Wiley StatsRef:
Statistics Reference Online

Stochastic Programming, Scenario Generation in

By *Werner Römisch*

Keywords: *stochastic programming, scenario, stability, quadrature, Monte Carlo, quasi-Monte Carlo, sparse grid, quantization, moment matching*

Abstract: Stability-based methods for scenario generation in stochastic programming are reviewed. In particular, we discuss Monte Carlo sampling, quasi-Monte Carlo methods, and quadrature rules based on sparse grids, optimal quantization, and moment matching methods. In addition, we provide some results on convergence rates. Scenario reduction and scenario tree generation are briefly mentioned.

Many **stochastic programming** problems may be stated in the form

$$\text{minimize} \quad \int_{\Xi} f_0(x,\xi)P(d\xi) \quad \text{such that} \quad x \in X \text{ and } \int_{\Xi} f_1(x,\xi)P(d\xi) \leq 0 \tag{1}$$

where X is a closed subset of $\mathbb{R}^m$, Ξ is a closed subset of $\mathbb{R}^d$, the functions f_0 and f_1 map from $\mathbb{R}^m \times \Xi$ to the extended real numbers $\mathbb{R} \cup \{\pm\infty\}$, and P is a probability distribution on Ξ. The set X is used to describe all constraints not depending on P and the set Ξ to contain the support of P. The integrands f_0 and f_1 are assumed to be lower semicontinuous jointly in (x, ξ) implying that all integrals in Equation (1) are well defined (although possibly infinite).

Classical examples are *two-stage stochastic (integer) programs* and *optimization models with probabilistic constraints* (see Ref. 1 for the specific form of integrands). The corresponding integrands f_0 and f_1, respectively, are nondifferentiable or even discontinuous. (We assume here that the integrals in problem (1) are finite for every $x \in X$.)

The most important approach to solve problem (1) computationally consists of replacing the integrals in Equation (1) by **numerical integration** formulas. This leads to the optimization problems

$$\text{minimize} \quad \sum_{i=1}^{n} w_i f_0(x,\xi^i) \quad \text{such that} \quad x \in X \text{ and } \sum_{i=1}^{n} w_i f_1(x,\xi^i) \leq 0 \tag{2}$$

for some $n \in \mathbb{N}$, weights $w_i \in \mathbb{R}$ and elements $\xi^i \in \Xi$, $i = 1, \ldots, n$. Such numerical integration schemes result by replacing P in problem (1) by a finite signed measure Q_n with support $\{\xi^1, \ldots, \xi^n\}$ and $Q_n(\{\xi^i\}) = w_i$. If Q_n is a probability measure, then ξ^i are called *scenarios*. An extension of known stability results for problem (1) with respect to approximations of the original probability distribution P (see Ref. 2

Humboldt-University Berlin, Berlin, Germany

DOI: 10.1002/9781118445112.stat08075

for a survey) states that infimal values $v(P)$ and $v(Q_n)$ and solution sets $S(P)$ and $S(Q_n)$ of Equation (1), respectively, get close if the distance

$$d_{\mathcal{F}}(P, Q_n) = \sup_{f\in\mathcal{F}}|P(f) - Q_n(f)| = \sup_{f\in\mathcal{F}}\left|\int_{\Xi} f(\xi)P(\mathrm{d}\xi) - \sum_{i=1}^{n} w_i f(\xi^i)\right| \tag{3}$$

of P and Q_n with $\mathcal{F} = \{f_j(x,\cdot) : x \in X,\ j = 0, 1\}$ gets small, the set X is compact, the objective function $x \mapsto \int_{\Xi} f_0(x, \xi)P(\mathrm{d}\xi)$ is Lipschitz continuous on X and a metric regularity condition for the constraint set is satisfied. The latter two conditions are not needed if the constraints of problem (1) do not depend on P.

Approximate stochastic programming problems (problem (2)) represent standard (linear, nonlinear, and integer) optimization models. The only difficulty consists in the need of many evaluations of the functions f_j at the pairs (x, ξ^i), $i = 1, \ldots, n$, if the number n gets large. But, large n are often unavoidable when recalling that the dimension d is often large in applied stochastic programming models in energy, transportation, and finance (see Ref. 3).

The behavior of the error $e_n(P, \mathcal{F}) = d_{\mathcal{F}}(P, Q_n)$ with respect to $n \in \mathbb{N}$ depends heavily on the set $\mathcal{F}$ of integrands as well as on the probability distribution P. As the set $\mathcal{F}$ in its present form is often not convenient to handle, it might be an alternative to enlarge $\mathcal{F}$ in Equation (3). However, if, for example, $\mathcal{F}$ is enlarged to coincide with the unit ball $\mathbb{B}$ in the Banach space $\mathrm{Lip}(\mathbb{R}^d)$ of Lipschitz continuous functions on $\mathbb{R}^d$, $d_{\mathcal{F}}$ represents the dual of the Wasserstein metric of order 1 (see Equation (8) for $r = 1$) and the convergence rate of $e_n(P, \mathbb{B})$ is at most $O(n^{-\frac{1}{d}})$ if P has a density with respect to the Lebesgue measure on $\mathbb{R}^d$.

Next, we discuss five scenario generation techniques for solving problem (1):

1 Monte Carlo Methods

Monte Carlo (MC) methods are based on drawing independent identically Distributed (iid) Ξ-valued random samples $\xi^1(\cdot), \ldots, \xi^n(\cdot), \ldots$ (defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$) from some probability distribution P (on Ξ)(*see* **Monte Carlo Methods**). It is well-known (see Ref. 4, Theorem 2.1) that

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} f(\xi^i(\omega)) - \int_{\Xi} f(\xi)P(\mathrm{d}\xi)\right)^2 = \frac{1}{n}\mathbb{E}\left(f(\xi^1(\omega)) - \int_{\Xi} f(\xi)P(\mathrm{d}\xi)\right)^2 \tag{4}$$

holds for any real function f on Ξ that is quadratically integrable with respect to P. Uniform versions of estimate (Equation (4)) imply that $e_n(P, \mathcal{F})$ has the typical mean square convergence rate $O(n^{-\frac{1}{2}})$ of MC methods which is slow in practice. Such iid samples are approximately obtained by pseudo random number generators first as uniform samples in $[0, 1]^d$ and later transformed to more general probability distributions (for example based on the Rosenblatt transformation[5]). We refer to Ref. 1 (Chapter 5) and the references therein for applying MC methods in stochastic programming.

2 Quasi-Monte Carlo Methods

The basic idea of quasi-Monte Carlo (QMC) methods is to replace random samples in MC methods by a sequence $(\xi^i)_{i\in\mathbb{N}}$ of deterministic points that are *uniformly distributed* in $[0, 1]^d$. The latter property means that the *star-discrepancy* $e_n(\lambda^d, \mathcal{F}_b)$ of the set $\{\xi^1, \ldots, \xi^n\}$ converges to zero if $n \to \infty$. Here, $\mathcal{F}_b$ denotes the class of characteristic functions $\mathbf{1}_{[0,\xi]}$ of boxes $[0, \xi]$ for each $\xi \in [0, 1]^d$, λ^d the d-dimensional Lebesgue measure and Q_n the uniform probability measure with support $\{\xi^1, \ldots, \xi^n\}$. It is known that there exist sequences such that $e_n(\lambda^d, \mathcal{F}_b) = O(n^{-1}(\log n)^d)$ holds and that this rate carries over to $e_n(\lambda^d, \mathcal{F})$ if $\mathcal{F}$ is a

DOI: 10.1002/9781118445112.stat08075

bounded subset of the tensor product Sobolev space $W_2^{(1,\dots,1)}([0,1]^d)$ of functions with mixed first derivatives belonging to $L_2[0,1]$. The important examples are *Sobol' sequences* and certain *lattice point sets* (see Refs 4, Section 5, 6, Chapter 8). Later, it turned out that proper randomizations of QMC point sets lead to improvements of the convergence rate. For example, if $\mathcal{F}$ is a bounded subset of $W_2^{(1,\dots,1)}([0,1]^d)$, there exists a weighted norm on that space such that randomly shifted lattice rules attain the following estimate

$$\sqrt{\mathbb{E}[e_n^2(\lambda^d,\mathcal{F})]} \le C(\delta)\, n^{-1+\delta} \tag{5}$$

for the root mean square error, where $\delta \in (0, \frac{1}{2}]$, $C(\delta)$ is independent of n and d (see Ref. 4, Section 5). It is shown in Ref. 7 that the results apply to linear two-stage stochastic programs if the integrands have low effective dimension[8]. For further information, we refer to Refs 4 and 6.

3 Quadrature Rules Using Sparse Grids

We consider again the unit cube $[0,1]^d$. Let nested sets of grids $\Xi^i = \{\xi_1^i, \dots, \xi_{m_i}^i\} \subset \Xi^{i+1}$ in $[0,1]$ be given for $i \in \mathbb{N}$, for example, the dyadic grids $\Xi^i = \{\frac{j}{2^i} : j = 0, 1, \dots, 2^i\}$. Then, the point set suggested by Smolyak

$$H(q,d) := \bigcup_{|\mathbf{i}|=\sum_{j=1}^d i_j = q} \Xi^{i_1} \times \cdots \times \Xi^{i_d} \tag{6}$$

is called a *sparse grid* in $[0,1]^d$. In the case of dyadic grids in $[0,1]$, $H(q,d)$ consists of all d-dimensional dyadic grids with the product of mesh size given by $\frac{1}{2^q}$.

The corresponding sparse grid quadrature rules for $q \ge d$ on $[0,1]^d$ with respect to the Lebesgue measure λ^d are of the form

$$Q_{n(q,d)}(f) = \sum_{q-d+1\le|\mathbf{i}|\le q} (-1)^{q-|\mathbf{i}|} \binom{d-1}{q-|\mathbf{i}|} \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_d=1}^{m_{i_d}} f(\xi_{j_1}^{i_1}, \dots, \xi_{j_d}^{i_d}) \prod_{l=1}^{d} a_{j_l}^{i_l} \tag{7}$$

where $n = n(q,d)$ is the number of summands in Equation (7), and the coefficients a_j^i, $j = 1, \dots, m_i$ are weights of one-dimensional quadrature rules based on Ξ^i, $i \in \mathbb{N}$. Even if the one-dimensional weights are positive, some of the weights in Equation (7) become negative. Hence, an interpretation as scenario-based probability measure is no longer possible. However, if the class $\mathcal{F}$ is a bounded subset of $W_2^{(1,\dots,1)}([0,1]^d)$, the error $e_n(\lambda^d,\mathcal{F})$ has the convergence rate $O(n^{-1}(\log n)^{2(d-1)})$. For details and further information see Refs 9 and 10.

4 Optimal Quantization of Probability Measures

Let $r \ge 1$ and consider the Wasserstein metric W_r on the set of all probability measures defined on Ξ that have r-th order moments, that is,

$$W_r(P,Q) = \left(\inf \left\{ \int_{\Xi\times\Xi} \|\xi - \tilde{\xi}\|^r \eta(\mathrm{d}\xi\, \mathrm{d}\tilde{\xi}) : \pi_1\eta = P, \pi_2\eta = Q \right\} \right)^{\frac{1}{r}} \tag{8}$$

where π_1 and π_2 denote the projections onto the first and the second component, respectively. Let P be a given probability distribution on Ξ. For given $n \in \mathbb{N}$, the support $\{\xi^1, \dots, \xi^n\}$ of a discrete probability measure Q_n on Ξ is called *optimal n-quantizer* of P of order r if Q_n is the best approximation to P with

DOI: 10.1002/9781118445112.stat08075

respect to W_r. The rth power of the distance $W_r(P, Q_n)$ enables a reformulation as a real function $\Phi_{n,r}$ on Ξ^n by defining

$$\Psi_{n,r}(\xi^1, \dots, \xi^n) = W_r^r(P, Q_n) = \int_\Xi \min_{i=1,\dots,n} \| \xi - \xi^i \|^r P(d\xi)$$

The function $\Psi_{n,r}$ is continuous and has compact level sets, hence, optimal n-quantizers of order r exist, but $\Psi_{n,r}$ is nonconvex and has probably many local minimizers. Hence, its global minimization on Ξ^n is not an easy task.

Moreover, the probability w_i of each scenario ξ^i has to be computed by $w_i = P(A_i)$, where $A_i, i = 1, \dots, n$, is a Voronoi partition[11] of Ξ. It is known that $W_r(P, Q_n)$ converges to zero, but not faster than $O(n^{-\frac{1}{d}})$ (see Ref. 11). Global minimization of $\Psi_{n,r}$ may be done by stochastic gradient algorithms or stochastic approximation methods (see Refs 12, 13).

5 Moment-Matching Methods

The idea consists of determining a discrete probability measure Q_n with scenarios $\xi^i \in \Xi$ and probabilities $w_i \geq 0, i = 1, \dots, n$, such that P and Q_n have identical moments defined by certain multivariate monomials, that is, it holds

$$\int_\Xi \prod_{j=1}^{d} (\xi_j)^{\alpha_j} P(d\xi) = \sum_{i=1}^{n} w_i \prod_{j=1}^{d} (\xi_j^i)^{\alpha_j}, \; (w_i, \xi^i) \in \mathbb{R}_+ \times \Xi, \; i = 1, \dots, n \tag{9}$$

for any $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathcal{I}$ with a prescribed multi-index set $\mathcal{I} \subset \mathbb{N}_0^d$. The matching polynomial moment conditions represent a system of nonlinear (polynomial) equations with $n(d+1)$ unknowns, $|\mathcal{I}|$ equations, and additional constraints (see Ref. 14). The connection to numerical integration rules that are exact for a class of d-variate polynomials, for example, for the class $\mathcal{P}_r$ of all d-variate polynomials of degree up to $r \in \mathbb{N}$, is revealed in Ref. 15. There it is also shown that for any given set $\mathcal{I}$, the system (9) is solvable for $n = |\mathcal{I}|$. Solution methods are developed and convergence for $r \to \infty$ is studied in Ref. 15, too.

Stability-based methods for *scenario reduction* are reviewed in Ref. 16 (see also Ref. 17 for original work). The *generation of scenario trees* for multistage stochastic programs is studied in [13, Chapter 4] and in Refs 14, 18 and 19.

References

[1] Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2014) *Lectures on Stochastic Programming*, MOS-SIAM Series on Optimization, 2nd edn, SIAM, Phildelphia, PA.
[2] Römisch, W. (2003) Stability of stochastic programming problems, in *Stochastic Programming*, Handbooks in Operations Research and Management Science, **vol. 10** (eds A. Ruszczyński and A. Shapiro), Elsevier, Amsterdam, pp. 483–554.
[3] Wallace, S. W. and Ziemba, W. T. (eds) (2005) *Applications of Stochastic Programming, MOS-SIAM Series on Optimization*, SIAM, Phildelphia, PA.
[4] Dick, J., Kuo, F.Y. and Sloan, I.H. (2013) High-dimensional integration – the Quasi-Monte Carlo way. *Acta Numer.* **22**, 1–153.
[5] Rosenblatt, M. (1952) Remarks on a multivariate transformation. *Ann. Math. Stat.* **23**, 470–472.
[6] Dick, J. and Pillichshammer, F. (2010) *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press, Cambridge.
[7] Leövey, H. and Römisch, W. (2015) Quasi-Monte Carlo methods for linear two-stage stochastic programming problems. *Math. Program.* **151**, 315–345.

DOI: 10.1002/9781118445112.stat08075

[8] Owen, A.B. (2003) The dimension distribution and quadrature test functions. *Stat. Sin.* **13**, 1–17.
[9] Gerstner, T. and Griebel, M. (1998) Numerical integration using sparse grids. *Numer. Algorithms* **18**, 209–232.
[10] Novak, E. and Ritter, K. (1996) High dimensional integration of smooth functions over cubes. *Numer. Math.* **75**, 79–97.
[11] Graf, S. and Luschgy, H. (2000) *Foundations of Quantization for Probability Distributions*, Lecture Notes in Mathematics, **Vol. 1730**, Springer, Berlin.
[12] Pagès, G., Pham, H., and Printems, J. (2004) Optimal quantization methods and applications to numerical problems in finance, in *Handbook of Computational and Numerical Methods in Finance*, (ed S. T. Rachev), Birkhäuser, Boston, pp. 253–297.
[13] Pflug, G.Ch. and Pichler, A. (2014) *Multistage Stochastic Optimization*, Springer, Cham.
[14] Dupačová, J., Consigli, G., and Wallace, S.W. (2000) Scenarios for multistage stochastic programs. *Ann. Oper. Res.* **100**, 25–53.
[15] Mehrotra, S. and Papp, D. (2013) Generating moment matching scenarios using optimization techniques. *SIAM J. Optim.* **23**, 963–999.
[16] Römisch, W. (2010) Scenario generation, in *Wiley Encyclopedia of Operations Research and Management Science*, (ed J. J. Cochran), John Wiley & Sons, Inc., Hoboken.
[17] Dupačová, J., Gröwe-Kuska, N., and Römisch, W. (2003) Scenario reduction in stochastic programming: an approach using probability metrics. *Math. Program.* **95**, 493–511.
[18] Heitsch, H. and Römisch, W. (2009) Scenario tree modeling for multistage stochastic programs. *Math. Program.* **118**, 371–406.
[19] Pennanen, T. (2009) Epi-convergent discretizations of multistage stochastic programs via integration quadratures. *Math. Program.* **116**, 461–479.

DOI: 10.1002/9781118445112.stat08075