Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

ROBUST RARE-EVENT PERFORMANCE ANALYSIS WITH NATURAL NON-CONVEX CONSTRAINTS

Jose Blanchet

Industrial Engineering and Operations Research
Columbia University
340 W. Mudd Building
500 W 120th Street
New York, NY 10027, USA

Christopher Dolan

Department of Statistics
Columbia University
1021 School of Social Work
1255 Amsterdam Avenue
New York, NY 10027, USA

Henry Lam

Department of Mathematics and Statistics
Boston University
111 Cummington Mall
Boston, MA 02215, USA

ABSTRACT

We consider a common type of robust performance analysis that is formulated as maximizing an expectation among all probability models that are within some tolerance of a baseline model in the Kullback-Leibler sense. The solution of such concave program is tractable and provides an upper bound which is robust to model misspecification. However, this robust formulation fails to preserve some natural stochastic structures, such as i.i.d. model assumptions, and as a consequence, the upper bounds might be pessimistic. Unfortunately, the introduction of i.i.d. assumptions as constraints renders the underlying optimization problem very challenging to solve. We illustrate these phenomena in the rare event setting, and propose a large-deviations based approach for solving this challenging problem in an asymptotic sense for a natural class of random walk problems.

1 INTRODUCTION

Robust performance analysis is concerned with the problem of evaluating the worst case performance measure of interest (typically described as an expectation) among all plausible probability models, such as those within certain tolerance of a baseline model which is believed to be reflective of reality. Taken literally, this problem formulation can be challenging because it gives rise to an infinite dimensional optimization problem (note that we mentioned "all models" within certain tolerance). When the tolerance region is described in terms of Kullback-Leibler divergence (and other related notions; see, for example, Pardo 2005), this apparently daunting optimization problem is often tractable, and this tractability feature has been exploited in a range of literature in recent years, for example in control theory (Iyengar 2005; Nilim and El Ghaoui 2005; Petersen, James, and Dupuis 2000), distributionally robust optimization (Ben-Tal et al. 2013), finance (Glasserman and Xu 2014), economics (Hansen and Sargent 2008) and queueing (Jain, Lim, and Shanthikumar 2010).

Tolerance regions based on the Kullback-Leibler divergence, however, fail to incorporate information that is often quite natural to assume in common stochastic settings, and that should be added in terms of constraints in the underlying robust performance analysis formulation. One such natural and important constraint is the i.i.d. property, often arising in models involving random walk input. Failing to inform the i.i.d. property even in simple situations involving random walk models can have important consequences in terms of the accurate assessment of worst case performance measures of interest.

Unfortunately, however, a robust formulation in which the i.i.d. property is added as an extra constraint on top of the Kullback-Leibler imposed tolerance gives rise to an optimization problem which is no longer easy to handle.

The main contribution of this paper is to show that in the context of performance analysis associated a class of large deviations events, such robust formulation gives rise to a problem for which asymptotically optimal solutions can be constructed. We illustrate our ideas in the setting of i.i.d. random walks.

The rest of the paper is organized as follows. In Section 2 we provide a precise mathematical formulation of the robust performance analysis problem with i.i.d. constraints and explain why the problem is very challenging. In Section 3 we provide a strategy that allows to solve this challenging problem asymptotically in a large deviations regime. Finally, we provide numerical examples which illustrate the performance of our proposed solution and the impact of adding i.i.d. constraints in the robust formulation.

2 PROBLEM FORMULATION

Let $\{X_k : k \geq 0\}$ be a sequence of zero mean i.i.d. random variables. Define $S_0 = 0$ and put $S_n = X_1 + \cdots + X_n$. Let us use $F(\cdot)$ to denote the CDF (Cumulative Distribution Function) of X_i, that is, $P(X_i \leq x) = F(x)$ and we use $P_F(\cdot)$ to denote the product measure generated by $F(\cdot)$. We use $P_F^n(\cdot)$ to denote the projection of $P_F(\cdot)$ onto its n first coordinates. Simply put, P_F^n describes the joint distribution of the random variables $(X_1, \ldots, X_n)$. The expectation operator associated to $P_F(\cdot)$ and $P_F^n(\cdot)$ is denoted by $E_F(\cdot)$ and $E_F^n(\cdot)$, respectively. We define $\psi_F(\theta) = \log E_F^1 \exp(\theta X_1)$ and assume that $\psi_F(\theta) < \infty$ for θ in a neighborhood of the origin.

Now, define $A_n = \{S_n/n \in A\}$ for a closed set A which does not contain the mean of X_k. We are concerned with the problem of estimating $P_F(A_n)$. Observe that $P_F(A_n) \to 0$ as $n \to \infty$ because of the law of large numbers. Moreover, because $\psi_F(\cdot)$ is finite in a neighborhood of the origin we have that $P_F(A_n) \leq \exp(-\delta n)$ for some $\delta > 0$ for all n sufficiently large.

In contrast to standard rare event estimation problems, however, here we assume that $F(\cdot)$ is unknown. Nevertheless, based on some evidence (for example based on data or expert knowledge) let us assume that we have obtained a CDF $G(\cdot)$, which approximates $F(\cdot)$ in a suitable sense, for example in the Kullback-Leibler sense which we shall review momentarily. Let us write $P_G(\cdot)$ to denote the product measure associated to $G(\cdot)$ and we use $E_G(\cdot)$ for the expectation operator corresponding to $P_G(\cdot)$. Similarly as before, $P_G^n(\cdot)$ is the projection of $P_G(\cdot)$ onto its n first coordinates and we use $E_G^n(\cdot)$ to denote the expectation operator associated to $P_G^n(\cdot)$.

We assume that the likelihood ratio dP_F^n/dP_G^n is well defined and therefore the Kullback-Leibler divergence of P_F^n with respect to P_G^n is defined via

$$R(P_F^n||P_G^n) = E_F^n \log\left(\frac{dP_F^n}{dP_G^n}\right) = nE_F^1 \log\left(\frac{dP_F^1}{dP_G^1}(X_1)\right) = n\int \log\left(\frac{dF}{dG}(x)\right) dF(x).$$

If dP_F^n/dP_G^n fails to exist (i.e. P_F^n is not absolutely continuous with respect to P_G^n), then the Kullback-Leibler divergence is defined as infinity. It is elementary to verify that $R(\cdot\,||P_G^n)$ is convex (actually $R(\cdot||\cdot)$ is convex in both of its arguments; Dupuis and Ellis 2011). The associated robust performance analysis problem with Kullback-Leibler constraint consists in solving

$$\max_{Q^n}\{Q^n(A_n) : R(Q^n||P_G^n) \leq \eta_n\}, \tag{1}$$

where η_n should be chosen to satisfy

$$\eta_n \approx n \int \log\left(\frac{dF}{dG}(x)\right) dF(x).$$

One might select η_n by estimating $\int \log(dF/dG(x))\,dF(x)$ using available data.

The optimization problem (1) is a concave mathematical program; the objective function to maximize is linear (in particular concave) in the variable Q^n and, as mentioned earlier, the constraint is convex. Moreover, as we shall see in the body of the paper (see equations (5) and (6)), the optimal solution to (1), $Q_*^n(\cdot)$, can be characterized as a suitable mixture between $P_G^n(\cdot|A_n)$ and $P_G^n(\cdot|A_n^c)$. As the next result shows, it turns out that $Q_*^n(A_n)$ might differ substantially from $P_F^n(A_n)$ even if F is close to G in the Kullback-Leibler sense, that is, even in cases in which $\eta_n = o(n)$ as $n \to \infty$. In more detail, typically we will have $P_F^n(A_n) = \exp(-n\bar{I}_F(A) + o(n))$, for some positive constant $\bar{I}_F(A)$, whereas the next result indicates that typically $Q_*^n(A_n) \geq \delta\eta_n/n$ for some $\delta > 0$ and large enough n. So, for example, if one builds an approximation G to F from data, one would need an exponentially large sample size (in n) in order to obtain an accurate estimate of the probability of interest using only the relative entropy constraint without recognizing that the data might have come from an i.i.d. model.

Theorem 1 Suppose that $\eta_n = o(n)$ and that $\eta_n > \delta > 0$ for some $\delta > 0$ uniformly over n. Assume also that $P_G^n(A_n) \in (\exp(-n/\delta'), \exp(-\delta' n))$ for some $\delta' > 0$ and all n sufficiently large. Then the optimal value of (1), $Q_*^n(A_n)$, satisfies

$$Q_*^n(A_n) = \frac{\eta_n}{-\log P_G^n(A_n)}(1 + o(1))$$

as $n \to \infty$.

One of the main reasons for such a disparity, as we shall establish in the next section, is that the feasible region (i.e. $\{Q^n : R(Q^n || P_G^n) \leq \eta_n\}$) fails to recognize that we are interested only in models for which the i.i.d. property of the X_i's is preserved. So, introducing the i.i.d. constraint transforms problem (1) into the alternative form

$$\max_H \left\{ P_H^n(A_n) = \int \cdots \int I\left(\frac{x_1 + \cdots + x_n}{n} \in A\right) dH(x_1) \cdots dH(x_n) : n \int \log\left(\frac{dH}{dG}(x)\right) dH(x) \leq \eta_n \right\}. \quad (2)$$

Observe that the previous problem is not a concave program because the objective function to maximize is no longer concave. Unfortunately, in general (2) is very challenging to solve. In the next secion we explain how to use large deviations theory to solve problem (2) in an asymptotic sense. We finish this section with a proof of our first theorem.

2.1 Proof of Theorem 1

To solve (1), we rewrite it in terms of the likelihood ratio between Q^n and G^n, namely $L = dQ^n/dP_G^n$, as

$$\begin{array}{ll} \max & E_G^n[L; A_n] \\ \text{subject to} & E_G^n[L \log L] \leq \eta_n, \end{array} \quad (3)$$

where the maximization is over $L \in \mathscr{L} = \{L \geq 0 : E_G^n L = 1\}$ and consider the Lagrangian relaxation

$$\max_{L \in \mathscr{L}} E_G^n[L; A_n] - \alpha(E_G^n[L \log L] - \eta_n). \quad (4)$$

Our goal is to find $\alpha^* \geq 0$ such that there is an L^* that solves (3) and moreover that $E_G^n[L^* \log L^*] = \eta_n$. Then this L^* will be optimal for (3) (c.f. Luenberger 1997, Theorem 1, p. 220).

First, note that when $\alpha = 0$, the optimal solution to (4) is clearly $L^* = I(A_n)/P_G^n(A_n)$, where $I(A_n)$ denotes the indicator function of the set A_n, which yields the optimal value 1. But then $E_G^n[L^* \log L^*] =$

$-\log P_G^n(A_n) = \Omega(n)$ by our assumption in Theorem 1, and since we assume $\eta_n = o(n)$ we cannot have $E_G^n[L^* \log L^*] = \eta_n$ as n increases. Therefore the case $\alpha^* = 0$ is discriminated.

Now, given any fixed $\alpha > 0$, it can be verified by a convexity argument that the solution to the maximization (1) is given by

$$L^* \propto e^{I(A_n)/\alpha} = e^{1/\alpha} I(A_n); \tag{5}$$

see, for example, Hansen and Sargent (2008). Now we write α_n for α to highlight the role of n, and introduce $\beta_n = 1/\alpha_n$ for convenience. We also write $p_n = P_G^n(A_n)$ and put $q_n = P_G^n(\bar{A}_n) = 1 - p_n$. Then (5) can be written as

$$L^* = \begin{cases} \frac{e^{\beta_n}}{p_n e^{\beta_n} + q_n} \text{ on } A_n \\ \frac{1}{p_n e^{\beta_n} + q_n} \text{ on } A_n^c \end{cases}. \tag{6}$$

We now proceed to find $\alpha_n^* > 0$, or $\beta_n^* = 1/\alpha_n^*$, such that

$$E_G^n[L^* \log L^*] = \eta_n. \tag{7}$$

Using the form of (6), (7) becomes

$$\beta_n \frac{p_n e^{\beta_n}}{p_n e^{\beta_n} + q_n} - \log(p_n e^{\beta_n} + q_n) = \eta_n. \tag{8}$$

Since $\eta_n > \delta > 0$ and $p_n \to 0$ as $n \to \infty$, we must have that all β_n satisfying (8) must also satisfy $\beta_n \to \infty$ as $n \to \infty$. Otherwise the left hand side converges to zero on some subsequence while the right hand side stays positively bounded away from zero. Now, we claim that $\limsup p_n e^{\beta_n} = 0$. Let us proceed assuming this claim for the moment and come back to this issue at the end of our proof. Then, by a Taylor series expansion applied to the left hand side of (8), we have that

$$\beta_n p_n e^{\beta_n}(1 + o(1)) = \eta_n, \tag{9}$$

which gives

$$\log \beta_n + \log p_n + \beta_n + o(1) = \log \eta_n.$$

Heuristically, we must have

$$\beta_n = \log \eta_n - \log p_n - \log \beta_n + o(1) = (\log \eta_n - \log p_n)(1 + o(1)). \tag{10}$$

To verify (10) rigorously, note that when we choose $\beta_n = \log(\eta_n/p_n)$, the left hand side of (9) becomes $\eta_n(\log(\eta_n/p_n))(1+o(1))$ which is much more larger than η_n for n large enough. On the other hand, setting $\beta_n = 0$ gives the left hand side $o(1)$. Therefore, by continuity there must be a solution to (9) in the range $[0, \log(\eta_n/p_n)]$. Consequently, we have that

$$\beta_n = \log \eta_n - \log p_n - \log \beta_n + o(1) = \log \eta_n - \log p_n + r_n, \tag{11}$$

where the remainder term r_n satisfies $|r_n| \leq \log(\log(\eta_n/p_n)) + o(1)$, or equivalently we obtain that $r_n = o(\log(\eta_n/p_n))$, and hence (10).

Iterating the first equality in (10) using (11), we get further that

$$\beta_n = \log \eta_n - \log p_n - \log(\log \eta_n - \log p_n + r_n) + o(1) = \log \eta_n - \log p_n - \log(\log \eta_n - \log p_n) + o(1).$$

Finally, the optimal value is

$$E_G^n[L; A_n] = \frac{p_n e^{\beta_n}}{p_n e^{\beta_n} + q_n} \sim p_n e^{\beta_n} = \frac{\eta_n}{\log \eta_n - \log p_n}(1 + o(1)) \sim \frac{\eta_n}{-\log p_n}.$$

Now we must verify that indeed $\limsup p_n e^{\beta_n} = 0$. Assuming that $\limsup p_n e^{\beta_n} > 0$, then $\beta_{n_k} \geq \delta n_k$ along a subsequence $n_k \to \infty$ for $\delta > 0$. But then we must have from (8) that $\eta_{n_k} \geq \delta' n_k$ for some $\delta' > 0$, contradicting our assumption that $\eta_n = o(n)$. We therefore conclude the statement of our theorem.

3 OUR MAIN RESULT

3.1 A Large Deviations Rate Characterization

In order to prove our main result we shall impose additional technical conditions. We assume that $\psi_G(\cdot)$ is steep in the sense that for all $a \in (-\infty, \infty)$ there is θ_a such that $\psi'_G(\theta_a) = a$. Under this assumption we have that

$$P_G^n(A_n) = \exp\left(-n\bar{I}_G(A) + o(n)\right),$$

where

$$\bar{I}_G(A) = \inf_{x\in A} I_G(x) = \inf_{x\in A} \sup_{\theta} \left(\theta x - \psi_G(\theta)\right).$$

Similarly, for any CDF H, we define $\psi_H(\theta) = \log E_H^1 \exp(\theta X)$ and $I_H(x) = \sup_\theta (\theta x - \psi_H(\theta))$. Similarly as for the definition of $\bar{I}_G(A)$, we write $\bar{I}_H(A) = \inf_{x\in A} I_H(x)$. Now we are ready to state our main result.

Theorem 2 Let $\mathrm{int}(A)$ denote the interior of the closed set A. Suppose that $\bar{I}_H(A) = \bar{I}_H(\mathrm{int}(A)) \in (0, \infty)$ for any $H \in \mathscr{P}$ for some feasible set $\mathscr{P}$. We have

$$\lim_{n\to\infty} \frac{1}{n} \log \max_{H\in\mathscr{P}, X_i \overset{i.i.d.}{\sim} H} P_H^n(A) = -\min_{H\in\mathscr{P}} \bar{I}_H(A).$$

Proof. The proof follows from a large deviations argument. First,

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \max_{H\in\mathscr{P}, X_i \overset{i.i.d.}{\sim} H} P_H^n(A_n) &= \liminf_{n\to\infty} \max_{H\in\mathscr{P}, X_i \overset{i.i.d.}{\sim} H} \frac{1}{n} \log P_H^n(A_n) \\ &\geq \max_{H\in\mathscr{P}, X_i \overset{i.i.d.}{\sim} H} \liminf_{n\to\infty} \frac{1}{n} \log P_H^n(A_n) \\ &\geq -\min_{H\in\mathscr{P}} \bar{I}_H(\mathrm{int}(A)) = -\min_{H\in\mathscr{P}} \bar{I}_H(A).\end{aligned}$$

Next, since A is closed, using Chebycheff inequality (as in the proof of Cramer's theorem; see Dembo and Zeitouni 1998, Remark (c), p. 27) gives

$$P_H(A_n) \leq 2\exp\left(-n\bar{I}_H(A)\right)$$

and hence

$$\limsup_{n\to\infty} \frac{1}{n} \log \max_{H\in\mathscr{P}, X_i \overset{i.i.d.}{\sim} H} P_H^n(A_n) \leq -\min_{H\in\mathscr{P}} \bar{I}_H(A).$$

Combining the upper and lower bounds we get our conclusion. □

The significance of the previous result is that the optimization problem that must be solved can now be cast as a concave program and thus it is more tractable. In order to have a concrete class of examples, let us focus on the case in which $A = [a, \infty)$. Really the key property that holds using this specific selection is that we can identify a specific element $a \in A$ such that $I_H(a) = \bar{I}_H(A)$. We consider the general problem of finding the minimum rate function over a class of distributions, namely

$$\inf_{H\in\mathscr{P}} I_H(a) = \inf_{H\in\mathscr{P}} \sup_{\theta} \{\theta a - \psi_H(\theta)\}. \tag{12}$$

Lemma 1 If $\mathscr{P}$ is a convex set, then the optimization program (12) is convex.

Proof. The inner objective function in (12) is concave as a function of θ, since $\psi_H(\cdot)$ is convex. For the outer objective function, note that $\psi_H(\theta)$ is concave in H, and so $\theta a - \psi_H(\theta)$ is convex in H. As the maximum over a set of convex functions (indexed by θ), the outer objective function $\sup_\theta\{\theta a - \psi_H(\theta)\}$ is also convex as a function of H. Therefore both the inner and outer optimizations in (12) are convex programs. □

3.2 Numerical Procedure

For our numerical procedure, in order to avoid the issue of optimization over infinite-dimensional variables, we concentrate on the case of discrete H. Also we focus on $A=[a,\infty)$. For convenience, we write $\mathbf{p}=(p_1,\ldots,p_m)$ as the weights over support points $\{x_1,\ldots,x_m\}$. Moreover, we write

$$Z(\mathbf{p})=\max_\theta\left\{\theta a-\log\sum_{i=1}^m p_ie^{\theta x_i}\right\} \tag{13}$$

as the outer objective function in (12). We concentrate on the case when a lies between $\min_{i=1,\ldots,m}x_i$ and $\max_{i=1,\ldots,m}x_i$; if $a>\max_{i=1,\ldots,m}x_i$, then the worst-case probability of interest $\max_{P_H\in\mathscr{P},X_i\overset{i.i.d.}{\sim}P_H}P_H^n(A_n)$ is trivially 0, whereas if $a<\min_{i=1,\ldots,m}x_i$ then one can replace a by $\min_{i=1,\ldots,m}x_i$ without changing the probability of interest. In the case $a\in[\min_{i=1,\ldots,m}x_i,\max_{i=1,\ldots,m}x_i]$, the optimal solution for θ in (13) can be solved simply by finding the root of

$$\frac{\sum_{i=1}^m p_ix_ie^{\theta x_i}}{\sum_{i=1}^m p_ie^{\theta x_i}}=a.$$

We now focus on $\min_{\mathbf{p}\in\mathscr{P}}Z(\mathbf{p})$. Suppose that $\mathscr{P}$ is a feasible region dictated by Kullback-Leibler divergence constraint, i.e. $\mathscr{P}=\{\mathbf{p}\geq 0:\sum_{i=1}^m p_i\log(p_i/p_i^0)\leq\eta,\sum_{i=1}^m p_i=1\}$ for some baseline distribution $\mathbf{p}^0=(p_1^0,\ldots,p_m^0)$. A particularly convenient procedure to approximate the optimal solution is to use the conditional gradient (or Frank-Wolfe) method (Frank and Wolfe 1956). This lies on the stepwise optimization, given the current solution $\mathbf{p}^k=(p_1^k,\ldots,p_m^k)$ at step k,

$$\min_{\mathbf{p}\in\mathscr{P}}\nabla Z(\mathbf{p}^k)(\mathbf{p}-\mathbf{p}^k). \tag{14}$$

This subroutine can be easily solved. In fact, we have

$$\nabla Z(\mathbf{p}^k)=\left(-\frac{d}{dp_i}\left(\log\sum_{j=1}^m p_je^{\theta^kx_j}\right)\Bigg|_{p_i=p_i^k}\right)_i=\left(-\frac{e^{\theta^kx_i}}{\sum_{j=1}^m p_je^{\theta^kx_j}}\right)_i$$

by simple arithmetic or by the use of the envelope theorem, where θ^k is the solution to

$$\frac{\sum_{i=1}^m p_i^kx_ie^{\theta x_i}}{\sum_{i=1}^m p_i^ke^{\theta x_i}}=a.$$

For convenience, we let

$$\xi_i(\mathbf{p}^k)=-\frac{e^{\theta^kx_i}}{\sum_{j=1}^m p_j^ke^{\theta^kx_j}}$$

be the i-th coordinate of $\nabla Z(\mathbf{p}^k)$. The solution to (14) is given by $\mathbf{q}^{k+1}=(q_1^{k+1},\ldots,q_m^{k+1})$, where

$$q_i^{k+1}=\frac{p_i^0e^{\beta\xi_i(\mathbf{p}^k)}}{\sum_{j=1}^m p_j^0e^{\beta\xi_j(\mathbf{p}^k)}}$$

and $\beta<0$ satisfies the equation

$$\frac{\sum_{i=1}^m\beta p_i^0\xi_i(\mathbf{p}^k)e^{\beta\xi_i(\mathbf{p}^k)}}{\sum_{j=1}^m p_j^0e^{\beta\xi_j(\mathbf{p}^k)}}-\log\sum_{j=1}^m p_j^0e^{\beta\xi_j(\mathbf{p}^k)}=\eta.$$

If there is no negative root to this equation, then $\mathbf{q}^{k+1}$ is plainly a degenerate mass on $\text{argmin}\{\nabla Z(\mathbf{p}^k)\}$.

Therefore, the iterative procedure is the following:

Iterative Procedure: Start from the baseline distribution $\mathbf{p}^0$ (or any other distribution). At each iteration k, given $\mathbf{p}^k$, do the following:

1. Compute the root θ that solves
$$\frac{\sum_{i=1}^m p_i^k x_i e^{\theta x_i}}{\sum_{i=1}^m p_i^k e^{\theta x_i}} = a.$$
2. Compute
$$\xi_i = -\frac{e^{\theta x_i}}{\sum_{j=1}^m p_j^k e^{\theta x_j}} \quad \text{for } i = 1, \ldots, m.$$
3. Compute, if any, the negative root of
$$\frac{\sum_{i=1}^m \beta p_i^0 \xi_i e^{\beta \xi_i}}{\sum_{j=1}^m p_j^0 e^{\beta \xi_j}} - \log \sum_{j=1}^m p_j^0 e^{\beta \xi_j} = \eta.$$
4. If there is a negative root β, then
$$q_i^{k+1} = \frac{p_i^0 e^{\beta \xi_i}}{\sum_{j=1}^m p_j^0 e^{\beta \xi_j}} \quad \text{for } i = 1, \ldots, m.$$
Otherwise $q_i^{k+1} = 1$ if $i = \text{argmin}\{\xi_i\}$, and 0 for all other i's.
5. Update $\mathbf{p}^{k+1} = (1 - \varepsilon^{k+1})\mathbf{p}^k + \varepsilon^{k+1}\mathbf{q}^{k+1}$ for some step size ε^{k+1}.

There are several choices for the step size ε^{k+1} in the above procedure. It can be a constant, or one can use the so-called limited minimization rule or the Armijo rule (see Bertsekas 1999, p. 217). The latter two choices guarantee convergence to the optimal solution, in the sense that every limit point of the sequence $\mathbf{p}^{k+1}$, as computed by the procedure above with the chosen rule, will be optimal for minimizing $Z(\mathbf{p})$ (Bertsekas 1999, Proposition 2.2.1 and Section 2.2.2).

4 NUMERICAL EXPERIMENTS

We will apply our algorithm to the case of two standard discrete distributions with finite support, namely the binomial distribution and a discrete distribution with random weights. We compare the outcome of robust performance analysis with i.i.d. constraints and without i.i.d. constraints, respectively.

Figure 1 shows the log-probabilities of the event $A_n = \{S_n/n > a\}$ with $a = 8$ associated with a binomial model with parameters $m = 10, p = 0.5$ as n increases. The true model is assumed to be binomial with $m = 10, p = 0.55$. This gives us $\eta = .05$, which is relatively low and chosen for illustrative purposes only. In both optimizations, we simply used step size $\varepsilon^k = k^{-\frac{2}{3}}$ which resulted in empirical convergence of our procedure.

Figure 2 shows the log-probabilities of the event $A_n = \{S_n/n > a\}$ with $a = 8$ associated with a discrete distribution with on the integer support of $\{1, 2, \ldots, 10\}$ with the vector of weights

$$(.05, .12, .08, .13, .06, .04, .14, .13, .13, .12)$$

obtained by random assignment (truncated to two decimal places here). We simulated $N = 300$ i.i.d. replications from the model. We took the standpoint of a modeler who does not have access to the true model, but instead uses maximum likelihood estimation (MLE) to estimate the weights and thus obtain a baseline distribution. This gives us $\eta \approx .02$, which is consistent for data-driven selection. As it can be seen, in both cases the upper bound with i.i.d constraint provides a much tighter bound to the real model than otherwise.

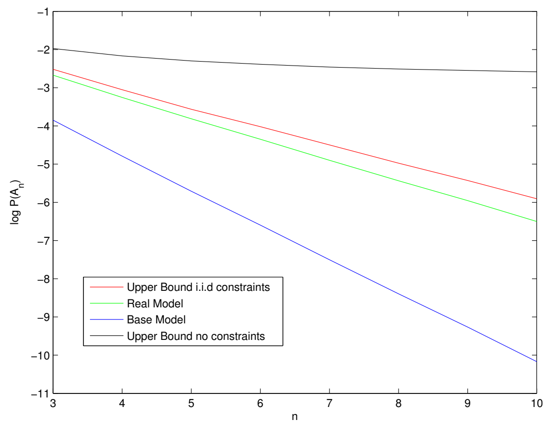

Figure 1: Binomial.

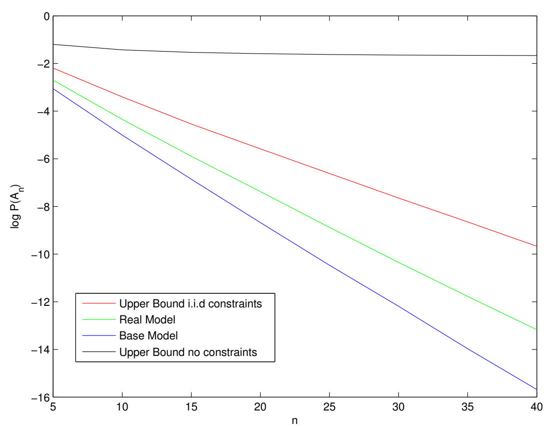

Figure 2: Random Weights.

ACKNOWLEDGMENTS

We thank the referees for their careful reading of our manuscript and for their comments and suggestions which helped us improve our presentation. We also gratefully acknowledge support from NSF grant CMMI 1436700.

REFERENCES

Ben-Tal, A., D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. 2013. "Robust solutions of optimization problems affected by uncertain probabilities". *Management Science* 59 (2): 341–357.

Bertsekas, D. P. 1999. *Nonlinear programming*. Belmont, MA: Athena Scientific.

Dembo, A., and O. Zeitouni. 1998. *Large deviations techniques and applications*. 2nd ed. New York: Springer-Verlag.

Dupuis, P., and R. S. Ellis. 2011. *A weak convergence approach to the theory of large deviations*. New York: John Wiley and Sons.

Frank, M., and P. Wolfe. 1956. "An algorithm for quadratic programming". *Naval Research Logistics Quarterly* 3 (1-2): 95–110.

Glasserman, P., and X. Xu. 2014. "Robust risk measurement and model risk". *Quantitative Finance* 14 (1): 29–58.

Hansen, L. P., and T. J. Sargent. 2008. *Robustness*. Princeton, NJ: Princeton University Press.

Iyengar, G. N. 2005. "Robust dynamic programming". *Mathematics of Operations Research* 30 (2): 257–280.

Jain, A., A. E. B. Lim, and J. G. Shanthikumar. 2010. "On the optimality of threshold control in queues with model uncertainty". *Queueing Systems* 65 (2): 157–174.

Luenberger, D. G. 1997. *Optimization by vector space methods*. New York: John Wiley and Sons.

Nilim, A., and L. El Ghaoui. 2005. "Robust control of Markov decision processes with uncertain transition matrices". *Operations Research* 53 (5): 780–798.

Pardo, L. 2005. *Statistical inference based on divergence measures*. Boca Raton, FL: Chapman & Hall/CRC Press.

Petersen, I. R., M. R. James, and P. Dupuis. 2000. "Minimax optimal control of stochastic uncertain systems with relative entropy constraints". *IEEE Transactions on Automatic Control* 45 (3): 398–412.

AUTHOR BIOGRAPHIES

JOSE BLANCHET is an Associate Professor in the Departments of IEOR and Statistics at Columbia University. Jose holds a Ph.D. in Management Science and Engineering from Stanford University. Prior to joining Columbia he was a faculty member in the Statistics Department at Harvard University. Jose is a recipient of the 2009 Best Publication Award given by the INFORMS Applied Probability Society and of the 2010 Erlang Prize. He also received a PECASE award given by NSF in 2010. He worked as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability and Monte Carlo methods. He serves in the editorial board of Advances in Applied Probability, Journal of Applied Probability, Mathematics of Operations Research and QUESTA. His email is jose.blanchet@columbia.edu.

CHRISTOPHER DOLAN is a PhD student in the Department of Statistics at Columbia University. He received his BA and MS in Mathematics from the Courant Institute at NYU. His research interests are in applied probability and rare event simulation. His email address is dolan@stat.columbia.edu.

HENRY LAM is an Assistant Professor in the Department of Mathematics and Statistics at Boston University. He graduated from Harvard University with a Ph.D. degree in statistics in 2011. His research focuses on large-scale stochastic simulation, rare-event analysis, and simulation optimization, with application interests in service systems and risk management. His email is khlam@bu.edu.