 & \text{if } M = 1 \\ 1/2 & \text{if } M > 1 \end{cases},$$

$$p_1 = 1 + U_1$$

and

$$p_i = 1 + \sum_{j=1}^{i-1} \sin((2+j)p_j) + U_i, \quad i = 2, 3, \ldots, M$$

where U_i, $i = 1, 2, \ldots, M$ are i.i.d. variables each having the uniform distribution on $[0, 1]$.

Since the optimality gap of (3.52) can hardly be computed exactly, we had to approximate it: the quantity

$$\hat{z}_M := E_P g_M(\dot{x}_1, \dot{x}_2(\dot{x}_1), \ldots, \dot{x}_M(\dot{x}_1, \ldots, \dot{x}_{M-1}))$$

was computed using the discretization with $M_n = 50$ while

$$\bar{z}_{M,n} := E_P g_M(\hat{x}_1, \hat{x}_2(\hat{x}_1), \ldots, \hat{x}_M(\hat{x}_1, \ldots, \hat{x}_{M-1}))$$

was approximated using MC with sample size 10000 (we recall that the optimality gap equals to $\bar{z}_{M,n} - \hat{z}_M$, see (3.11)).

The following tables show results for the three methods, examined in this work, i. e. discretization (the optimality denoted by η_n, we have approximated the exact distribution by a grid discretization at each stage), Monte Carlo (the expected optimality gap is denoted by $\tilde{\eta}_n = \tilde{\eta}_{\sqrt[M]{n},\ldots,\sqrt[M]{n}}$) and Quasi Monte Carlo (the expected optimality gap is denoted by $\acute{\eta}_n$) for various numbers of stages M. The numbers typed by the small font show estimated standard errors of the MC computation of the optimality gap.

M=1			
n	η_n	$\tilde{\eta}_n$	$\acute{\eta}_n$
4	0.0007 $_{0.0029}$	0.0448 $_{0.0044}$	0.0368 $_{0.0042}$
49	0.0017 $_{0.0029}$	0.0449 $_{0.0042}$	-0.0032 $_{0.0029}$
144	0.0032 $_{0.0029}$	0.0353 $_{0.0041}$	0.0034 $_{0.0029}$
289	-0.0001 $_{0.0029}$	0.0342 $_{0.004}$	-0.0011 $_{0.0029}$
484	-0.0004 $_{0.0029}$	0.0285 $_{0.0039}$	0.0013 $_{0.0029}$
729	-0.0012 $_{0.0029}$	0.0238 $_{0.0038}$	0.0048 $_{0.0029}$
1024	0.0061 $_{0.0029}$	0.0303 $_{0.0038}$	-0.0007 $_{0.0029}$
1369	0.0029 $_{0.0029}$	0.0204 $_{0.0037}$	0.0041 $_{0.0029}$

M=2 n	η_n	$\tilde{\eta}_n$	$\acute{\eta}_n$
8	0.0021 0.0018	0.0431 0.0022	0.0156 0.0019
27	0.0005 0.0018	0.0312 0.0021	7.2424e-06 0.0018
64	-0.0003 0.0017	0.03 0.0021	-0.0016 0.0018
125	0.0003 0.0018	0.0191 0.002	-0.0018 0.0018
216	-0.0015 0.0018	0.0177 0.0019	0.0012 0.0018
343	0.0013 0.0018	0.0104 0.0019	-0.0037 0.0018
512	-0.0021 0.0018	0.0081 0.0019	-0.0017 0.0018
729	0.0026 0.0018	0.008 0.0019	0.0012 0.0018
1000	0.0007 0.0018	0.0082 0.0018	0.0009 0.0018
1331	-0.0009 0.0018	0.0038 0.0018	-0.0011 0.0018
1728	0.001 0.0018	0.0061 0.0018	-0.0011 0.0018
2197	0.0013 0.0018	0.0047 0.0018	0.0005 0.0018
2744	0.0005 0.0018	0.0065 0.0018	-0.0009 0.0018
3375	0.004 0.0018	0.0006 0.0018	-0.0022 0.0018
4096	-0.0001 0.0018	0.0015 0.0018	0.0018 0.0018
4913	-0.0005 0.0018	0.0023 0.0018	-0.0026 0.0018
5832	0.0002 0.0018	0.0028 0.0018	-0.0001 0.0018
6859	0.0012 0.0018	0.0044 0.0018	0.0011 0.0018
8000	-0.0004 0.0018	0.0015 0.0018	-0.0003 0.0018
9261	0.0009 0.0018	0.0038 0.0018	-0.004 0.0018
10648	0.0018 0.0018	0.0029 0.0018	0.0034 0.0018
12167	-0.0012 0.0018	0.0025 0.0018	0.002 0.0018
13824	0.0034 0.0018	0.0028 0.0018	-0.0004 0.0018
15625	-0.0015 0.0018	0.0023 0.0018	0.0003 0.0018
17576	-0.0021 0.0018	-0.0009 0.0018	-0.0017 0.0018
19683	0.002 0.0018	-6.6383e-05 0.0018	-0.0012 0.0018
21952	-0.0008 0.0018	0.0023 0.0018	-0.0027 0.0018
24389	-0.0014 0.0018	0.0021 0.0018	-0.0007 0.0018
27000	0.0013 0.0018	-0.0001 0.0018	-0.0015 0.0018
29791	0.0015 0.0018	0.0029 0.0018	0.001 0.0018
32768	0.0021 0.0018	-0.0017 0.0018	0.0008 0.0018
35937	-0.0014 0.0018	-0.0024 0.0018	-0.0015 0.0018
39304	-0.0027 0.0018	0.0001 0.0018	-0.0011 0.0018
42875	0.0008 0.0018	-0.0002 0.0018	0.0005 0.0018
46656	-0.0004 0.0018	-0.0002 0.0018	-0.0014 0.0018
50653	-8.5972e-05 0.0018	0.0017 0.0018	-0.0019 0.0018
54872	-0.0024 0.0018	0.0012 0.0018	0.0002 0.0018
59319	-0.0009 0.0018	0.0027 0.0018	-0.002 0.0018
64000	-0.0027 0.0018	0.002 0.0018	-0.0032 0.0018

M=3 n	η_n	$\tilde{\eta}_n$	$\acute{\eta}_n$
16	0.0011 0.0017	0.0331 0.002	0.0157 0.002
81	-0.0015 0.0017	0.0252 0.002	0.0074 0.0019
256	0.0004 0.0017	0.0174 0.002	0.0057 0.0018
625	-0.002 0.0017	0.02 0.002	0.0034 0.0018
1296	0.001 0.0017	0.02 0.0019	-0.0005 0.0018
2401	-0.0003 0.0017	0.0196 0.0019	0.0004 0.0017
4096	0.0001 0.0017	0.0155 0.0019	0.0028 0.0017
6561	0.0009 0.0017	0.0163 0.0019	-0.0013 0.0017
10000	-0.0018 0.0017	0.0155 0.0019	-0.0004 0.0017
14641	-0.0014 0.0017	0.0095 0.0019	0.0021 0.0017
20736	-0.0035 0.0017	0.0135 0.0019	-0.0008 0.0017
28561	0.0012 0.0017	0.0106 0.0019	0.0017 0.0017
38416	0.0007 0.0017	0.0095 0.0019	-0.0024 0.0017
50625	-0.0025 0.0017	0.0112 0.0019	-0.0017 0.0017
65536	-0.0007 0.0017	0.0149 0.0019	0.0012 0.0017
83521	-0.0035 0.0018	0.0093 0.0019	-0.0033 0.0017
104976	0.0007 0.0017	0.0091 0.0019	0.0036 0.0017
130321	0.0016 0.0017	0.0091 0.0019	0.0026 0.0017
1.6e+05	-0.0018 0.0017	0.0074 0.0019	-0.002 0.0017
1.9448e+05	0.0002 0.0017	0.0077 0.0019	0.0009 0.0017

M=4 n	η_n	$\tilde{\eta}_n$	$\acute{\eta}_n$
32	0.0014 0.0017	0.0333 0.0021	0.0114 0.002
243	0.0008 0.0017	0.0254 0.0021	0.0075 0.002
1024	-0.0022 0.0023	0.0185 0.0021	0.0059 0.002
3125	0.0008 0.0017	0.0193 0.0021	0.0072 0.002
7776	-0.0005 0.0022	0.0179 0.0021	0.0024 0.002
16807	-6.7995e-06 0.0017	0.0163 0.0021	-0.001 0.002
32768	0.0019 0.0017	0.0141 0.0021	0.0042 0.002
59049	0.0015 0.0023	0.0144 0.002	-0.0012 0.002
100000	0.0005 0.0017	0.0116 0.002	-0.0014 0.002

M=5 n	η_n	$\tilde{\eta}_n$	$\acute{\eta}_n$
64	-0.0003 0.0023	0.0427 0.0022	0.0221 0.0021
729	0.0009 0.0022	0.0388 0.0021	0.0091 0.0021
4096	-9.2547e-05 0.0022	0.0313 0.0021	0.0062 0.0021
15625	0.0038 0.0023	0.0253 0.0021	0.0057 0.0021
46656	0.0007 0.0023	0.026 0.0021	0.0053 0.0022

To estimate the actual rate of convergence $\bar{\eta}$ we have used the logarithmic regression $\bar{\eta} = \beta n^{\alpha}$(to be able to do so, we have treated non-positive values as outliers, i.e. we did not include them into the regression[5]). The results are shown in the following table - symbol s_α denotes standard error of the regressor α.

[5]Of course, the problem of negative values can be solved by improving accuracy. However, the computation

M	discr.		MC		QMC	
	α	s_α	α	s_α	α	s_α
1	-7.56	0.44	-2.79	0.17	-3.02	0.99
2	-7.29	0.62	-2.1	0.45	-7.17	1.43
3	-7.25	0.95	-3	0.13	-4.22	0.68
4	-6.75	0.68	-3.04	0.09	-3.91	0.49
5	-7.11	3.85	-2.76	0.12	-3.1	0.33

The estimated rate of convergence of each method is $O(n^\alpha)$.

The results are indicating that the rates of convergence, predicted by Theorem 3.2. by Theorem 3.5. and by Theorem 3.6., are much outperformed by our test problem: it is probably thanks to a "smooth" structure of the problem. The most remarkable is the fact that the convergence rates seem to be independent on the number of stages M.

took more than. 10 hours and to improve the accuracy by one digit we would have to enlarge the sample 100 times...

Chapter 4

Conclusion

The main goal of the present work was to study and compare various techniques of the stochastic programming problem's approximation, namely the discretization, the Monte Carlo and the Quasi Monte Carlo (in sense of subsection 1.2.3).

We have formulated the upper bounds of the discretization and the Monte Carlo approximation error (measured by the optimality gap) for both the one-stage and the multistage programming. However, both they are probably very conservative so that they could not be used for the comparison.

Then we have studied the convergence rate of the approximation error at one-stage stochastic programming. As to the discretization error, we have proved that if the value function is uniformly Lipschitz in the random argument then the approximation error is $O(n^{1/k(-1+1/a)})$ where n is the number of atoms, k is the dimension of the random argument of the objective function and $-a$ is the convergence rate of the marginal distributions' tails. Further, we have shown that the Monte Carlo approximation error is $O(n^{-1/2})$ under the assumption that the objective function is $\mathcal{L}_2$-Lipschitz in the decision variable and that the feasibility set is compact. Finally, we have concluded that the Quasi Monte Carlo error is $o\left(n^{-1/2-(1-b)/k+\delta}\right)$ for each $\delta > 0$ where b is a constant dependent on the distribution. The sufficient conditions for this result were that the value function is both uniformly Lipschitz both in the random argument and the decision variable and that the feasibility set is compact. It is obvious that the QMC is no worse then the Monte Carlo and that the Monte Carlo is better then the discretization when $k > 2$ (the case $k \leq 2$ depends on the constant a). However, the result of the QMC has stronger assumptions then that of the Monte Carlo and the assumptions of the Monte Carlo assertion are different from those of the "discretization" assertion. Moreover, it is very important to stress here that we are comparing upper bounds of the errors, not the errors themselves. The opinion of the author is that the bound of the discretization error is much more conservative then those of the MC and of the QMC.

In the case of multistage problems, we found that the discretization error is $O\left(n^{1/k(-1+1/a}\right)$ provided that all the functions defining the problem are uniformly Lipschitz in the chance parameter (a has a meaning similar to the preceding paragraph, k is the total number of the random parameter's dimension). Further, we have found that the MC error is $o_P\left(n^{-\frac{1}{2M}+\delta}\right)$ given that the functions defining the problem are uniformly Lipschitz in decision variable, that their difference with their conditional expectations are bounded and that the union of the feasibility sets of all the (sub)problems in compact (M denotes the number of the stages of the problem, $\delta > 0$ is an arbitrarily small constant). Further, we have constructed an example of an M-stage MSSPP with the functions defining the problem being uniformly Lipschitz both in the

"chance" parameter and in the decision variable and with the expected Monte Carlo optimality gap being $O(n^{-1/M})$. Hence, if the chance parameter is one-dimensional and bounded at each stage, then the convergence rate, guaranteed under the given assumptions, is no worse for the discretization then for the Monte Carlo. The case of unbounded and/or more-dimensional random parameter depends on the constants M, k and a. Again we have to say, however, that our results ar only the "worst" bounds and that the actual results may be better. Finally, we have estimated the convergence rate of the Quasi Monte Carlo optimality gap. If the functions defining the problem are uniformly Lipschitz both in the "chance" parameter and in the decision variable and their difference with their conditional expectation is bounded, if the union of the feasibility sets is compact subset of an Euclidean space and if the random parameter at each stage is a κ-dimensional real vector, then the QMC optimality gap is $o_P(n^{-(2-2a+\kappa)/(2M\kappa)+\delta})$ where a has a similar meaning as in one-stage case. Therefore, the QMC is no worse that MC under the given assumptions.

Hence, the comparison has no definite winner. However, if the author should solve a one stage SPP then he would choose the QMC or, if the strong assumptions of the QMC result were not fulfilled, he would use the discretization.

In addition to the results mentioned above, some other ones were achieved in the present work:

- a lower bound of the Wasserstein distance of an absolutely continuous distribution and a discrete distribution (Theorem 1.2.),
- a probabilistic upper bound of the MC estimate of the integral (Theorem 1.5.),
- the definition of the optimality gap for multistage problems (subsections 3.2.1 and 3.2.2), its decomposition (Lemma 3.1., Lemma 3.2., Lemma 3.3.) and its upper bounds (Theorem 3.1., Theorem 3.4.).

Appendix A

Some Quoted Results

In the Appendix, some results, used in the present work, are cited.

Theorem A.1. *([3], Lemma 3.1.)* Let Y be such random variable that $E\exp\{\theta Y\}$ exists and is finite for $0 \leq \theta \leq \theta_0$ where $\theta_0 > 0$. Then

$$P[Y \geq y] \leq E\exp\{\theta Y\} \times \exp\{-\theta y\}$$

for each $0 \leq \theta \leq \theta_0$.

Theorem A.2. *(Continuous Mapping Theorem, [37], p.175)* Let T be continuous map from $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ to $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ where $\mathcal{X}$ and $\mathcal{Y}$ are metric spaces and $\mathcal{B}$ denotes Borel sets. If $\mathcal{P}$, $\mathcal{P}_1$,$\mathcal{P}_2$,... are probability measures defined on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that $\mathcal{P}_n \to \mathcal{P}$ in distribution then $T(\mathcal{P}_n) \to T(\mathcal{P})$ in distribution.

Theorem A.3. *(Central Limit Theorem for $\mathcal{C}(\mathcal{X})$, [28], Appdix, Prop. A5).* Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and let g be a measurable mapping from Ω into $\mathcal{C}(K)$ where K is a compact subset of R^m. Suppose that g satisfies the $\mathcal{L}^2$-Lipschitz condition on K, i.e. there exists a in $\mathcal{L}^2(\Omega, \mathcal{P})$ such that

$$|g(x_1) - g(x_2)| \leq a\|x_1 - x_2\|_2$$

for all $x_1, x_2 \in K$. Suppose also that $E|g(x)|^2 < \infty$ for some $x \in K$. Let g_n, $n = 1, 2, \ldots$, be independent, identically distributed observations of g. Then there exists a Gaussian random variable Ξ taking values in $\mathcal{C}(K)$ such that

$$\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N} g_n - Eg\right] \xrightarrow{\mathcal{D}} \Xi$$

and where for all $x \in K$,

$$\Xi(x) \sim N(0, \mathrm{cov}(g(x))).$$

Theorem A.4. *([44], "Bennet's Inequality 3.", pp. 951-2,[1] and Proposition 1., p. 441)* Let $X_1, X_2, \ldots, X_n$ be independent random variables with $X_i \leq b$, $EX_i = 0$ and $\mathrm{var}\, X_i = \sigma_i^2$, for $i = 1, 2, \ldots, n$. Let $\sigma^2 = (\sigma_1^2 + \ldots + \sigma_n^2)/n$ and $\bar{X} = (X_1 + \ldots + X_n)/n$. Then for all $\lambda \geq 0$

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

[1] There is a misprint in the assertion. The correct version follows from the proof.

where

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

The function ψ has the following properties:

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

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

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

Bibliography

[1] J. Anděl. *Matematical Statistics (in Czech)*. SNTL, Ptague, 1985.

[2] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.

[3] L. Dai, C. H. Chen, and J. R. Birge. Convergence properties of two-stage stochastic programming. *Journal of Optimization Theory and Applications*, 108(3):489–509, 2000.

[4] G. B. Dantzig. Planning under uncertainty. *Annals of Operations Research*, 85:preface, 1999.

[5] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration*. Academic Press, New York, 1975.

[6] M. A. H. Dempster and G. W. P. Thompson. Dynamic portfolio replication using stochastic programming. In M. Dempster, editor, *Risk Management: Value at Risk and Beyond*, pages 100–128. Cambridge University Press, Cambridge, 2002.

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

[8] J. Dupačová. Multistage stochastic programs: the state-of-the-art and selected bibliography. *Kybernetika (Prague)*, 31(2):151–174, 1995.

[9] J. Dupačová, N. Gröve-Kuska, and V. Römisch. Scenario reducttion in stochastic programming, an approach using probability metrics. *Mathematical Programmming ser. A*, 95:493–511, 2003.

[10] J. Dupačová, J. Hurt, and J. Štěpán. *Stochastic Modelling in Economics and Finance*. Kluwer, Dodrecht, 2002.

[11] J. Dupačová and W. Römisch. Quantitative stability for scenario-based stochastic programs. In *Prague Stochastics 98*, pages 119–124. JCMF, 1998.

[12] J. Dupačová and R. Wets. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. statist.*, 16:1517–1549, 1988.

[13] Y. Ermoliev. Stochastic quasigradient methods. In Ermoliev Y and Wets R. J.-B, editors, *Numerical Techniques for Stochastic Optimization*, pages 33–64. Springer, Berlin, 1988.

[14] J. Hoffmann-Jørgensen. *Probability with a View Towards to Statistics I.* Chapman and Hall, New York, 1994.

[15] P. Kall. *Stochastic Linear Programming.* Springer, Berlin, 1976.

[16] P. Kall, P. Ruszczyński, and K. Frauendorfer. Approximation techniques in stochastic programming. In Ermoliev Y and Wets R. J.-B, editors, *Numerical Techniques for Stochastic Optimization*, pages 33–64. Springer, Berlin, 1988.

[17] P. Kall and S. W. Wallace. *Stochastic Programming.* Willey, Chichester, 1994.

[18] Yu. M. Kaniovski, A. J. King, and R. J.-B. Wets. Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems. *Ann. Oper. Res.*, 56:189–208, 1995.

[19] V. Kaňková. Optimum solution of a stochastic programming problem witth unknown parameters. In *Transactions of the Seventh Prague Conference 1974*, pages 234–244, Praha, 1977. Academia.

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

[21] V. Kaňková. Approximative solution of problems of two-stage stochastic nonlinear programming (in czech). *Ekonomicko-matematický obzor*, 16:64–76, 1980.

[22] V. Kaňková. On distribution stability in stochastic programming. Technical Report 1826, Instiute of Information Theory and Automation, Czech Acad. Sci., 1994.

[23] V. Kaňková. A note on extimates in stochastic programming. *Journal of Computational and Applied Mathematic*, 56:97–112, 1995.

[24] V. Kaňková. A note on multistage stochastic programming. In *Proceedings of the 11th Joint Czech-German-Slovak Conference "Mathematical Methods in Economy and Industry". Liberec (Czech Republic) 1998*, pages 45–52, Liberec, 1999. Technická univerzita Liberec, Vysokoškolský podnik s.r.o.

[25] V. Kaňková. Multistage stochastic programming: Stability, approximation and markov dependence. In *Operations Research Proceedings 1999*, pages 136–141, Berlin, 2000. Springer.

[26] V. Kaňková. A remark on empirical estimates in multistage stochastic programming. *Bulletin of the Czech Econometric Society*, 17:32–51, 2002.

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

[28] A. J. King. *Asymptotic Behavior of Solutions in Stochastic Optimization: Nonsmooth Analysis And the Derivation of Non-Normal Limit Distributions.* PhD thesis, University of Washington, 1986.

[29] J. Lukeš. *Notes on Functional Analysis (in Czech).* Karolinum, Prague, 1998.

[30] J. Lukeš and J. Malý. *Measure and Integral.* Matfyzpress, Prague, 1995.

[31] W. K. Mak, D. P. Morton, and R. K. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24:47–56, 1999.

[32] D. Morton and E. Popova. Monte Carlo simulations for stochastic optimization. In P.M. Pardalos C.A. Floudas, editor, *Encyclopedia of Optimization*. Kluwer Academic Publishers, 2001.

[33] H. Niedereiter. *Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, PA, 1992.

[34] M. P. Nowak and W. Römisch. Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty. *Annals of Operations Research*, 100:251–272, 2000.

[35] T. Pennanen and M. Koivu. Epi-convergent discretizations of stochastic programs via integration quadratures. Preprint published at http://www.speps.info as article no. 16/2003, 2003.

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

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

[38] A. Prékopa. *Stochastic Programming*. Kluwer and Academiai Kiado, Dordrecht and Budapest, 1995.

[39] S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models*. J. Wiley and Sons, New York, 1991.

[40] A. Ruszczyński and A. Shapiro, editors. *Stochastic Programming*. Elsevier, Amsterodam, 2003.

[41] A. Shapiro. Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30:169–186, 1991.

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

[43] A. Shapiro. Statistical inference of multistage stochastic programming problems. *Optimization Online (www.optimization-online.org)*, 2002.

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

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

[46] S. S. Vallander. Calculation of the Wasserstein distance between probability distributions on the line. *Theor. Probab. Appl.*, 18:784–786, 1973.

[47] M. H. van der Vlerk. Stochastic programming bibliography. World Wide Web, http://mally.eco.rug.nl/spbib.html, 1996-2003.

[48] R. J.-B. Wets. A statistical approach to the solution of stochastic programs with (convex) simple recourse. *Research report, Univ. Kentucky*, 1979.