Ann Oper Res
DOI 10.1007/s10479-008-0355-9

The Expected loss in the discretization of multistage stochastic programming problems—estimation and convergence rate

Martin Šmíd

Abstract In the present paper, the approximate computation of a multistage stochastic programming problem (MSSPP) is studied. First, the MSSPP and its discretization are defined. Second, the expected loss caused by the usage of the "approximate" solution instead of the "exact" one is studied. Third, new results concerning approximate computation of expectations are presented. Finally, the main results of the paper—an upper bound of the expected loss and an estimate of the convergence rate of the expected loss—are stated.

Keywords Multistage stochastic programming problems · Approximation · Discretization · Monte Carlo

1 Introduction

In practice, many decision problems may be realistically modeled by means of the multistage stochastic programming models (see Dupačová et al. 2002, Chap. II, or Ruszczyński and Shapiro 2003 for examples). Unfortunately, the realistic models often lead to unsolvable optimization problems, hence an approximation has to be done.

1.1 Expected loss

Usually, the distance between the optimal values of the exact and the approximate problem is used to measure the error of the approximation. Contrary to this approach, we use the *expected loss*, i.e. the mean difference between the value function with the "approximate" solution as argument and the optimal value.

The expected loss is a very natural measure of the approximation error since it measures the error in the "units" of the original optimization problem. If the costs are minimized, for instance, the expected loss can be interpreted as the price paid for the approximation.

M. Šmíd (✉)
Institute of Information Theory and Automation of the ASCR, Pod Vodárenskou věží 4, 182 08 Prague, Czech Republic
e-mail: smid@utia.cas.cz

There are several papers devoted to the expected loss in (one stage) stochastic programming. Let us mention Mak et al. (1999) suggesting estimates of the expected loss based on the Monte Carlo methods or Pflug (2001) devoted to the minimization of the expected loss by means of minimizing the Wasserstein distance.

In the present article, we are linking up to the article of Pflug (2001) which is based on the following idea: If we are approximating a one-stage problem

$$\min_{x\in\mathcal{X}} \mathbb{E}_{P(\omega)} h(x,\omega), \tag{1}$$

where $h(x,\omega)$ is uniformly Lipschitz in ω, by

$$\min_{x\in\mathcal{X}} \mathbb{E}_{Q(\omega)} h(x,\omega), \tag{2}$$

where Q is an "approximate" probability distribution, then the expected loss is

$$\eta_Q^1 := \mathbb{E}_{P(\omega)} h(\tilde{x},\omega) - \mathbb{E}_{P(\omega)} h(\hat{x},\omega), \tag{3}$$

where $\hat{x}$ is an optimal solution of (1) and $\tilde{x}$ is an optimal solution of (2). Even if η_Q^1 cannot be computed exactly due to its dependence on $\hat{x}$, it may be easily estimated:

$$\begin{aligned}
\eta_Q^1 &= \mathbb{E}_{P(\omega)} h(\tilde{x},\omega) - \mathbb{E}_{Q(\omega)} h(\tilde{x},\omega) + \mathbb{E}_{Q(\omega)} h(\tilde{x},\omega) - \mathbb{E}_{P(\omega)} h(\hat{x},\omega) \\
&\le \mathbb{E}_{P(\omega)} h(\tilde{x},\omega) - \mathbb{E}_{Q(\omega)} h(\tilde{x},\omega) + \mathbb{E}_{Q(\omega)} h(\hat{x},\omega) - \mathbb{E}_{P(\omega)} h(\hat{x},\omega) \\
&\le 2 \sup_{x\in\mathcal{X}} |\mathbb{E}_{P(\omega)} h(x,\omega) - \mathbb{E}_{Q(\omega)} h(x,\omega)| \le 2K d_W(P,Q)
\end{aligned} \tag{4}$$

where d_W is the Wasserstein distance and K is the Lipschitz constant of h (see Pflug 2001 for details).

Unfortunately, this result cannot be easily adapted to the "multistage" situation. The main reason is that the "approximate" solutions need not be feasible to the "exact" problem (see Sect. 2 of the present article for an explanation). Therefore, the "multistage" expected loss has to be defined differently from the "one-stage" one and new bounds have to be derived to estimate it.

1.2 Discretization

There could be more ways how to approximate the multistage stochastic programming problems, let us mention the Monte Carlo estimation (see Ruszczyński and Shapiro 2003, Chap. 6, for instance) or sequential approximation techniques (see Kall et al. 1988 for the two-stage case); in the present paper, however, we concentrate on the simple discretization, i.e. the replacement of the "exact" probability distribution by a discrete one.

Great attention has been already paid to the discretization of the (multistage) stochastic programming problems (let us mention Dupačová et al. 2003; Kaňková 2000; Kaňková and Šmíd 2004 or Pennanen and Koivu 2005). The main contribution of the present article, in addition to the use of the expected loss instead of the distance of the optimal values, is that weaker assumptions are required from the data process: it must neither be bounded (as in Kaňková 2000) nor have the Markov structure with the exponential type densities (as in Kaňková and Šmíd 2004) and it may be continuous (contrary to Dupačová et al. 2003); our only assumption is that the tails of marginal distributions converge at a "reasonable" rate.

2 Approximation of MSSPP problems

The definitions of the multistage stochastic programming problem vary in the literature. In the present article, we define the MSSPP equivalently to the recursive definition by Ruszczyński and Shapiro (2003) (Chap. 1, (3.17)):

Definition 1 Suppose that

$$M \in \mathbb{N}, \qquad d_1, d_2, \ldots, d_M \in \mathbb{N}, \qquad k_1, k_2, \ldots, k_M \in \mathbb{N},$$

$$g_M : \mathbb{R}^{d_1+d_2+\cdots+d_M} \times \mathbb{R}^{k_1+k_2+\cdots+k_M} \longrightarrow \mathbb{R},$$

$\xi_1 \in \mathbb{R}^{k_1}$ is a constant,[1]

$\xi_2 \in \mathbb{R}^{k_2}, \xi_3 \in \mathbb{R}^{k_3}, \ldots, \xi_M \in \mathbb{R}^{k_M}$ are random vectors,

and that

$$\mathcal{X}_i : \mathbb{R}^{d_1+d_2+\cdots+d_{i-1}} \times \mathbb{R}^{k_1+k_2+\cdots+k_i} \longrightarrow 2^{\mathbb{R}^{d_i}}, \quad i = 1, 2, \ldots, M,$$

are point-to-set mappings.

If we put $\bar{\xi}_i := (\xi_1, \xi_2, \ldots, \xi_i)$ and $\bar{x}_i := (x_1, x_2, \ldots, x_i)$, we may define the ($M$-stage) multistage stochastic programming problem (MSSPP) as

$$\min_{x_1 \in \mathcal{X}_1(\xi_1)} \mathbb{E} g_1(x_1, \bar{\xi}_2), \tag{5}$$

where

$$g_i(\bar{x}_i, \bar{\xi}_{i+1}) = \min_{x_{i+1} \in \mathcal{X}_{i+1}(\bar{x}_i, \bar{\xi}_{i+1})} \mathbb{E}(g_{i+1}(\bar{x}_{i+1}, \bar{\xi}_{i+2}) | \bar{\xi}_{i+1}) \tag{6}$$

for each $i = 1, 2, \ldots, M-2$ and

$$g_{M-1}(\bar{x}_{M-1}, \bar{\xi}_M) = \min_{x_M \in \mathcal{X}_M(\bar{x}_{M-1}, \bar{\xi}_M)} g_M(\bar{x}_M, \bar{\xi}_M). \tag{7}$$

In case that we are unable to evaluate the (conditional) expectations in (5) or in (6) exactly, an approximation has to be done. Usually, we replace $\bar{\xi}_M$ by a discrete random vector $\bar{\zeta}_M := (\zeta_1, \zeta_2, \ldots, \zeta_M)$ with a "tractable" number of atoms[2] which leads to the problem

$$\min_{x_1 \in \mathcal{X}_1(\zeta_1)} \mathbb{E} \tilde{g}_1(x_1, \bar{\zeta}_2), \tag{8}$$

where

$$\tilde{g}_i(\bar{x}_i, \bar{\zeta}_{i+1}) = \min_{x_{i+1} \in \mathcal{X}_{i+1}(\bar{x}_i, \bar{\zeta}_{i+1})} \mathbb{E}(\tilde{g}_{i+1}(\bar{x}_{i+1}, \bar{\zeta}_{i+2}) | \bar{\zeta}_{i+1})$$

[1] Even if this constant has no function in the present article, we include it for the sake of compatibility with the widely respected source.

[2] For convenience in notation, we take ξ_1 as a random vector with the Dirac probability measure and we assume that $\zeta_1 = \xi_1$.

for each $i = 1, 2, \dots, M-2$ and

$$\tilde{g}_{M-1}(\bar{x}_{M-1}, \bar{\zeta}_M) = \min_{x_M \in \mathcal{X}_M(\bar{x}_{M-1}, \bar{\zeta}_M)} g_M(\bar{x}_M, \bar{\zeta}_M).$$

In the sequel, we shall suppose the distribution of $\bar{\zeta}_M$ to possess a tree structure, i.e. we assume that there exist $n_2 \in \mathbb{N}, \dots, n_M \in \mathbb{N}$ such that the number of atoms of $\mathcal{L}(\zeta_i|\bar{z}_{i-1})$ is n_i for each atom $\bar{z}_{i-1}$ of $\bar{\zeta}_{i-1}$, $i = 2, 3, \dots, M$.

To make our analysis of the MSSPP approximation complete, we have to deal with one more problem (reported by Dempster and Thompson 2002, for instance) which is the possible infeasibility of the interior solutions: Suppose that we have applied the "approximate" first stage decision $\tilde{x}_1$ (obtained from (8)) and that the second stage random parameter has realized itself as $\bar{\xi}_2$. In this situation, the second stage approximate solution $\tilde{x}_2$ need not be feasible to the exact problem since the second stage feasibility set depends on ξ_2, which may possibly realize itself outside the set of the atoms of ζ_2. Obviously, we are facing analogous problems at the stages $3, 4, \dots, M$ as well.

Hence, for a successful application of the MSSPP, it does not suffice to approximate (5) by (8) but, in addition, one has to be able to solve (at least approximately) a (parametric multistage) stochastic programming problem (6) for each $1 \le i \le M-2$ and each feasible $\bar{x}_i = \bar{x}_i(\bar{\xi}_i)$ (i.e. such that $x_1 \in \mathcal{X}_1(\xi_1), x_2 \in \mathcal{X}_2(x_1, \bar{\xi}_2), \dots, x_i \in \mathcal{X}_i(\bar{x}_{i-1}, \bar{\xi}_i)$). Clearly, the problems (6) may contain non-computable expectations, hence they also have to be approximated.

There could be more ways how to approximate (6): we can take the "nearest" feasible solution of (8), for instance. However, the most natural approach is to make a new approximation of (6).

Obviously, if we want to do so, we have to be able to approximate each conditional distribution $\mathcal{L}(\bar{\xi}_i|\bar{\xi}_{i-1})$ for each $3 \le i \le M$, i.e. to construct a mapping

$$\Pi_i : \operatorname{supt}(\bar{\xi}_{i-1}) \to \mathcal{S}_{d_i, n_i}, \tag{9}$$

where $\mathcal{S}_{d_i,n_i}$ is the space of n_i-atom discrete distributions on $\mathbb{R}^{d_i}$, for each $i = 3, 4, \dots, M$.

Definition 2 In the sequel, an ordered set $(\Pi_2, \Pi_3, \dots, \Pi_M)$, where $\Pi_2 = \Pi_2(\xi_1)$ is an n_2-atom discrete probability distribution on $\operatorname{supt}(\xi_2)$ and where Π_i, $i = 3, 4, \dots, M$, are mappings of the type (9), will be called approximation scheme with the dimensions $n_2, n_3, \dots, n_M$ (we shall write $\Pi_i(\varsigma_i|\bar{\varsigma})$ instead of $(\Pi_i(\bar{\varsigma}))(\varsigma_i)$ for the sake of readability). The quantity $\prod_{i=2}^k n_i$ will be called number of atoms of the approximation scheme Π.

Having an approximating scheme Π, we may approximate each conditional expectation

$$\mathbb{E}(h(\bar{\xi}_{i+1})|\bar{\xi}_i),$$

where h is a function of $\bar{\xi}_{i+1}$, by

$$\mathbb{E}_{\Pi_{i+1}(z_{i+1}|\bar{\xi}_i)} h(\bar{\xi}_i, z_{i+1})$$

which may simply be rewritten as

$$\mathbb{E}(h(\bar{\xi}_i, \zeta^i_{i+1})|\bar{\xi}_i)$$

where ζ^i_{i+1} is a random vector defined by $\Pi_{i+1}(\bar{\xi}_i)$ (i.e. such that $\mathcal{L}(\zeta^i_{i+1}|\bar{\xi}_i) = \Pi_{i+1}(\bar{\xi}_i)$ a.s.).

Using this reformulation, we may write the problem approximating (6) as

$$\tilde{g}_i(\bar{x}_i, \bar{\xi}_{i+1}) = \min_{x_{i+1} \in \mathcal{X}_{i+1}(\bar{x}_i, \bar{\xi}_{i+1})} E(\tilde{g}_{i+1}(\bar{x}_{i+1}, \bar{\xi}_{i+1}, \zeta_{i+2}^{i+1}) | \bar{\xi}_{i+1}). \tag{10}$$

To avoid technical problems, we assume throughout that:

(a) the problem (5) is well defined, i.e.
  (i) $\mathbb{E} g_1(x_1, \bar{\xi}_2)$ exists and is finite for each feasible x_1 and the minimum in (5) exists,
  (ii) $\mathbb{E}(g_{i+1}(\bar{x}_{i+1}, \bar{\xi}_{i+2}) | \bar{\xi}_{i+1})$ exists and is finite for each feasible $\bar{x}_{i+1}$ and the minimum in (6) exists almost surely for each feasible $\bar{x}_i$, $i = 1, 2, \ldots, M-2$,
  (iii) the minimum in (7) exists almost surely.
(b) the problem (8) is well defined for each discrete random vector ζ defined on $\mathrm{supt}(\bar{\xi}_M)$.

3 Expected loss in MSSPP

Let us define a multistage counterpart of the one-stage expected loss now. Denote

$\hat{x}_1(\xi_1)$ (an arbitrary) solution of (5),
$\tilde{x}_1(\xi_1)$ (an arbitrary) solution of (8),
$\hat{x}_{i+1}(\bar{x}_i) = \hat{x}_{i+1}(\bar{x}_i, \bar{\xi}_{i+1})$ (an arbitrary) solution of (6) with the parameters i, $\bar{x}_i$ and $\bar{\xi}_{i+1}$,
$\tilde{x}_{i+1}(\bar{x}_i) = \tilde{x}_{i+1}(\bar{x}_i, \bar{\xi}_{i+1})$ (an arbitrary) solution of (10) with the parameters i, $\bar{x}_i$ and $\bar{\xi}_{i+1}$.

If we were able to solve (5) exactly then the sequence of our decisions would be

$$\hat{x}_1(\xi_1), \hat{x}_2(\hat{x}_1, \bar{\xi}_2), \ldots, \hat{x}_M(\hat{x}_1, \ldots, \hat{x}_{M-1}, \bar{\xi}_M);$$

however, since we are approximating (5) by (8) and (6) by (10) for each $i = 2, 3, \ldots, M-2$, our decisions are

$$\tilde{x}_1(\xi_1), \tilde{x}_2(\tilde{x}_1, \bar{\xi}_2), \ldots, \hat{x}_M(\tilde{x}_1, \ldots, \tilde{x}_{M-1}, \bar{\xi}_M)$$

(we assume that we are able to solve (7) exactly because there is no expectation there).

Definition 3 We define the expected loss of the approximation of a MSSPP by the scheme $\Pi = (\Pi_2, \Pi_3, \ldots, \Pi_M)$ as

$$\begin{aligned} \eta_\Pi := &\mathbb{E} g_M(\tilde{x}_1, \tilde{x}_2(\tilde{x}_1), \ldots, \hat{x}_M(\tilde{x}_1, \ldots, \tilde{x}_{M-1}), \bar{\xi}_M) \\ &- \mathbb{E} g_M(\hat{x}_1, \hat{x}_2(\hat{x}_1), \ldots, \hat{x}_M(\hat{x}_1, \ldots, \hat{x}_{M-1}), \bar{\xi}_M). \end{aligned} \tag{11}$$

The interpretation of the "multistage" expected loss is analogous to the "one-stage" one: if the costs are minimized, for instance, η_Π measures the expected amount of money lost due to the approximation.

The multistage expected loss may be decomposed into the sum of the one-stage losses of the nested problems as follows:

Lemma 1 *Denote*

$$\begin{aligned} \eta_i(\bar{\xi}_{i+1}) := &g_i(\tilde{x}_1, \tilde{x}_2(\tilde{x}_1), \ldots, \tilde{x}_{i-1}(\tilde{x}_1, \ldots, \tilde{x}_{i-2}), \tilde{x}_i(\tilde{x}_1, \ldots, \tilde{x}_{i-1}), \bar{\xi}_{i+1}) \\ &- g_i(\tilde{x}_1, \tilde{x}_2(\tilde{x}_1), \ldots, , \tilde{x}_{i-1}(\tilde{x}_1, \ldots, \tilde{x}_{i-2}), \hat{x}_i(\tilde{x}_1, \ldots, \tilde{x}_{i-1}), \bar{\xi}_{i+1}). \end{aligned} \tag{12}$$

Then

$$\eta_\Pi = \sum_{i=1}^{M-1} \mathbb{E}\eta_i(\bar{\xi}_{i+1}). \tag{13}$$

Proof We may write

$$\begin{aligned}
\eta_\Pi &= \mathbb{E}g_M(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_{M-1}, \hat{x}_M, \bar{\xi}_M) - \mathbb{E}g_M(\tilde{x}_1, \tilde{x}_2, \dots \hat{x}_{M-1}, \hat{x}_M, \bar{\xi}_M) \\
&\quad + \mathbb{E}g_M(\tilde{x}_1, \tilde{x}_2, \dots \hat{x}_{M-1}, \hat{x}_M, \bar{\xi}_M) - \dots - \mathbb{E}g_M(\tilde{x}_1, \hat{x}_2, \dots, \hat{x}_M, \bar{\xi}_M) \\
&\quad + \mathbb{E}g_M(\tilde{x}_1, \hat{x}_2, \dots, \hat{x}_M, \bar{\xi}_M) - \mathbb{E}g_M(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_M, \bar{\xi}_M) \\
&\overset{(6)}{=} \mathbb{E}g_{M-1}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_{M-1}, \bar{\xi}_M) - \mathbb{E}g_{M-1}(\tilde{x}_1, \tilde{x}_2, \dots, \hat{x}_{M-1}, \bar{\xi}_M) + \cdots \\
&\quad + \mathbb{E}g_2(\tilde{x}_1, \tilde{x}_2, \bar{\xi}_3) - \mathbb{E}g_2(\tilde{x}_1, \hat{x}_2, \bar{\xi}_3) + \mathbb{E}g_1(\tilde{x}_1, \bar{\xi}_2) - \mathbb{E}g_1(\hat{x}_1, \bar{\xi}_2). \qquad \square
\end{aligned}$$

Obviously, the evaluation of η_Π depends on the unknown exact solutions, so it has to be estimated. The following Lemma provides an upper bound of η_Π.

Lemma 2 *Let $\Pi = (\Pi_2, \Pi_3, \dots, \Pi_M)$ be an approximation scheme and let there exist constants $b_2, b_3, \dots, b_M$ such that*

$$\left|\mathbb{E}(g_i(\bar{x}_i, \bar{\xi}_i, \xi_{i+1})|\bar{\xi}_i) - \mathbb{E}(g_i(\bar{x}_i, \bar{\xi}_i, \zeta_{i+1}^i)|\bar{\xi}_i)\right| \le b_{i+1} \tag{14}$$

almost surely for each $i = 1, 2, \dots, M-1$ and each feasible $\bar{x}_i$. Then

$$\eta_\Pi \le 2\sum_{i=1}^{M-1} i b_{i+1}.$$

Proof For the sake of better readability, we shall omit the constant ξ_1 and write

$$\begin{aligned}
G_i(\bar{x}_i, \xi_2, \xi_3, \dots, \xi_{i+1}) &:= g_i(\bar{x}_i, \xi_1, \xi_2, \dots, \xi_{i+1}), \\
\tilde{G}_i(\bar{x}_i, \xi_2, \xi_3, \dots, \xi_{i+1}) &:= \tilde{g}_i(\bar{x}_i, \xi_1, \xi_2, \dots, \xi_{i+1})
\end{aligned}$$

and

$$\Xi_i(\bar{x}_{i-1}, \xi_2, \xi_3, \dots, \xi_i) := \mathcal{X}_i(\bar{x}_{i-1}, \xi_1, \xi_2, \dots, \xi_i)$$

for each $i = 1, 2, \dots, M$.

First, we show that

$$\sup_{x_1 \in \Xi_1} \left|\mathbb{E}G_1(x_1, \xi_2) - \mathbb{E}\tilde{G}_1(x_1, \zeta_2)\right| \le \sum_{\nu=1}^{M-1} b_{\nu+1} \tag{15}$$

by induction according to M: If $M = 2$ then $\tilde{G}_1$ coincides with G_1 so (15) follows directly from (14). If $M > 2$ and (15) holds for $M-1$ then

$$\begin{aligned}
&\sup_{x_1 \in \Xi_1} \left|\mathbb{E}G_1(x_1, \xi_2) - \mathbb{E}\tilde{G}_1(x_1, \zeta_2)\right| \\
&\quad \le \sup_{x_1 \in \Xi_1} \left|\mathbb{E}G_1(x_1, \xi_2) - \mathbb{E}G_1(x_1, \zeta_2)\right| + \sup_{x_1 \in \Xi_1} \left|\mathbb{E}G_1(x_1, \zeta_2) - \mathbb{E}\tilde{G}_1(x_1, \zeta_2)\right|
\end{aligned}$$

$$\leq b_2 + \mathbb{E} \sup_{x_1 \in \Xi_1} \left| G_1(x_1, \zeta_2) - \tilde{G}_1(x_1, \zeta_2) \right|$$

(to derive the inequality above, we have used the triangular inequality similarly to Kaňková 2000, pp. 137,138). Further, by exploiting the fact that $\min f(x) - \min g(x) = f(\arg\min f(x)) - g(\arg\min g(x)) \leq f(\arg\min g(x)) - g(\arg\min g(x)) \leq \sup_x [f(x) - g(x)]$, we get

$$\begin{aligned}
&\mathbb{E} \sup_{x_1 \in \Xi_1} \left| G_1(x_1, \zeta_2) - \tilde{G}_1(x_1, \zeta_2) \right| \\
&\quad = \mathbb{E} \sup_{x_1 \in \Xi_1} \Big| \min_{x_2 \in \Xi_2(x_1, \zeta_2)} \mathbb{E}(G_2(x_1, x_2, \xi_2, \xi_3) | \xi_1 = \zeta_2) \\
&\qquad - \min_{x_2 \in \Xi_2(x_1, \zeta_2)} \mathbb{E}(\tilde{G}_2(x_1, x_2, \zeta_2, \zeta_3) | \zeta_2) \Big| \\
&\quad \leq \mathbb{E} \sup_{x_1 \in \Xi_1} \sup_{x_2 \in \Xi_2(x_1, \zeta_2)} \left| \mathbb{E}(G_2(x_1, x_2, \xi_2, \xi_3) | \xi_2 = \zeta_2) - \mathbb{E}(\tilde{G}_2(x_1, x_2, \zeta_2, \zeta_3) | \zeta_2) \right| \\
&\quad \leq \sum_{\nu=2}^{M-1} b_{\nu+1}
\end{aligned} \tag{16}$$

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

$$\min_{x_2 \in \Xi_2(x_1, , \xi_2)} \mathbb{E}(G_2(x_1, x_2, \xi_2, \xi_3) | \xi_2 = \zeta_2).$$

Hence, (15) is proved and we may use the construction (4) to get

$$\begin{aligned}
\mathbb{E}\eta_1(\xi_2) &= \mathbb{E}G_1(\tilde{x}_1, \xi_2) - \mathbb{E}G_1(\hat{x}_1, \xi_2) \\
&\leq \mathbb{E}G_1(\tilde{x}_1, \xi_2) - \mathbb{E}\tilde{G}_1(\tilde{x}_1, \zeta_2) + \mathbb{E}\tilde{G}_1(\hat{x}_1, \zeta_2) - \mathbb{E}G_1(\hat{x}_1, \xi_2) \\
&\leq 2 \sup_{x_1 \in \Xi_1} |\mathbb{E}G_1(x_1, \xi_2) - \mathbb{E}\tilde{G}_1(x_1, \zeta_2)| \overset{(15)}{\leq} 2 \sum_{\nu=1}^{M-1} b_{\nu+1}
\end{aligned} \tag{17}$$

(η_1 is defined by (12)). When we apply (17) to the interior problems (6) we get, analogously, that

$$\mathbb{E}\eta_i(\bar{\xi}_{i+1}) \leq 2 \sum_{\nu=i}^{M-1} b_{\nu+1} \tag{18}$$

so

$$\eta_\Pi \overset{\text{Lemma 1}}{=} \sum_{i=1}^{M-1} \mathbb{E}\eta_i(\bar{\xi}_{i+1}) \overset{(17),(18)}{\leq} 2 \sum_{i=1}^{M-1} \sum_{\nu=i}^{M-1} b_{\nu+1} = 2 \sum_{i=1}^{M-1} i \cdot b_{i+1}. \qquad \square$$

4 Some extensions to numerical integration

Before we state our main results, we will formulate several assertions concerning the approximate computation of integrals with respect to a probability measure using a discretization of the measure.

Discretization (i.e. the replacement of the original probability measure by a discrete one) is a widely used technique in the approximation of integrals. Assertions similar to the ones presented in the present work exist (see Davis and Rabinowitz 1975, Chap. 5), however, these results do not cover the situation with an unbounded support of the probability measure, common in stochastic programming. It should be also noted that there exist (in some sense) more accurate approximation techniques of the numerical integration (e.g. various quadrature rules, see Davis and Rabinowitz 1975 for a survey); however, we do not work with them because they usually require the differentiability of the integrated function while many popular multistage problems do not satisfy this requirement (e.g. multistage linear programming problems).

Let us consider the expression

$$\mathbb{E}_{P(\omega)}h(\omega) = \int h(\omega)dP(\omega)$$

where $h : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz with the constant K with respect to the l_1 norm (l_1-K-Lipschitz) and P is some k-dimensional distribution with finite first absolute moment. Since

$$h(0) - K\sum_{i=1}^{k}\mathbb{E}_{P(\omega)}|\omega^i| \le \mathbb{E}_{P(\omega)}h(\omega) \le h(0) + K\sum_{i=1}^{k}\mathbb{E}_{P(\omega)}|\omega^i|$$

(the symbol ω^i denotes the i-th component of the vector ω) and since the first moments of all the marginal distributions are finite according to our assumption, the expectation $\mathbb{E}_{P(\omega)}h(\omega)$ exists and is finite.

Let Π be a discrete distribution defined on supt(P). Denote

$$e(\Pi, h) = |\mathbb{E}_{\Pi(\omega)}h(\omega)) - \mathbb{E}_{P(\omega)}h(\omega)| \tag{19}$$

the error arising from replacing P by Π and

$$e(\Pi) = \sup_{h \text{ is } l_1\text{-}K\text{-Lipschitz}} e(\Pi, h) \tag{20}$$

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

Lemma 3 *If* $k = 1$ *then*

$$e(\Pi) = K \cdot d_W(P, \Pi) \tag{21}$$

where

$$d_W(P, \Pi) = \sup_{h \text{ is } l_1\text{-}1\text{-}Lipschitz} |\mathbb{E}_\Pi h - \mathbb{E}_P h|$$

and it holds that

$$d_W(P, \Pi) = \sum_{i=0}^{n}\int_{x_i}^{x_{i+1}}\left|F(x) - \sum_{j=1}^{i}p_j\right|dx \tag{22}$$

where $x_1 \le x_2 \le \cdots \le x_n$ *are the atoms of* Π ($F(x) = P\{(-\infty, x)\}$ *denotes the distribution function of* P, *we take* $x_0 = -\infty$, $x_{n+1} = \infty$ *and* $\sum_{j=1}^{0} p_j = 0$.)

Proof The formula (21) is straightforward. Let us prove (22): According to Pflug (2001), Theorem 1, it holds that

$$d_W(P, \Pi) = \int_{-\infty}^{\infty} |F(x) - G(x)| dx \tag{23}$$

where G is the distribution function of Π ((23) is proved in Vallander 1973). When we note that $G|_{(x_i, x_{i+1}]} = \sum_{j=1}^{i} p_j$ we get (22). □

The quantity $d_W(P, \Pi)$ is called the Wasserstein distance of measures P and Π (see Pflug 2001, Definition 1, and Rachev 1991, Chap. 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 4 Let $n = m_1 \cdot m_2 \cdots m_k$ for some positive integers $m_1, m_2, \ldots, m_k$. A distribution $\Pi = \{x_i, p_i\}_{i=1}^n$ is a grid discretization of P with dimensions $m_1, m_2, \ldots, m_k$ if there exist

$$-\infty = c_{1,0} < x_{1,1} \le c_{1,1} \le x_{1,2} \le \cdots \le x_{1,m_1} < c_{1,m_1} = \infty,$$
$$-\infty = c_{2,0} < x_{2,1} \le c_{2,1} \le x_{2,2} \le \cdots \le x_{2,m_2} < c_{2,m_2} = \infty,$$
$$\cdots$$
$$-\infty = c_{k,0} < x_{k,1} \le c_{k,1} \le x_{k,k} \le \cdots \le x_{k,m_k} < c_{k,m_k} = \infty$$

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

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

and

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

where

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

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

Remark 1 Denote P_i the i-th marginal distribution of P. If Π is a 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 $1 \le i \le k$. Moreover, Π_i is a grid discretization of P_i.

Remark 2 Conversely, if

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

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

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

is grid discretization of P with dimensions $m_1, m_2, \ldots, m_k$ such that $\Pi_1, \Pi_2, \ldots, \Pi_k$ are its marginal distributions.

The following Lemma provides an easy way of computing the error of an approximation of P by a grid discretization:

Lemma 4 *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 the marginal distributions of P, Π respectively.

Proof For each l_1-K-Lipschitz function h it holds that

$$\begin{aligned}
&|E_P h - E_\Pi h| \\
&\quad = \left| \int_{\mathbb{R}^k} h\,dP - \sum_{j_1=1, j_2=1,\ldots,j_k=1}^{m_1,m_2,\ldots,m_k} h(x_{1,j_1}, x_{2,j_2}, \ldots, x_{k,j_k})\, p_{j_1,j_2,\ldots,j_k} \right| \\
&\quad = \left| \sum_{j_1,j_2,\ldots,j_k=1}^{m_1,m_2,\ldots,m_k} \left[\int_{M_{j_1,j_2,\ldots,j_k}} h\,dP - h(x_{1,j_1}, x_{2,j_2}, \ldots, x_{k,j_k}) \int_{M_{j_1,j_2,\ldots,j_k}} dP \right] \right| \\
&\quad = \left| \sum_{j_1,j_2,\ldots,j_k=1}^{m_1,m_2,\ldots,m_k} \int_{M_{j_1,j_2,\ldots,j_k}} [h(y_1, y_2, \ldots, y_k) \right. \\
&\qquad \left. - h(x_{1,j_1}, x_{2,j_2}, \ldots, x_{k,j_k})]\, dP(y_1, y_2, \ldots, y_k) \right| \\
&\quad \le \sum_{j_1,j_2,\ldots,j_k=1}^{m_1,m_2,\ldots,m_k} \int_{M_{j_1,j_2,\ldots,j_k}} K \sum_{i=1}^{k} |y_i - x_{i,j_i}|\, dP(y_1, y_2, \ldots, y_k) \\
&\quad = \sum_{i=1}^{k} \left[K \sum_{j_1,j_2,\ldots,j_k=1}^{m_1,m_2,\ldots,m_k} \int_{M_{j_1,j_2,\ldots,j_k}} |y_i - x_{i,j_i}|\, dP(y_1, y_2, \ldots, y_k) \right] \\
&\quad = \sum_{i=1}^{k} \left[K \sum_{j_i=1}^{m_i} \int_{\bigcup_{j_1,\ldots,j_{i-1},j_{i+1},\ldots,j_m} M_{j_1,j_2,\ldots,j_k}} |y_i - x_{i,j_i}|\, dP(y_1, y_2, \ldots, y_k) \right] \\
&\quad = \sum_{i=1}^{k} \left[K \sum_{j=1}^{m_i} \int_{\mathbb{R}\times\cdots\times\mathbb{R}\times[c_{i,j-1}, c_{i,j})\times\mathbb{R}\times\cdots\times\mathbb{R}} |y_i - x_{i,j}|\, dP(y_1, y_2, \ldots, y_k) \right] \\
&\quad = \sum_{i=1}^{k} K D_i
\end{aligned} \tag{24}$$

where

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

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

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

where the constants $a_{i,j}$ are chosen so that h_i is continuous, then we get the "=" instead of the "$\leq$" in (24). Therefore and since $\bar{h}$ is l_1-K-Lipschitz, we have

$$\sup_{h \text{ is } l_1\text{-}K\text{-Lipschitz}} |E_P h - E_\Pi h| = |E_P \bar{h} - E_\Pi \bar{h}| = \sum_{i=1}^{k} K D_i. \tag{25}$$

Moreover, it holds that

$$d_W(P_i, \Pi_i) \overset{(21)}{=} \sup_{h \text{ is } l_1\text{-}1\text{-Lipschitz}} |E_{P_i} h - E_{\Pi_i} h| \overset{(25) \text{ with } k=1}{=} D_i$$

for each $i = 1, 2, \ldots, k$. By the combination of (25) and (26) we get the assertion of the Lemma. □

It is known (see Davis and Rabinowitz 1975, Theorem on p. 267) that if $h : \mathbb{R}^k \to \mathbb{R}$ is a function with a finite variation and $H \subset \mathbb{R}^k$ is a bounded set then the error of approximation of $\int_H h(x)dx$ by $\frac{1}{n}\sum_{i=1}^{n} h(x_i)$, where $x_1, x_2, \ldots, x_n$ are suitably chosen points, is $O(n^{-1/k})$. The present work generalizes this result.

Theorem 1 *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 grid discretizations Π^n may be constructed such that*

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

Before we prove the Theorem we will formulate a useful auxiliary assertion:

Lemma 5 *Let Q be a probability distribution on $\mathbb{R}$ with the distribution function F and let $m > 1$ be an integer. Then a grid discretization Γ of Q with at most m atoms exists such that*

$$d_W(Q, \Gamma) \leq \int_{-\infty}^{-C} F(x)dx + \frac{2C}{m-2} + \int_C^{\infty} [1 - F(x)]dx \tag{26}$$

for each $C \geq 0$. Here, $F^{-1}(\alpha) = \inf\{e \in \mathbb{R} : F(e) \geq \alpha\}$ denotes the quantile function of F.

Proof of the Lemma Define

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

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

and

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

It is easy to verify that $F^{-} + \tilde{F} + F^{+} = F$.

Further, define

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

and

$$\tilde{G}(\xi) := \tilde{F}(\tilde{F}^{-1}(D)) I_{(-C,\tilde{F}^{-1}(D)]}(\xi) + \sum_{i=2}^{m-2} \tilde{F}(\tilde{F}^{-1}(iD)) I_{(\tilde{F}^{-1}((i-1)D),\tilde{F}^{-1}(iD)]}(\xi)$$
$$+ \tilde{F}(C) I_{(\tilde{F}^{-1}((m-2)D),\infty)}(\xi), \quad \xi \in \mathbb{R},$$

where $D = \frac{\tilde{F}(C)}{m-2}$ and $\tilde{F}^{-1}(y) = \inf\{e \in \mathbb{R} : \tilde{F}(e) \geq y\}$. It is possible to show that $\tilde{F}(\tilde{F}^{-1}(d)) \leq d$ and that $\tilde{F}(\tilde{F}^{-1}(d)+) \geq d$ for each $d \in (0, \tilde{F}(C)]$; using it, we can estimate: If $\xi \in (-C, \tilde{F}^{-1}(D)]$ then

$$|\tilde{G}(\xi) - \tilde{F}(\xi)| = |\tilde{F}(\tilde{F}^{-1}(D)) - \tilde{F}(\xi)| = \tilde{F}(\tilde{F}^{-1}(D)) - \tilde{F}(\xi)$$
$$\leq \tilde{F}(\tilde{F}^{-1}(D)) \leq D. \tag{27}$$

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

$$|\tilde{G}(\xi) - \tilde{F}(\xi)| = |\tilde{F}(\tilde{F}^{-1}(iD)) - \tilde{F}(\xi)| = \tilde{F}(\tilde{F}^{-1}(iD)) - \tilde{F}(\xi)$$
$$\leq \tilde{F}(\tilde{F}^{-1}(iD)) - \tilde{F}(\tilde{F}^{-1}((i-1)D)^{+})$$
$$\leq iD - (i-1)D = D. \tag{28}$$

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

$$|\tilde{G}(\xi) - \tilde{F}(\xi)| = |\tilde{F}(C) - \tilde{F}(\xi)| = \tilde{F}(C) - \tilde{F}(\xi)$$
$$\leq \tilde{F}(C) - \tilde{F}(\tilde{F}^{-1}((m-2)D)^{+})$$
$$\stackrel{\text{def. of } D}{=} \tilde{F}(C) - \tilde{F}(\tilde{F}^{-1}(\tilde{F}(C))^{+}) \tag{29}$$
$$\leq \tilde{F}(C) - \tilde{F}(C) = 0. \tag{30}$$

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

$$|\tilde{G}(\xi) - \tilde{F}(\xi)| = |0 - 0| = 0 \tag{31}$$

for $\xi \in (-\infty, -C] \cup (C, \infty)$. When we summarize (27), (28), (30) and (31) we get

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

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

$$|\tilde{G}(\xi) - \tilde{F}(\xi)| \le D \cdot I_{(-C,C]}(\xi). \tag{32}$$

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 $x_1 := -C, x_2 := \tilde{F}^{-1}(D), x_3 := \tilde{F}^{-1}(2D), \ldots, x_{m-1} := \tilde{F}^{-1}((m-2)D).x_m := C$, some of the values may coincide), G is the distribution function of an at most m-atom discrete probability distribution—let us denote it by Γ. Using Vallander (1973), triangular inequality, formula (32) and the definitions of F^-, F^+, G^- and G^+, we gradually obtain

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

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 $x_1 \le c_1 \le x_2 \le c_2 \le x_3 \le \cdots \le c_{m-1} \le x_m$ and $\Gamma\{x_i\} = F(c_i) - F(c_{i-1})$, i.e.

$$G(x_{i+1}) - G(x_i) = F(c_i) - F(c_{i-1}) \tag{33}$$

for each $i = 1, 2, \ldots, m$ (we take $x_{m+1} = \infty$, $G(x_{m+1}) = 1$, $c_0 = -\infty$, $F(c_0) = 0$, $c_m = \infty$ and $F(c_m) = 1$). Clearly, (33) holds if

$$G(x_{i+1}) = F(c_i) \tag{34}$$

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

$$\begin{aligned}
G(x_1) &= 0 = F(c_0), \\
G(x_{i+1}) &= G(\tilde{F}^{-1}(iD)) = F(-C) + \tilde{F}(\tilde{F}^{-1}(iD)) = F(\tilde{F}^{-1}(iD)) = F(c_i)
\end{aligned}$$

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

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

i.e. (34) is verified. □

Proof of the Theorem By using Lemma 5, we get that, for each $1 \le i \le k$ and $m \in \mathbb{N}$, there exists an at most m-atom grid discretization Π_i^m such that

$$d_W(\Pi_i^m, Q_i) \le \bar{d}_{m,i}, \qquad \bar{d}_{m,i} = \int^{-m^{1/a}} F_i(\xi)d\xi + \frac{2m^{1/a}}{m+1} + \int_{m^{1/a}} [1 - F_i(\xi)]d\xi \quad (35)$$

(we have put $C = m^{1/a}$ in Lemma 5).

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

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

for some $A < \infty$, there exists $\epsilon > 0$ and x_0 such that $F_i(-x) \le (A+\epsilon)x^{-a}$, for each $x \ge x_0$. Therefore,

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

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

Clearly

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

so, due to (35),

$$\bar{d}_{m,i} = O(m^{-1+1/a}) + O(m^{-1+1/a}) + O(m^{-1+1/a}) = O(m^{-1+1/a}). \quad (36)$$

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

$$\begin{aligned}
e(\Pi^n) &\overset{\text{Lemma 4}}{=} \sum_{i=1}^{k} K \cdot d_W(\Pi_i^{\lfloor n^{1/k} \rfloor}, P_i) \\
&\overset{(35)}{\le} K \sum_{i=1}^{k} \bar{d}_{\lfloor n^{1/k} \rfloor, i} \overset{(36)}{=} O\big(\lfloor n^{\frac{1}{k}} \rfloor^{-1+1/a}\big) = O\big(n^{-\frac{1}{k}+\frac{1}{ka}}\big).
\end{aligned}$$

□

Corollary *Let P have marginal densities $f_i(x) = e^{-\phi_i(x)}$, $i = 1, 2, \dots, k$, 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*

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

for each arbitrarily small $\delta > 0$.

Proof Since

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

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

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

i.e. $F_i(m)=O(m^{-\frac{2}{k\delta}})$ as $m\to\infty$, and it follows from Theorem 1 that there exists a sequence of at most n-atom distributions Π_n such that

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

5 Upper bound and convergence rate of the expected loss

Adding to the assumptions from Sect. 1, we assume that

(c) $g_i(\bar{x}_i,\bar{\xi}_i,\bullet)$ is uniformly l_1-Lipschitz with a constant K_i for each $i=1,2,\dots,M-1$ (i.e. there exists K_i such that $|g_i(\bar{x}_i,\bar{\xi}_i,z_1)-g_i(\bar{x}_i,\bar{\xi}_i,z_2)|\le K_i\|z_1-z_2\|_1$ almost surely for each feasible $\bar{x}_i$ and each $z_1,z_2\in\operatorname{supt}(\xi_{i+1})$).

The following theorem provides an upper bound on the expected loss:

Theorem 2 *Let Π be an approximation scheme with dimensions $n_2,n_3,\dots,n_M$ such that $\Pi_i(\bullet|\bar{\xi}_{i-1})$ is a grid discretization of $\mathcal{L}(\xi_i|\bar{\xi}_{i-1})$ almost surely for each $i=2,3,\dots,M$. Moreover, let there exist constants $d_{i,\nu}$, $i=2,3,\dots,M$, $\nu=1,2,\dots,k_i$ such that*

$$d_W(\mathcal{L}(\xi_i^\nu|\bar{\xi}_{i-1}),\Pi_{i,\nu}(\bar{\xi}_{i-1}))\le d_{i,\nu} \tag{37}$$

almost surely (ξ_i^ν denotes the ν-th component of the vector ξ_i, $\Pi_{i,\nu}(\bar{\xi}_{i-1})$ denotes ν-th marginal distribution of $\Pi_i(\bar{\xi}_{i-1})$). Then

$$\eta_\Pi\le 2\sum_{i=1}^{M-1}\left(i\cdot K_i\cdot\sum_{\nu=1}^{k_{i+1}}d_{i+1,\nu}\right)$$

independently of the choice of the approximating problem's solutions.

Proof Since

$$\left|\mathbb{E}(g_i(\bar{x}_i, \bar{\xi}_i, \xi_{i+1})|\bar{\xi}_i) - \mathbb{E}(g_i(\bar{x}_i, \bar{\xi}_i, \zeta^i_{i+1})|\bar{\xi}_i)\right|$$

$$\overset{\text{(c), Lemma 4}}{\leq} K_i \sum_{\nu=1}^{k_{i+1}} d_W(\mathcal{L}(\xi^\nu_{i+1}|\bar{\xi}_i), \Pi_{i+1,\nu}(\bar{\xi}_i)) \overset{(37)}{\leq} K_i \sum_{\nu=1}^{k_{i+1}} d_{i+1,\nu}$$

almost surely for each $1 \leq i \leq M-1$ and each feasible $\bar{x}_i$, the assertion of the present Theorem follows directly from Lemma 2. □

The last assertion of the present article states a convergence rate of the expected loss achievable by a suitable sequence of the approximation schemes:

Theorem 3 *Assume that there exists a constant $a > 1$ and a function $h(x) = O(x^{-a})$ as $x \to \infty$ such that*

$$\max(F^{\bar{\xi}_{i-1}}_{i,\nu}(-x), 1 - F^{\bar{\xi}_{i-1}}_{i,\nu}(x)) \leq h(x) \tag{38}$$

almost surely for each $i = 2, 3, \ldots, M$ and $\nu = 1, 2, \ldots, k_i$ (the symbol $F^{\bar{\xi}_{i-1}}_{i,\nu}$ denotes the distribution function of $\mathcal{L}(\xi^\nu_i|\bar{\xi}_{i-1})$). Then there exists a sequence $\{\Pi^n\}_{n=1}^\infty$ of at most n-atom approximation schemes such that

$$\eta_{\Pi_n} = O\left(n^{-\frac{1}{k}+\frac{1}{ka}}\right), \quad k = \sum_{i=2}^M k_i,$$

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

Proof Denote $m_n = \lfloor n^{1/k} \rfloor$. When we put $C = m_n^{1/a}$ in Lemma 5, we get that for each $2 \leq i \leq M$, $1 \leq \nu \leq k_i$ and each realization of $\bar{\xi}_{i-1}$ there exists a grid discretization $\Pi^{m_n}_{i,\nu,\bar{\xi}_{i-1}}$ of $\mathcal{L}(\xi^\nu_i|\bar{\xi}_{i-1})$ with the dimension at most m_n such that

$$d_W(\Pi^{m_n}_{i,\nu,\bar{\xi}_{i-1}}, \mathcal{L}(\xi^\nu_i|\bar{\xi}_{i-1}))$$

$$\overset{\text{Lemma 5}}{\leq} \int_{-\infty}^{-m_n^{1/a}} F^{\bar{\xi}_{i-1}}_{i,\nu}(x)dx + \frac{2m_n^{1/a}}{m_n+1} + \int_{m_n^{1/a}}^\infty [1 - F^{\bar{\xi}_{i-1}}_{i,\nu}(x)]dx \overset{(38)}{\leq} \bar{\delta}_{m_n} \tag{39}$$

where

$$\bar{\delta}_{m_n} = 2\int_{m_n^{1/a}}^\infty h(x)dx + \frac{2m_n^{1/a}}{m_n+1}$$

and it can be shown analogously to the proof of Theorem 1 that

$$\bar{\delta}_{m_n} = \bar{\delta}_{\lfloor n^{1/k} \rfloor} = O\left(n^{-\frac{1}{k}+\frac{1}{ka}}\right). \tag{40}$$

Let $\Pi^{m_n}_{i,\bar{\xi}_{i-1}}$ be the grid discretization defined by $\Pi^{m_n}_{i,1,\bar{\xi}_{i-1}}, \Pi^{m_n}_{i,2,\bar{\xi}_{i-1}}, \ldots, \Pi^{m_n}_{i,k_i,\bar{\xi}_{i-1}}$ for each $i = 2, 3, \ldots, M$ (see Remark 2 for an explanation how $\Pi^{m_n}_{i,\bar{\xi}_{i-1}}$ is defined using its marginal distributions) and let Π^n be the approximation scheme defined by $\Pi^{m_n}_{2,\bullet}, \Pi^{m_n}_{3,\bullet}, \dots, \Pi^{m_n}_{M,\bullet}$. Clearly, Π^n has

$$m_n^{k_1} \cdot m_n^{k_2} \cdots m_n^{k_M} = m_n^k \le n$$

atoms and it holds that

$$\eta_{\Pi_n} \overset{\text{Theorem 2}}{\le} \bar{\delta}_{m_n} \sum_{i=1}^{M-1} i K_i k_{i+1} \overset{(40)}{=} O\left(n^{-\frac{1}{k}+\frac{1}{ka}}\right).$$

□

6 Conclusion

In the present article, generalized results concerning the approximate computation of expectations were presented and the behavior of the approximation error (measured by the expected loss) of the discretization of multistage stochastic programming problems was studied under rather general conditions. It was found that the behavior of the expected loss depends only on the Lipschitz properties of the MSSPP and on the asymptotic behavior of the tails of the marginal distributions of the random parameters.

In particular, the properties of the numerical estimators of the expectations in the MSSPP are inherited by the expected loss.

Based on these facts, new results concerning the quantitative stability of the multistage stochastic programs were achieved in the present article.

Acknowledgements This research was partially supported by Grant Agency of the Czech Republic under Grants No. 402/04/1294 and No. 402/06/1417.

The author thanks an anonymous referee for valuable remarks and suggestions.

References

Davis, P. J., & Rabinowitz, P. (1975). *Methods of numerical integration*. New York: Academic Press.

Dempster, M. A. H., & Thompson, G. W. P. (2002). Dynamic portfolio replication using stochastic programming. In M. Dempster (Ed.), *Risk management: value at risk and beyond* (pp. 100–128). Cambridge: Cambridge University Press.

Dupačová, J., Hurt, J., & Štěpán, J. (2002). *Stochastic modelling in economics and finance*. Dordrecht: Kluwer.

Dupačová, J., Gröve-Kuska, N., & Römisch, V. (2003). Scenario reduction in stochastic programming, an approach using probability metrics. *Mathematical Programming Series A*, *95*, 493–511.

Kall, P., Ruszczyński, P., & Frauendorfer, K. (1988). Approximation techniques in stochastic programming. In Y. Ermoliev & R. J.-B. and Wets (Eds.), *Numerical techniques for stochastic optimization* (pp. 33–64). Berlin: Springer.

Kaňková, V. (2000). Multistage stochastic programming: Stability, approximation and Markov dependence. In *Operations research proceedings 1999* (pp. 136–141). Springer: Berlin.

Kaňková, V., & Šmíd, M. (2004). On approximation in multistage stochastic programs: Markov dependence. *Kybernetika*, *40*(5), 625–638.

Mak, W. K., Morton, D. P., & Wood, R. K. (1999). Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, *24*, 47–56.

Pennanen, T., & Koivu, M. (2005). Epi-convergent discretizations of stochastic programs via integration quadratures. *Numerische Mathematik*, *100*(1), 141–163.

Pflug, G. Ch. (2001). Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming Series B*, *89*(2), 251–271.

Rachev, S. T. (1991). *Probability metrics and the stability of stochastic models*. New York: Wiley.

Ruszczyński, A., & Shapiro, A. (Eds.). (2003). *Stochastic programming*. *Handbooks in operations research and management science* (Vol. 10). Amsterdam: Elsevier.

Vallander, S. S. (1973). Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability and Its Applications*, *18*, 784–786.