Insurability of Catastrophic Risks: The Stochastic Optimization Model

Y.M. Ermoliev
T.Y. Ermolieva
G.J. MacDonald
International Institute for Applied Systems Analysis, Laxenburg, Austria

V.I. Norkin
Institute of Cybernetics, Kiev, Ukraine

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.

INSURABILITY OF CATASTROPHIC RISKS: THE STOCHASTIC OPTIMIZATION MODEL

Y. M. ERMOLIEV[a,]*, T. Y. ERMOLIEVA[a], G. J. MACDONALD[a] and V. I. NORKIN[b]

[a] *IIASA, Laxenburg, Austria;* [b] *Inst. Cybernetics, Kiev, Ukraine*

Catastrophes produce losses highly correlated in space and time, which break the law of large numbers. We derive the insurability of dependent catastrophic risks by calculating conditions that would aid insurers in deliberate selection of their portfolios. This paper outlines the general structure of a basic stochastic optimization model. Connections between the probability of ruin and nonsmooth risk functions, as well as adaptive Monte Carlo optimization procedures and path dependent laws of large numbers, are discussed.

Keywords: Stochastic optimization; catastrophic risk; portfolio selection; ruin probability; nonsmooth risk functions; adaptive Monte Carlo optimization; path-dependent laws of large numbers

AMS 1980 Subject Classification: 90 C 15

1. INTRODUCTION

Catastrophic risks challenge both conventional economic and existing risk (insurance) theory. Although risk is a key concept to characterize future uncertain outcomes of any socio-economic and environmental changes, existing economic theory does not provide appropriate risk-related approaches. The standard economic theory is dominated by the universal power of the price system to reveal all uncertainties and bring the economic system to an efficient equilibrium. Insurers are

---

then able to derive the possibility of pricing and spreading risks through markets over the whole society in such a way that those who have enjoyed gains will compensate those who have suffered losses. Under such assumptions of certainty, catastrophes pose no special problems [1].

Insurance risk theory has developed independently of the fundamental economic ideas [1, 11]. The central problem of this theory is modeling the probability distribution of total future claims [14], which is then used to evaluate ruin probabilities, premiums, reinsurance arrangements *etc.* This theory essentially relies on the assumption of independent, frequent, low-consequence (conventional) risks, such as car accidents, for which decisions on premiums, estimates of claims and likelihood of insolvency (probability of ruin) can be calculated by using rich historical data. The frequent conventional risks also permit simple "trial-and-error" or "learning-by-doing" procedures for adjusting default decisions, for instance, premiums and coverage.

Catastrophes produce claims highly correlated in space and time. The law of large numbers does not operate (in general) and the probability of ruin can be reduced only if insurers deliberately select the dependent catastrophic risks they will cover. So-called catastrophe modeling [15] is becoming increasingly important to insurance companies for estimating dependent catastrophic losses as they analyze alternatives on the allocation of coverage, premiums, reinsurance agreements, and the effects of mitigation measures.

The aim of this paper is to show that the choice of insurance decisions in the presence of catastrophic risks can be regarded as a stochastic optimization problem. Section 2 illustrates the peculiarities of emerging stochastic optimization problems by using a typical model of risk theory. Section 3 discusses a rather general model attempting to bridge decision-oriented economic theory with risk theory and catastrophe modeling. This discussion closely follows papers [5–7]. The concept of risk emphasizes the variability of outcomes, the possibility of gains and, at the same time, the chances of losses. This type of "hit-or miss" situation often leads to nonsmooth models [4], challenging the standard paradigm of smooth utilities and their marginal values. In our model the risk of ruin is modeled by nonsmooth risk functions. Section 4 discusses the connections between the nonsmooth risk functions and the chance constraints. Section 5 and 6 analyze the adaptive Monte Carlo

optimization procedure on the basis of path-dependent laws of large numbers. Section 7 gives concluding remarks.

2. THE STANDARD RISK MODEL

To begin with, let us consider a simple model of an insurance business [14]. The main variable of concern is the risk reserve r^t at time t: $r^t = r_0 + \pi^t - A^t, t \geq 0$, where π^t, A^t are aggregated premiums and claims, and r_0 is the initial risk reserve. The process $A^t = \sum_{k=1}^{N(t)} S_k$, where $N(t), t \geq 0$, is a counting process (*e.g.*, a Poisson process) with $N(0) = 0$, and $\{S_k\}_1^\infty$ is a sequence of independent and identically distributed random variables (claims), in other words, replicates of a random variable S.

Assume that $N(t), S_k$ are independent, $N(t)$ has intensity α, *i.e.*, $\mathbf{E}[N(t)] = \alpha t$, and $\pi^t = \pi t$, $\pi > 0$. Then the expected profit over the interval $[0, t]$ is $(\pi - \alpha \mathbf{E}S)t$, that is, the expected profit increases in time for $\pi - \alpha \mathbf{E}S > 0$. The difference $\pi - \alpha \mathbf{E}S$ is the "safety loading". It follows from the strong law of large numbers that $\lim_{t \to \infty}[\pi^t - A^t]/t = \pi - \alpha \mathbf{E}S$ with probability 1. Therefore, in the case of positive safety loading, $\pi > \alpha \mathbf{E}L$, we have to expect that the real random profit $\pi^t - A^t$ for large enough t would also be positive under the appropriate choice of premium $\pi = (1 + \rho)\alpha \mathbf{E}S$, where ρ is the "relative safety" loading, $\rho = (\pi - \alpha \mathbf{E}S)/\alpha \mathbf{E}S$. This is a basic actuarial principle: premiums are calculated by relying on the mean value of aggregated claims increased by the (relative) safety loading. Thus, practical actuarial approaches ignore complex interdependencies among timing of claims, their sizes, and the possibility of ruin, $r^t \leq 0$. The random jumping process r^t is simply replaced by a linear in t function $\bar{r}^t = r_0 + (\pi - \alpha \mathbf{E}S)t$.

The main problem of the risk theory [14] is the evaluation of the ruin probability $\Psi = \mathbf{P}\{r^t \leq 0 \text{ for some } t, t > 0\}$ under different assumptions on π^t and A^t. There are several cases where Ψ can be explicitly given, or at least given in a form suited for numerical calculations. An important case arises when the claim distribution is a mixture of exponential distributions and claims occur according to a Poisson process. There are numerous approximations for the probability distribution of A^t. Most of them provide satisfactory results only in the area of mean values and cannot be applied to catastrophes.

Various decision variables affect Ψ. Claim size S depends on the coverages of the insurer from different locations. Important decision variables are r_0, π, and reinsurance arrangements, for example, the "excess of loss" reinsurance contract. In this case the insurer retains only a portion, $S(x) = \min\{S, x\}, x \geq 0$, of a claim S and the remaining portion is passed to the reinsurer. The reinsurance contracts with deductibles are defined by two variables $x = (x_1, x_2)$. In this case $S(x) = \max\{x_1, \min[S, x_2]\}, x_1 \geq 0, x_2 \geq 0$ is retained by the insurer. The reduction of Ψ to acceptable levels can be viewed as the well-known chance constraint problem [17, 22]. The complexity is associated with the jumping process A^t.

Consider $t = 0, 1, \ldots$, and assume that r^t can be subdivided into a "normal" part (including r_0) M^t, associated with ordinary claims, and a "catastrophic" part B^t; $\pi^t = \pi t$, where π corresponds to premiums related to catastrophes; the probability of a catastrophic event p is characterized by a probability distribution in an interval $[\underline{p}, \bar{p}]$, and the probability distribution $V_t(z) = \mathbf{P}\{M^t < z\}$ can be evaluated. Assume also that ruin may only occur due to a catastrophe. Then the probability of ruin after the first catastrophe and with the "excess of loss" contract is defined as

$$\Psi(x) = \mathbf{E} \sum_{t=1}^{\infty} p \, (1 - p)^{t-1} V_t(\min\{x, B^t\} - \pi t).$$

The problem is to choose the reinsurance contract x that guarantees a firm will not exceed a given level of ruin, $\Psi(x) = \gamma, \gamma > 0$.

We can define the following stochastic approximation type procedure (see, for example [21] for general definitions). Let x^k be an approximate solution calculated after k simulations, where x^0 is an arbitrary nonnegative value. Step $k + 1$: choose t_k with a probability $\mu_t, \sum_{t=1}^{\infty} \mu_t = 1$, from the set $t \in \{1, 2, \ldots\}$; generate $p_k \in [\underline{p}, \bar{p}]$ and simulate claim $B_k^{t_k}$ by a catastrophe model. Adjust x^k according to the feedback

$$x^{k+1} = \max \left\{ 0, x^k + \frac{\rho}{k+1} \mu_{t_k}^{-1} \big[p \, (1 - p)^{t_k - 1} V_{t_k}(\min\{x^k, B_k^{t_k}\} \right.$$
$$\left. - \pi t^k) - \gamma \big] \right\},$$

where $k = 0, 1, \ldots, \rho$ is a positive constant. The sequence x^k, $k = 0, 1, \ldots$, converges to the desired value since $\mu_{t_k}^{-1} p (1 - p)^{t_k - 1} V_{t_k}$ ($\min\{x^k, B_k^{t_k}\} - \pi t^k$) is an estimate of $\Psi(x^k)$, i.e., its expected value is $\Psi(x^k)$. This type of approach is discussed in Sections 5 and 6 for general problems with many insurers and complex dependent claim processes. The choice of probabilities $\mu_t, t = 0, 1, \ldots$ can regulate the efficiency of the importance sampling (Section 6).

3. THE STOCHASTIC OPTIMIZATION (STO) MODEL

To deal with the insurability of catastrophic risks one should characterize patterns of possible catastrophic events, geographical locations of current and possible new coverages of insurers, available loss reduction (mitigation) measures and insurance-linked securities. Assume that the study region is divided into subregions or locations $j = 1, 2, \ldots, m$. For each location j there exists an estimate W_j^t of the "wealth" at time intervals $t = 0, 1, \ldots$ that includes values of houses, factories, etc. A sequence of random, possibly catastrophic, events $\omega = \{\omega_t, t = 0, 1, \ldots\}$ affects different locations j and generates at each t losses $L_j^t(\omega)$. In the following we do not use a specific structure of the sequence ω: we simply assume that ω is an element of a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where Ω is a set of all possible ω; $\mathcal{F}$ is a σ-algebra of measurable (with respect to probability measure $\mathbf{P}$) events from Ω, and $\{\mathcal{F}_t\}$ is an increasing family of σ-algebras, $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}, \mathcal{F}_t \subseteq \mathcal{F}$. Random variables $L_j^t(\omega)$ are assumed to be $\mathcal{F}_t$ – measurable, that is, they depend on the observable "history" till t. In the following we also assume the existence of all necessary mathematical expectations without specifying the standard requirements each time.

Losses $L_j^t(\omega)$, in contrast to conventional risks, are shared by many participants, such as governments, insurers, reinsurers, banks, and brokers. In the model these are called "insurers".

For each insurer i, the risk reserve r_i^t at time $t = 0, 1, \ldots$ is defined, in general, as $r_i^{t+1} = r_i^t + I_i^t - O_i^t - S_i^t, t \geq 0$, where r_i^0 is a fixed amount of the initial risk reserve; I_i^t represents incomes such as premiums from insurance contracts and other assets; and O_i^t stands for transaction costs and other outcomes, e.g., debts, loans, etc. In

the case of catastrophic risks there are strong dependencies among variables I_i^t, O_i^t, S_i^t defined by patterns of catastrophes and decision variables.

Let us denote by q_{ij}^t a fraction of $L_j^t(\omega)$ covered by insurer $i, \sum_{i=1}^n q_{ij}^t \leq 1, q_{ij}^t \geq 0$. Then the claim process can be written as $S_i^t = \sum_{j=1}^m q_{ij}^t L_j^t$. The decision variables q affect also I_i^t, O_i^t. For example, the income of insurer i from premiums π_i^t at time t and the outcome from transactions costs C_i^t may be linear functions of q. $\pi_i^t = \sum_{j=1}^m \pi_{ij}^t q_{ij}^t$, $C_i^t = \sum_{j=1}^m C_{ij}^t q_{ij}^t$. The claim process S_i^t has a more complex, nonsmooth structure in the case of reinsurance arrangements. Thus insurer i retains only the part $S_i^t = \min \left\{ u_{il}, \sum_{j=1}^n q_{ij}^t L_j^t \right\}$ in the case of "excess of loss" contract with reinsurer l, which is defined by the decision variables u_{il}. The reinsurer l absorbs the claim $S_i^t = \sum_i \max \left\{ 0, \sum_{j=1}^n q_{ij}^t L_j^t - u_{il} \right\}$, where the sum is taken with respect to all insurers having this type of contract with l. Apart from decision variables q, the distribution of L_j^t may depend on mitigation measures in location j. Variables I^t, O_i^t are also affected by decisions on other assets. Let us define all possible ex-ante decision variables by a vector x.

If W_j^0 is the initial wealth, then location j's wealth at time $t+1$ is $W_j^{t+1} = W_j^t + I_j^t - O_j^t - L_j^t$. We assume that the "aggregate individual" from j maximizes his expected wealth W_j^t, taking into account the risk of underestimating losses, $\min \left[0, W_j^t - \mathbf{E}W_j^t \right]$, and insolvency, $W_j^t < 0$. Therefore assume that decisions x are chosen from the maximization of the expectation function $H_j(x) = \mathbf{E}h_j^{\tau_j}(x, \omega), h_j^t = W_j^{t-1} + \gamma_j \min \left[0, W_j^{t-1} - \mathbf{E}W_j^{t-1} \right] + \phi_j \min \left[0, W_j^t \right]$, where γ_j, ϕ_j are substitution coefficients (or risk coefficients) between possible wealth, the risk of underestimating losses, and insolvency; τ_j is a stopping time, for example, the time of ruin not exceeding T, $\tau_j = \min \left[T, \min \left\{ t : W_j^t \leq 0, t > 0 \right\} \right]$. Similarly, insurer i maximizes his expected wealth r_i^t, taking into account the risk of overestimating profits and the risk of insolvency (ruin), that is, function $D_i(x) = \mathbf{E}d_i^{\lambda_i}(x, \omega), d_i^t = r_i^{t-1} + \varepsilon_i \min \left[0, r_i^{t-1} - \mathbf{E}r_i^{t-1} \right] + \delta_i \min \left\{ 0, r_i^t \right\}$, where ε_i, δ_i are substitution coefficients between profit and the risk of overestimating profits, $\min \left[0, r_i^t - \mathbf{E}R_i^t \right]$, and insolvency; λ_i is a stopping time. e.g., $\lambda_i = \min \left[T, \min \{ t : r_i^t \leq 0, t > 0 \} \right]$.

The problem is to analyze the Pareto optimal improvement of the risk situation with respect to goal functions $H_j(x), D_i(x)$ by maximizing $W(x) = \sum_{j=1}^m \alpha_j H_j(x) + \sum_{i=1}^n \beta_i D_i(x)$, where $\alpha_j \geq 0$, $\beta_i \geq 0$,

$\sum_{j=1}^{m} \alpha_j + \sum_{i=1}^{n} \beta_i = 1$. This problem is similar to economic welfare programs [13]. The general idea of this model is also close to the model proposed in [3] for designing portfolios of callable bonds. Let $w(x, \omega) = \sum_{j=1}^{m} \alpha_j h_j^{\tau_j}(x, \omega) + \sum_{i=1}^{n} \beta_i d_i^{\lambda_i}(x, \omega)$, then $W(x)$ can be written as $W(x) = \mathbf{E} w(x, \omega)$.

Remark 3.1 In general, functions H, D can be replaced by $H_j(x) = \mathbf{E} h_j(x, W_j^{0, \tau_j}, \mathbf{E} W_j^{0, \tau_j}, \omega), D_i(x) = \mathbf{E} d_i(x, r_i^{0, \lambda_i}, \mathbf{E} r_i^{0, \lambda_i}, \omega)$, where $a^{0,t} = (a^0, a^1, \dots, a^t)$ for a sequence $a^0, a^1, \dots$ Therefore, the maximization of $W(x)$ is an STO problem of the following form: maximize $F(x) = \mathbf{E} f(x, \mathbf{E} g(x, \omega), \omega)$.

4. PROBABILITY OF RUIN

$W(x)$ includes nonsmooth risk functions to model the need for risk management. The use of these risk functions corresponds to the Markovitz mean-semivariance model [16]. In [20] it was shown that the use of absolute deviations with the appropriate choice of risk coefficients is consistent with the stochastic dominance of random outcomes. The applicability of the well-known mean-variance model [16] is usually linked to the normality of the probability distribution summarizing different prospects, which cannot be assumed for catastrophic risks. The following theorem shows that if risk coefficients δ_i become large enough, then the probability of ruin drops below a given level [5].

The function $W(q)$ can be represented in the form $W(q) = V(q) + \mathbf{E} \sum_{i=1}^{n} \beta_i \delta_i \min \{0, r_i^{\lambda_i}\}$. If $\delta_i = N/\beta_i$, then $W(q) = V(q) + N \mathbf{E} \sum_{i=1}^{n} \min\{0, r_i^{\lambda_i}\}$. If N is large enough, then maximization of $W(q)$ approximates the maximization of $V(q)$ subject to the chance constraints $\mathbf{P} \{ \sum_{i=1}^{n} \min \{0, r_i^{\lambda_i}\} < 0 \} < \varepsilon$ for arbitrarily small $\varepsilon > 0$. This is due to the following general result, which, for linear chance constraints was, in fact, discussed in [25].

Consider two general problems: the chance constraint problem $F(x) \rightarrow \max_{x \in X}, \mathbf{P}\{g(x, \omega) < 0\} \le \varepsilon$, with optimal value F_ε^* and the problem $\Phi_N(x) = F(x) + NG_+(x) \rightarrow \max_{x \in X}$, with optimal value Φ_N^*. Here X is a compact set, $F(x)$ is a continuous function, $G_+(x) = \mathbf{E} \max\{0, g(x, \omega)\}$, and N is a penalty coefficient.

THEOREM 4.1   *Assume* [5] *that* $g(\cdot, \omega)$ *is almost sure (a.s.) continuous and* $|g(x, \omega)| \leq C(\omega)$ *for all* $x \in X$, $EC^{1+\lambda}(\omega) \leq C^{1+\lambda} < +\infty$, *for some* $C > 0, \lambda > 0$; $G_+(x') = 0$ *for some* $x' \in X$; $P\{g(x, \omega) = 0\} = 0 \ \forall x \in X$. *Then there exist non-negative functions* $\varepsilon(N), \alpha(N), \beta(\varepsilon)$ *and* $\gamma > 0$ *such that* $lim_{N \to +\infty} \varepsilon(N) = lim_{N \to +\infty} \alpha(N) = lim_{\varepsilon \to 0} \beta(\varepsilon) = 0$,

$$\Phi_N^* - \alpha(N) \leq F_{\varepsilon(N)}^* \leq \Phi_{1/\varepsilon^\gamma(N)}^* - \beta(\varepsilon(N)),$$
$$F_{1/N^{1/\gamma}}^* + \beta(1/N^{1/\gamma}) \leq \Phi_N^* \leq F_{\varepsilon(N)}^* + \alpha(N).$$

In other words, there always exist large enough N so that the maximization of penalty function $\Phi_N(x)$ generates optimal solutions, which also satisfy the chance constraints for any given $\varepsilon > 0$. The proof of this theorem provides some insights on the reasonable choice of N. This value can easily be adjusted in the process of calculations [6, 7] by using histograms of random outcomes $g(x, \omega)$.

5. THE ESTIMATION OF SUBGRADIENTS

The maximization of $W(x)$ has the following form. Let $f^t(x, \omega)$, $\zeta^t(x, \omega), t = 0, 1, \ldots$ be real-valued, random $\mathcal{F}_t$ – measurable functions, which guarantee the existance of necessary expectations; $\tau(x, \omega) = \min[T, \min\{t : \zeta^t(x, \omega) \leq 0, t > 0\}]$. Then the maximization of $W(x)$ can be viewed as the maximization of a function $F(x) = Ef(x, \omega), f(x, \omega) = f^\tau(x, \omega)$. If $f^t(x, \omega)$ are concave in x, then $F^t(x) = Ef^t(x, \omega)$ are also concave, but not $F(x)$ due to the dependence of $\tau(x, \omega)$ on x. The class of so-called generalized differentiable (GD)-functions is important for problems involving min, max operations as in "excess of loss" reinsurance contracts.

DEFINITION 5.1   Function $f: R^n \to R$ is called generalized differentiable at $x \in R^n$ if in some vicinity of x there exists an upper semicontinuous at x multivalued (subdifferential) mapping ∂f with closed convex compact values $\partial f(x)$ such that $f(y) = f(x) + \langle g, y - x \rangle + o(x, y, g)$, where $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors, $g \in \partial f(y)$ and $lim_k |o(x, y^k, g^k)|/\|y^k - x\| = 0$ for any sequences $y^k \to x, g^k \in \partial f(y^k)$. Function f is called generalized differentiable if it is generalized differentiable at each point $x \in R^n$.

It is important [19] that continuously differentiable, convex or concave functions are generalized differentiable: class GD-functions is closed with respect to max, min operations and superpositions, there is calculus of subgradients:

$$\partial \min(f_1, f_2)(x) = co\{\partial f_i | f_i(x) = \min(f_1(x), f_2(x))\}, \qquad (5.1)$$

where $co\{\cdot\}$ denotes a convex hull of $\{\cdot\}$; the subdifferential $\partial f_0(f_1, \ldots, f_m)$ of a composite function $f_0(f_1, \ldots, f_m)$, where $f_0(\cdot)$ is a GD-function, is calculated by the chain rule. In addition, the class of GD-functions is closed with respect to expectations. It is easy to see that $f^{\tau(x,\omega)}(x, \omega)$ may be function discontinuous in x for continuous $f^t(\cdot, \omega), \zeta^t(\cdot, \omega), t = 0, 1, \ldots$. Therefore, in general, the maximization of $F(x)$ requires the use of stochastic mollifiers [4]. For many important cases, for example, when τ does not depend on x, and $f^t(\cdot, \omega)$ are GD-functions, it is possible to show that $f^{\tau(x,\omega)}(x, \omega)$ is a GD-function and to derive the following estimator of a subgradient F_x:

THEOREM 5.2   *Assume* [5] *that X is a compact set and*

(i) *functions* $\zeta^t(x, \omega)$, $t = 0, 1, \ldots$ *are continiously differentiable in* $x \in X$ *for almost all* ω *and* $\sup \{|\zeta^t(x, \omega)| \ |x \in X\} \le L(\omega)$, *with integrable function* $L(\omega)$;
(ii) *generalized gradient (in x) mappings* $\partial f^t(x, \omega)$ *are measurable in* ω *and bounded by* $L(\omega)$ *for all* $x \in X$;
(iii) *for all* $x \in X$ *and* $t = 0, 1, \ldots$ *the probability* $P\{\zeta^t(x, \omega) = 0\} = 0$.

Then $\partial F(x) = E \partial f(x, \omega)$, *where*

$$\partial f(x, \omega) = \partial f^t(x, \omega)|_{t=\tau}. \qquad (5.2)$$

Assume now $X = \{x | \Psi(x) \le 0\}$, where Ψ is a GD-function, inf $\{\|g\| : g \in \partial \Psi(x)\} > 0$; $F(x) = E f(x, \omega)$, where $f(\cdot, \omega)$, $F(\cdot)$ are GD-functions. The following key result was proved in [8]. Consider the stochastic quasi-gradient (SQG) procedure:

$$x^{k+1} \in \Pi_X(x^k - \rho_k \xi^k), x^0 \in X, k = 0, 1, \ldots,$$
$$E\{\xi^k | x^0, \ldots, x^k\} \in \partial f(x^k, \omega^k), \qquad (5.3)$$

where $\Pi_X(y)$ is the (multivalued) projection of y on X, $\Pi_X(y) = \arg \min\{\|y - x\|^2, \ x \in X\}$; $\rho_k \ge 0, \sum_{k=0}^{\infty} \rho_k = \infty, \sum_{k=0}^{\infty} \rho_k^2 < \infty$; ω^k is a

sequence of independent observations (simulations) of ω. Define $X^* = \{x | 0 \in \partial F(x) + N_X(x)\}$, where $N_X(x) = \{u \partial \Psi(x) : u \geq 0\}$ if $\Psi(x) = 0$ and $N_X(x) = \emptyset$ if $\Psi(x) < 0$. Let X be a compact and $\|\xi^k\| \leq C < \infty$ (this usually follows from the compactness of X).

THEOREM 5.3   *All cluster points of* $\{F(x^k)\}$ *a.s. constitute an interval in* F^*. *If set* F^* *does not contain intervals (for example,* F^* *is finite or countable), then all cluster points of* $\{x^k\}$ *a.s. belong to a connected subset of* X^* *and* $\{F(x^k)\}$ *a.s. has a limit in* F^*.

From Theorem 5.3 follows the basic SQG procedure for maximizing $W(x)$. Assume that after k independent simulations of events $\omega^0, \ldots, \omega^{k-1}$ from $(\Omega, \mathcal{F}, \mathbf{P})$ we obtain an approximate solution x^k. Theorem 5.2 provides the following simple rule for calculating ξ^k for $F(x) := W(x)$. For given x^k simulate ω^k from $(\Omega, \mathcal{F}, \mathbf{P})$ independently of $\omega^0, \ldots, \omega^{k-1}$; observe stopping times $\tau_j^k = \tau_j(x^k, \omega^k)$, $\lambda_i^k = \lambda_i(x^k, \omega^k)$, and calculate subgradients of functions $h_j^t(x, \omega(k))$, for $t \leq \tau_j^k$, and $d_i^t(x, \omega(k))$, for $t \leq \lambda_i^k$. Compute

$$\xi^k = \sum_{j=1}^{m} \alpha_j h_{jx}^t(x^k, \omega^k)|_{t=\tau_j^k} + \sum_{i=1}^{n} \beta_i d_{ix}^t(x^k, \omega^k)|_{t=\lambda_i^k}. \qquad (5.4)$$

After that a new x^{k+1} is adjusted according to (5.3), and so on. The implementation of this basic procedure requires at least the exact calculation of functions h, d, which may be impossible for general cases. Let us consider this in more detail.

6. ADAPTIVE MONTE CARLO OPTIMIZATION

The maximization of $W(x)$, in general, is regarded as estimating the maximum value F^* of the integral (see Remark 3.1)

$$F(x) = \int f(x, \mathbf{E}g(x, \omega), \omega) \mathbf{P}(d\omega) \qquad (6.1)$$

and a corresponding optimal solution x^* from a subset X. "Adaptive Monte Carlo" usually means [23] a technique that makes on-line use of sampling information to improve sequentially the efficiency of the sampling itself. We use "adaptive Monte Carlo optimization" in a

rather broad sense, *i.e.*, when the efficiency of the sampling procedure is considered as a part of more general improvements with respect to different decisions and goals. The adaptive Monte Carlo optimization problem arises when the probability measure **P** and/or the sample function f are not explicitly known, but known only in terms of other explicitly known measures and functions.

The random function $f(x, \mathbf{E}g(x, \omega), \omega)$ in (6.1) may not be analytically tractable even for simple risk management problems (Section 2). The essential complexity is associated with the calculation of expectation $\mathbf{E}g(x, \omega)$. Assume that $f(x, \nu, \omega), g(x, \omega)$ are GD-functions with respect to x, ν, and $\mathcal{F}$ – measurable with respect to ω. A stochastic subgradient of $F(x)$, that is, an estimate of subgradient F_x, is $f_x(x, \mathbf{E}g(x, \omega), \omega) + f_\nu(x, \nu, \omega)|_{\nu = \mathbf{E}g(x, \omega)} g_x(x, \theta)$, where ω, θ, are independent samples from **P**. Unfortunately, we cannot use this formula directly because expectation $\mathbf{E}g$ is not explicitly known. Consider the sequence of estimates x^k, generated by (5.3) with

$$\xi^k = f(x^k, \nu^k, \omega^k) + f_z(x^k, \nu^k, \omega^k) g_x(x^k, \omega^k), \qquad (6.2)$$

where ν^k is an estimate of $E[g(x^k, \omega) \mid x^k]$ defined by the rule

$$\nu^{k+1} = \nu^k + \sigma_k(g(x^k, \omega^k) - \nu^k), \;\; z^0 = 0, k = 0, 1, 2, \ldots. \qquad (6.3)$$

Let us note that if $\sigma_k = 1/k + 1$, then $\nu^k = k^{-1} \sum_{s=1}^k g(x^s, \omega^s)$.

Assume that sequence $x^k, k = 0, 1, \ldots$, converges with probability 1. The convergence of ν^k to $E\{g(x^k, \omega)|x^k\}$, $\nu^k - E\{g(x^k, \omega)|x^k\} \to 0$, with probability 1, is then derived from known results [12, 18] on the law of large numbers for dependent random variables. Unfortunately, the convergence of $\{x^k\}$ (defined by (5.3), (5.4), (6.3)) itself is derived only from the convergence $\nu^k - E\{g(x^k, \omega) \mid x^k\} \to 0$, that is, if estimates ν^k "track" $E\{g(x^k, \omega)|x^k\}$ without assumption of the convergence of $\{x^k\}$. This is typical situation for the nonstationary optimization [10]. The following general theorem [9] shows that ν^k is able to track $E\{g(x^k, \omega)|x^k\}$ without the convergence of x^k.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space with a flow of nondecreasing σ-algebras $\mathcal{F}_k \subseteq \mathcal{F}_{k+1} \subseteq \mathcal{F}, k = 1, 2, \ldots; \zeta^k(\omega) : \Omega \to R^n$ are measurable with respect to $\mathcal{F}_k, k = 1, 2, \ldots, z^k(\omega) = E\{\zeta^k(\omega)|\mathcal{F}_{k-1}\}$. Consider the

following estimators:

$$\bar{\zeta}^{k+1}(\omega) = \Pi_Z(\bar{\zeta}^k(\omega) - \sigma_k(\bar{\zeta}^k(\omega) - \zeta^k(\omega))), \quad k = 0, 1, \ldots, \qquad (6.4)$$

$$\bar{z}^{k+1}(\omega) = \Pi_Z(\bar{z}^k(\omega) - \sigma_k(\bar{z}^k(\omega) - z^k(\omega))), \quad k = 0, 1, \ldots, \qquad (6.5)$$

where $\bar{\zeta}^0 = 0, \bar{z}^0 = 0, Z$ is a convex compact set from $R^n, \Pi_Z(y)$ is the projection of y on Z and random variables σ_k are $\mathcal{F}_{k-1}$ – measurable.

THEOREM 6.1   *Assume* [9] *for some* $0 < \varepsilon \leq 1$

$$0 \leq \sigma_k \leq 1, \lim_k \sigma_k = 0, \sum_{k=0}^{\infty} \sigma_k = \infty \sum_{k=0}^{\infty} \sigma_k^{1+\varepsilon} < +\infty \ a.s. :$$

$$E\|\zeta^k - z^k\|^{1+\varepsilon} \leq C < \infty.$$

Then $\lim_k(\bar{\zeta}^k(\omega) - \bar{z}^k(\omega)) = 0$ *a.s. Suppose additionally that*

$$\lim_k \|z^{k+1} - z^k\|/\sigma_k = 0 \ a.s.$$

Then $\lim_k(\bar{z}^k - \Pi_Z(z^k) = 0$ *a.s. and hence* $\lim_k(\bar{\zeta}^k(\omega) - \Pi_Z(z^k(\omega)) = 0$ *a.s.*

Thus, if $z^k(\omega) \in Z$, then estimator $\bar{\zeta}^k$ a.s. tracks a moving mean z^k as $k \to \infty$. Consider now again the procedures (5.3), (6.2), (6.3). We can always assume that $\|x^{k+1} - x^k\| \leq C_{\rho k}$ for a constant C. If $\lim_k \rho k / \sigma_k = 0$, then $|g^k - Eg(x^k, \omega)| \to 0$ for $k \to \infty$ (from Theorem 6.1). Therefore, this justifies the use of $f(x^k, \nu^k, \omega^k)$ as an estimate of $f(x^k, Eg(x^k, \omega), \omega^k)$ and ξ^k from (6.2) as an estimate of $F_x(x^k)$. The full convergence analysis of (5.3), (6.2) is similar to [8].

The fast simulation of rare catastrophes and the variance reduction of estimates $W(x^k), k = 0, 1, \ldots$ can be achieved by importance sampling. A general idea of adaptive improvement of the sampling procedure was introduced by Pugh [23]. Unfortunately, the proposed approach itself requires the additional estimation of some integrals. Stochastic optimization procedures of type (5.3), (6.2), (6.3) allow for sequential variance reduction without major additional computations.

The general idea is as follows [7]. Consider a probability measure $\mu(\omega)$ on the domain of $\mathbf{P}$ such that $W(x) = \int w(x, \omega)(d\mathbf{P}/d\mu)d\mu(\omega)$, where $(d\mathbf{P}/d\mu)$ is the Radon-Nycodym derivative [18]. The random variable $w(x, \omega)$ $(d\mathbf{P}/d\mu)$ is an estimate of $W(x)$. The variance of this estimator is $\Phi := \int w^2(x, \omega)(d\mathbf{P}/d\mu)^2 d\mu(\omega) - W^2(x)$. The aim is to find a μ that minimizes Φ.

Let the family of distributions μ be indexed by the vector parameter y. Then $\Psi, (d\mathbf{P}/d\mu)$ are functions of y. Let us assume that they are continuously differentiable. It is easy to see that $\partial\Phi/\partial y_l = 2 \int w^2(d/d_{yl})$ $(d\mathbf{P}/d\mu)d\mathbf{P} = 2 \int w^2 (d/dy_l) (d\mathbf{P}/d\mu) (d\mathbf{P}/d\mu) d\mu$. Together with procedures (5.3), (6.2), (6.3) consider a sequence of measures $\mu_k := \mu(y^k, \omega)$ defined by the sequence of vectors $\{y^k\}, y^{k+1} = y^k - \phi_k w^2(x^k, \bar{\omega}^k)$ $\mathrm{grad}(d\mathbf{P}/d\mu)|_{y=y^k}(d\mathbf{P}/d\mu)$, where $\bar{\omega}^k$ is a sample from μ_k; ϕ_k satisfies the same conditions as ρ_k. The convergence of the resulting procedure easily follows because $W(x)$ does not depend on y.

7. CONCLUDING REMARKS

Numerical experiments with the proposed model show a satisfactory convergence of the methods analyzed. The model has a rather general form suitable for so-called integrated catastrophe risk management [24], in other words, for the analysis of location-specific risk reduction measures combined with different risk spreading options. It takes into account differences in vulnerability between various insurance port-folios and geographically explicit, dependent losses from events occur-ing at different locations. Paper [24] discusses the need for integrated risk management as the natural further development of catastrophe modeling.

Our model can be used by one insurer or a pool of insurers. The importance of cooperative efforts of insurers ("pooling" of risks) was emphasized by Borch [2]. In contrast to [2], the model includes trans-action costs and deals with non-substitutable catastrophic risks. The solution of the resulting stochastic optimization problem by (5.3) can be organized in a decentralized manner by using a computer's network connecting all insurers.

The maximization of functions H_j, D_j generates the insurance demand and supply functions, which depend on premiums. The choice

of premiums must reflect the balance between insurance demand and supply that calls for an appropriate concept of stochastic equilibrium. This requires special attention.

References

[1] Arrow, K. J. (1996). The theory of risk-bearing: small and great risks, *Journal of Risk and Uncertainty*, **12**, 103–111.
[2] Borch, K. (1974). The mathematical theory of insurance, Lexington Books, Lexington.
[3] Consinglio, A. and Zenious, S. (1998). Designing Portfolios of financial products *via* integrated simulation and optimization models, University of Cyprus, Revised Report 96–05.
[4] Ermoliev, Y. M. and Norkin, V. I. (1997). On nonsmooth and discontinuous problems of stochastic systems optimization, *European Journal of Operation Research*, **101**, 230–244.
[5] Ermoliev, Y. M., Ermolieva, T. Y., MacDonald, G. J. and Norkin, V. I. (1998). On the design of catastrophic risk portfolios, IIASA Interim Report IR-98-056, Web: www.iiasa.ac.at.
[6] Ermolieva, T. Y., Ermoliev, Y. M. and Norkin, V. I. (1997). Spatial Stochastic Model for Optimization Capacity of Insurance Networks Under Dependent Catastrophic Risks: Numerical Experiments. IIASA Interim Report IR-97-028, Web: www.iiasa.ac.at.
[7] Ermolieva, T. Y. (1997). The Design of Optimal Insurance Decisions in the Presence of Catastrophic Risks, IIASA Interim Report IR-97-068, Web: www.iiasa.ac.at.
[8] Ermoliev, Y. and Norkin, V. (1997). Stochastic generalized gradient method with application to insurance risk management, IIASA Interim Report IR-97-021, Web: www.iiasa.ac.at.
[9] Ermoliev, Y. and Norkin, V. (1998). Monte Carlo Optimization and Path Dependent Nonstationary Laws of Large Numbers, IIASA Interim Report IR-98-009, Web: www.iiasa.ac.at.
[10] Gaivoronski, A. A. (1978). On nonstationary stochastic optimization problems, *Cybernetics*, **14**.
[11] Giarini, O. and Louberg, H. (1978). The Diminishing Returns of Technology, Pergamon Press, Oxford.
[12] Gihman, I. I. and Skorohod, A. V. (1971). Theory of random processes. **1**, Nauka, Moscow.
[13] Ginsburg, V. and Kaizer, M. A. (1997). The Structure of Applied General Equilibrium Models, Cambridge MA: MIT Press.
[14] Grandell, J. (1991). Aspects of Risk Theory, Springer Series in Statistics: Probability and Its Applications, Springer Verlag, New York / Berlin / Heidelberg.
[15] Insurance Service Office, The Impact of Catastrophes on Property Insurance, New York, 1994.
[16] Markowitz, H. M. (1987). Mean Variance Analysis in Portfolio Choice and Capital Markets, Blackwell, Oxford.
[17] Marti, K. (1994). Approximations and derivatives of probability functions, In: *Approximation, Probability and Related Fields* (Edited by Anastassion, G. and Rachev, S.), Plenum Publishing Corp., New York. pp. 367–377.
[18] Neveu, J. (1965). Mathematical Foundations of the Calculous of Probabilities, San Francisco: Holden-Day.

[19] Norkin, V. I. (1980). Generalized differentiable functions, *Kibernetika*, **1**, 9 – 11 (In Russian, English translation in *Cybernetics*, **16**, 1).

[20] Ogryczaki, W. and Ruszczynski, A. (1997). From Stochastic Dominance to Mean-Risk Models: Semideviations as Risk Measures, IIASA Interim Report 97 – 027, Web: www.iiasa.ac.at.

[21] Pflug, G. (1996). Optimization of stochastic models: the interface between simulation and optimization, Kluwer Academic Publishers, Boston/Dordrecht/London.

[22] Prekopa, A. (1988). Numerical solution of probabilistic constrained programming problems, In: Ermoliev, Y. and Wets, R. (Eds.), Numerical Techniques for Stochastic Optimization, pp. 123 – 139, Springer Verlag, Berlin.

[23] Pugh, E. L. (1966). A gradient technique of adaptive Monte Carlo, *SIAM Review*, **8**(3).

[24] Wall, G. R. (1997). Current developments in catastrophe modeling, In: Britton, N. and Oliver, J. (Eds.), Financial Risk Management for Natural Catastrophes, *Proceedings of a Conference sponsored by Aon Group Australia Limited*.

[25] Wets, R. J.-B. (1996). Challenges in stochastic programming, *Math. Progr.*, **75**, 115 – 135.