APPROXIMATION OF BOUNDS ON MIXED-LEVEL ORTHOGONAL ARRAYS

ALI DEVIN SEZER,[*][**] AND

FERRUH ÖZBUDAK,[*] *Middle East Technical University*

Abstract

Mixed-level orthogonal arrays are basic structures in experimental design. We develop three algorithms that compute Rao- and Gilbert–Varshamov-type bounds for mixed-level orthogonal arrays. The computational complexity of the terms involved in the original combinatorial representations of these bounds can grow fast as the parameters of the arrays increase and this justifies the construction of these algorithms. The first is a recursive algorithm that computes the bounds exactly, the second is based on an asymptotic analysis, and the third is a simulation algorithm. They are all based on the representation of the combinatorial expressions that appear in the bounds as expectations involving a symmetric random walk. The Markov property of the underlying random walk gives the recursive formula to compute the expectations. A large deviation (LD) analysis of the expectations provides the asymptotic algorithm. The asymptotically optimal importance sampling (IS) of the same expectation provides the simulation algorithm. Both the LD analysis and the construction of the IS algorithm use a representation of these problems as a sequence of stochastic optimal control problems converging to a limit calculus of a variations problem. The construction of the IS algorithm uses a recently discovered method of using subsolutions to the Hamilton–Jacobi–Bellman equations associated with the limit problem.

Keywords: Mixed-level orthogonal array; Rao bound; Gilbert–Varshamov bound; error block code; counting; importance sampling; large deviation; optimal control; asymptotic analysis; subsolution approach; Hamilton–Jacobi–Bellman equation

2010 Mathematics Subject Classification: Primary 05B15; 62K99; 65C05
Secondary 93E20; 49L99

1. Introduction

Mixed-level orthogonal arrays (OAs) are fundamental to experimental design. Each row of an array is thought of as a run of an experiment; each entry of the row is the value of a variable of the system being tested. The goal of the experiment is to test as wide a range of variable values of the system as possible. The number of blocks σ, and the block and alphabet sizes $l_i, s_i, i = 1, 2, 3, \ldots, \sigma$, of an OA are determined by the number of variables in the system being tested and which values these variables can take. The remaining parameters of an OA are its number of rows N and its strength t. The strength of an OA is t when the OA is capable of exploring all possible interactions for up to t variables of the system (see Definition 1); N is the number of experiments that the OA describes. A high t and a low N is desirable. The

following bound, called the Rao bound, gives a lower bound on N in terms of t:

$$N \geq \sum_{i=0}^{t/2} \sum_{u_1,u_2,...,u_\sigma \geq 0, \sum u_m=i} \prod_{m=1}^{\sigma} \binom{l_m}{u_m} (s_m - 1)^{u_m}. \tag{1}$$

The Rao bound was first proved for fixed-level orthogonal arrays by Rao [31]; for the proof of the general case, see [25, p. 201]. Our first object of study is this bound and the goal is to develop algorithms that compute its right-hand side exactly and approximately.

The Rao bound is a necessary bound, all OAs satisfy it. There are also sufficient bounds that arise from constructions. One well-known construction method for ordinary OAs is by taking the dual of error correcting codes [25]. Feng *et al.* [17] generalized this idea by defining error-block codes, which are error-correcting codes in which one can specify the alphabet to be used for each entry of the code word. Furthermore, Feng *et al.* [17] noted that the duals of error-block codes are mixed-level orthogonal arrays. This idea and construction of error-block codes were used in [26] to construct orthogonal arrays whose parameters satisfy the following Gilbert–Varshamov-type (GV-type) bound:

$$Nq \geq \sum_{i=0}^{t-1} \sum_{u_1,u_2,...,u_\sigma, \sum u_m=i} s_\sigma \binom{l_\sigma - 1}{u_\sigma - 1} (s_\sigma - 1)^{u_\sigma - 1} \prod_{m=1}^{\sigma-1} \binom{l_m}{u_m} (s_m - 1)^{u_m} \geq N. \tag{2}$$

Our second object of study is this bound.

In Subsection 2.2 we calculate the computational complexity of directly computing the Rao bound (1) and the GV bound (2). We see that this complexity is polynomial in the strength parameter, and the degree of the polynomial is one more than the number of different types of alphabet used in the OA. If many different types of alphabet are used in an OA, which is typical in real-life experimental designs, the Rao and the GV bounds become inefficient to compute directly from their original representations (1) and (2). This potentially high complexity of the direct computation of these bounds justifies the construction of new algorithms to compute them. In the present paper we develop three algorithms for this purpose. The fact that underlies these is an expectation representation of the Rao and the GV bounds that we derive in Sections 3 and 6. The expectation is that of a function of a random walk whose increments are either 0 or 1 with equal probability. The walk takes n steps, the row length of the OA, and accumulates a cost throughout its excursion as follows: if the walk goes up at the ith step, the accumulated cost increases by a factor of one less than the alphabet size of the ith variable of the OA. The aforementioned representation is the expectation of this accumulated cost over sample paths which are less than $t/2$ at the last step of the random walk for the Rao bound and less than $t-1$ for the GV bound.

Once these expectation representations are available, it is possible to use them in several ways to obtain algorithms to compute the bounds. The Markov property of the underlying walk gives the recursive formula (7). The complexity of this formula is a second-order polynomial in the strength parameter and is far less than the original formulae when the number of alphabets is large.

The analysis of the asymptotic behavior of bounds such as the Rao and the GV bounds is of basic interest. Typically, the number of elements in an error correcting code or an orthogonal array grows exponentially in the size of the code or the array [27], [28]. Therefore, in the asymptotic analysis of the size of these structures, the exponential growth rate is the natural asymptotic quantity to study [28, Chapter 17]. Ling and Özbudak [26] carried out an asymptotic analysis to compute the exponential growth rate of the GV bound for orthogonal arrays with two

alphabets. To the authors' knowledge, no results concerning the asymptotic behavior of either the GV or the Rao bound for general mixed-level orthogonal arrays is available in the current literature. With our expectation representation, an asymptotic analysis of these bounds becomes what is called a large deviation (LD) analysis in probability theory and we use the methods of LD theory to carry it out. In Section 4 we use the stochastic optimal control approach to LDs [8], [18], [20], to show that the right-hand side of the Rao bound (1) grows exponentially in the row length n and identify the growth rate. Following [8], we use a relative entropy representation of our expectation of interest to write it as a discrete-time stochastic optimal control problem. After a $1/n$ log transform, this control problem converges to a limit deterministic calculus of variations problem. Similar to [15] and [32], the connection between the prelimit and the limit problems is established using the Hamilton–Jacobi–Bellman (HJB) equation associated with the limit problem (see Section 4 for the Rao bound and Section 6 for the GV bound). This analysis provides our second approximation algorithm. To the authors' knowledge, the idea of using the limit HJB equation to compute LD limits first appeared in [18] in the context of the analysis of queueing systems.

The asymptotic analysis gives good approximations in an exponential scale. More accurate approximations can be obtained using simulation, which is possible because we have the expectation representations (6) and (44). However, these are expectations over sets with small probabilities (i.e. rare) for reasonable values of the strength parameter t. For such expectations, ordinary simulation would require a great number of samples for reliable estimates. A remedy to this is importance sampling (IS) whereby the sampling distribution is altered so that the set over which the expectation is computed is not rare anymore. We modify the estimator by multiplying it with a likelihood ratio to counteract the change in the sampling distribution. IS is a well-known idea, going back at least to 1949; see, for example, [9], [21], [24], [35], and the references therein.

For our problem, an IS distribution will be one under which our random walk remains below $t - 1$ or $t/2$ at its final step with high probability. There are many such distributions. Among these, we would like to choose a distribution that minimizes the variance of the IS estimator. It is well known that obtaining an exact solution to this optimization problem is as difficult as directly computing the expectation [23]. In situations such as the one covered in this paper where the object of study is a sequence of expectations decaying or growing exponentially in a parameter, a compromise is to choose a sequence of estimators whose variances decay or grow exponentially at a rate twice the asymptotic decay or growth rate of the expectation itself. Such a sequence is called asymptotically optimal; see [9], [35], and the references therein. To obtain such a sequence, we will follow [9] and [15], and represent the variance minimization problems in IS, once again, as a sequence of stochastic optimal control problems. Under proper scaling, these also converge to the same limit control problem as that which emerges in the LD analysis. Theorem 3 of Subsection 5.4 asserts that a simple change of measure based on a piecewise linear subsolution of the HJB equation of the limit control problem is asymptotically optimal. Throughout a simulation, the proposed change of measure depends only on the current block being simulated; thus, within each block it can be thought of as a static exponential tilt of the original independent and identically distributed (i.i.d.) distribution. The idea of using subsolutions to construct IS algorithms is from [11], [12], [15], and [32], and is called the subsolution approach to IS.

Another way to say that an IS estimator of the Rao bound is asymptotically optimal is to say that its coefficient of variation, which is the ratio of the standard deviation of the estimator to the bound itself, grows *subexponentially* in the row length. It is well known that, in the context of

the estimation of tail probabilities of the sample mean of a sequence of i.i.d. random variables, the coefficient of variation of the static IS change of measure based on exponential tilting grows at rate $n^{1/4}$ [3]. In Subsection 5.5 we provide numerical evidence that the coefficient of variation of the IS estimator of Subsection 5.4 exhibits the same growth behavior. We also comment on methods that may be useful in the proof of this observation.

The use of randomized algorithms for counting is one of the central ideas in statistics. The use of IS for this purpose seems to be relatively new. The paper by Chen *et al.* [6] is the first that we are aware of that uses IS for purposes of counting. More recent articles since [6] include [1], [4], and [5]. The present work seems to be the first to use the subsolution method to construct IS algorithms for counting.

The plan of this paper is as follows. In Section 2 we give the definition of an OA and restate the Rao and GV bounds. We also compute a lower bound on the computational complexity of the original combinatorial representations of these bounds. In Section 3 we derive the expectation representation of the Rao bound and state the exact recursive algorithm to compute it (see (7)). In Section 4 we carry out the LD analysis of the expectation representation of the Rao bound. The final result here is Theorem 2 which characterizes the growth rate of the bound as a finite-dimensional concave maximization problem. The dimension of the problem is the number of alphabets used in the OA. In Section 5 we use the ideas in the above paragraphs to construct an asymptotically optimal IS algorithm to estimate the Rao bound; the final result is Theorem 3. In Section 6 we do for the GV bound what we do for the Rao bound in Sections 4 and 5. This generalization requires only minor modifications. Section 7 provides numerical results which give evidence that the constructed algorithms are effective in practice as well.

2. Definitions and bounds

We begin with the following definition from [25].

Definition 1. A matrix A is said to be an OA$(N, s_1^{l_1} s_2^{l_2}, \ldots, s_\sigma^{l_\sigma}, t)$ if it has the following structure.

1. A has N rows.

2. The row length of A is $l_1 + l_2 + \cdots + l_\sigma$; the first l_1 components of each row are from the alphabet $\mathbb{Z}_{s_1}$, the second l_2 components are from $\mathbb{Z}_{s_2}, \ldots$, the last l_σ components are from $\mathbb{Z}_{s_\sigma}$.

3. Take any t columns $c_{i_1}, c_{i_2}, \ldots, c_{i_t}$ of A and call the matrix formed by these columns A'. Take any string s of length t such that the jth letter of s comes from the alphabet corresponding to column i_j. Count the number times s occurs as a row of A'. This count is the same for all s.

Item 3 of Definition 1 is the orthogonality property, and t is the *strength* of the orthogonal array. These types of array are called mixed-level because the columns are allowed to be from different alphabets (property 2 of Definition 1).

The parameters of any mixed-level orthogonal array have to satisfy the Rao bound:

$$N \geq \sum_{i=0}^{t/2} \sum_{u_1, u_2, \ldots, u_\sigma \geq 0, \sum u_m = i} \prod_{m=1}^{\sigma} \binom{l_m}{u_m} (s_m - 1)^{u_m}. \tag{3}$$

This bound corresponds to the sphere packing bound for error-block codes. As indicated in the introduction, for $\sigma = 1$, (3) was proved in [31]; for the proof of the general case, see [25, p. 201].

2.1. Sufficient bounds

The duality idea mentioned in the introduction and block-error code constructions implied by Theorem 3.1 of [26] give mixed-level orthogonal arrays whose parameters satisfy the following conditions: $s_i = q^{m_i}$, where q is a prime power,

$$Nq \geq \sum_{i=0}^{t-1} \sum_{u_1, u_2, \ldots, u_\sigma, \sum u_m = i} s_\sigma \binom{l_\sigma - 1}{u_\sigma - 1} (s_\sigma - 1)^{u_\sigma - 1} \prod_{m=1}^{\sigma - 1} \binom{l_m}{u_m} (s_m - 1)^{u_m} \geq N. \quad (4)$$

This is a sufficient bound, that is, it is known that OAs with these parameters do exist. Bounds like (4) are called GV-type bounds in coding theory. Using methods from algebraic geometry, it has been shown that it is possible to beat asymptotic GV bounds but only over relatively large alphabets. For small alphabet sizes, the asymptotic GV bound is still the best-known bound. We refer the reader to [29] and the references therein for more on these.

The right-hand side of (4) has essentially the same structure as that of (3). The key difference between these bounds is the upper limit of the outer sum: (3) goes up to $t/2$ whereas (4) goes up to $t - 1$.

2.2. Computational complexity of evaluating (3) and (4)

It follows from their definitions that the evaluations of (3) and (4) have the same computational complexity. Therefore, it is enough to consider one of them.

The right-hand side of (3) involves a partitioning of each i less than $t/2$ into a sum of σ integers. The number of such partitions is $\binom{\sigma + i - 1}{\sigma - 1}$. Then the number of operations needed to compute the right-hand side of (3) is bounded below by

$$\sum_{i=0}^{t/2} \sigma \binom{\sigma + i - 1}{\sigma - 1} \geq \sum_{i=0}^{t/2 - 1} i^\sigma \geq C \left(\frac{t}{2} \right)^{\sigma + 1}, \quad (5)$$

where C is a constant that depends only on σ. If the strength parameter t grows linearly in n, i.e. if $t = \mu n$, where $\mu \in (0, 1)$, a direct computation of (3) requires $\Omega(n^{\sigma + 1})$ operations. The present paper provides methods that compute (3) and (4) much more efficiently.

The next section presents a probabilistic representation of (3), which forms the basis for all of the results and algorithms presented in this paper.

3. Expectation representation

Let X_i be i.i.d. Bernoulli random variables with $P(X_i = 1) = P(X_i = 0) = \frac{1}{2}$. Let $S_k := X_1 + X_2 + \cdots + X_k$. Define the following 'running cost':

$$r(x, j) := \begin{cases} 1 & \text{if } x = 0, \\ s_i - 1 & \text{if } x = 1 \text{ and } \sum_{k=1}^{i-1} l_k + 1 \leq j \leq \sum_{k=1}^{i} l_k. \end{cases}$$

Equation (3) can be written in the form

$$N \geq 2^n E \left[\mathbf{1}_{\{S_n \leq t/2\}} \prod_{j=1}^{n} r(X_j, j) \right] = E \left[\mathbf{1}_{\{S_n \leq t/2\}} \prod_{j=1}^{n} 2r(X_j, j) \right]. \quad (6)$$

This is an expectation over the trajectories of S_k that stay below the level $t/2$ at step n. Each trajectory of S_k has probability 2^{-n} and the term 2^n in front of the first expectation in (6) cancels this probability. At each step the random walk accumulates a running cost $2r$; the cost depends on the step number and the current step. The random walk can be thought of as a scan of the letters of a row of the array. At each step we flip a coin to decide whether the current letter will be included in the computation. If the decision is yes, i.e. if $X_i = 1$ and the random walk goes up, then the current bound is multiplied with $2(s_i - 1)$, where s_i is the alphabet size of the letter we are going over (this is the term $2r$ in (6)). The first sum in (3) groups trajectories according to their positions at step n. For position $i \leq t/2$, the second sum in (3) partitions these i up-steps into different cost regions and the binomial coefficients count the number of possible ways u_m up-steps can be taken in l_m steps.

3.1. A simple recursive algorithm to compute the Rao bound

For integers $0 \leq x \leq t/2$ and $0 \leq k \leq n$, define

$$M(x, k) = \mathrm{E}\left[\mathbf{1}_{\{x+S_{n-j} \leq t/2\}} \prod_{j=k+1}^{n} 2r(X_i, j) \right].$$

The Rao bound (3) in terms of M is $N \geq M(0, 0)$. Because the X_i are i.i.d. and the S_i are their sum, M satisfies

$$M(x, k) = M(x + 1, k + 1)r(x, k) + M(x, k + 1) \tag{7}$$

for $x < t/2$ and $k < n$. In addition, we have the boundary conditions $M(x, n) = 0$ for $x \leq t/2$ and $M(t/2, k) = 0$ for $k \leq n$. These give an algorithm that takes only $tn/2$ steps to compute the Rao bound. If we write the strength parameter t as a fraction μ of n, i.e. $t = \mu n$, then the complexity analysis in the previous chapter implies that the direct evaluation of (3) will take at least $\Omega(n^{\sigma+1})$ operations. Whereas the computation of the same bound using (7) will only take $O(n^2)$ operations.

4. Large deviation analysis

The goal of this section is an asymptotic analysis of the right-hand side of (6) as $n \to \infty$. In order for this analysis to be meaningful, t and l_i need to grow with n. Therefore, we assume that

$$t = \mu n, \qquad \mu \in (0, 1), \qquad l_i = na_i, \qquad \sum a_i = 1. \tag{8}$$

The asymptotic analysis of (3) now consists of evaluating

$$\lim_n \frac{1}{n} \log \mathrm{E}\left[\mathbf{1}_{\{S_n \leq t/2\}} \prod_{j=1}^{n} 2r(X_j, j) \right]. \tag{9}$$

For the evaluation of (9), we will follow [8] and begin by representing the $\log \mathrm{E}[\cdots]$ term in it as a discrete-time stochastic optimal control problem as follows.

Proposition 1. *The identity*

$$\log \mathrm{E}\left[\mathbf{1}_{\{S_n \leq t/2\}} \prod_{j=1}^{n} 2r(X_j, j) \right] = \sup_{\substack{\bar{p}(\cdot \mid \cdot, \cdot) \\ \bar{\mathrm{P}}(S_n \leq \mu n/2)=1}} \bar{\mathrm{E}}\left[\sum_{j=1}^{n} \log r(X_j, j) - \log \bar{p}(X_j \mid j, S_{j-1}) \right] \tag{10}$$

holds, where the sup *is over all transition probabilities* $\bar{p}(\cdot \mid \cdot, \cdot) \colon \mathbb{Z}_2 \times \mathbb{N} \times \mathbb{N} \to [0, 1]$ *that*

give the probability of the steps 0 *and* 1 *given the current position and the current step number of the random walk S, and* $\bar{\text{P}}$ *is the probability distribution defined by these measures on the path space of the random walk.*

The proof of this result is similar to that of Proposition 1.4.2 of [8, p. 31], and is thus omitted. The sup on the right-hand side of (10) is over all Markov chains on the sample paths of S_k such that the nth step is less than $t/2$ with probability 1. The log term inside the sup corresponds to the entropy of $\bar{p}(\cdot \mid \cdot, \cdot)$. Define $A_i := \sum_{j=1}^{i} a_j$ and

$$\tilde{r}(t) := \log(s_i - 1), \qquad A_i \leq t < A_{i+1},$$

and let

$$H(\theta) := -\theta \log \theta - (1 - \theta) \log(1 - \theta)$$

be the entropy function. As observed earlier, the right-hand side of (10) is a stochastic optimal control problem. Upon dividing it by n, scaling the time and space parameters with $1/n$, and sending n to ∞, we obtain the following limit deterministic optimal control problem:

$$\sup_{\theta(\cdot)} \int_0^1 [\tilde{r}(t)\theta + H(\theta)] \, dt. \tag{11}$$

Here the sup is over all measurable functions on [0, 1] with values in [0, 1] such that $\int_0^1 \theta(t) \, dt \leq \mu/2$. The rigorous connection between this optimal control problem and (10) can be established in several ways. For example, we can use the weak convergence approach of [8]. Another approach is via the HJB equation associated with the limit control problem (11) and a verification argument, which is followed in [15]. In this paper we will take this second path because the same method will also allow us to prove the asymptotic optimality of an IS estimator based on a subsolution of the limit HJB equation.

4.1. Solution to the limit control problem

For $A_i \leq t \leq A_{i+1}$, $\tilde{r}(t)\theta + H(\theta) = \log(s_i - 1)\theta + H(\theta)$ is a strictly concave function with no t dependence. This and Jensen's inequality imply that, between times A_i and A_{i+1}, the optimal trajectory of (11) is a straight line. Therefore, it is enough to consider the optimization problem (11) over piecewise linear continuous paths and the sup in (11) equals

$$\sup \left\{ \sum_{i=1}^{\sigma} a_i (\theta_i \log(s_i - 1) + H(\theta_i)) \right\}, \tag{12}$$

where the sup is subject to

$$\theta_i \in (0, 1), \qquad \langle a, \theta \rangle \leq \tfrac{1}{2}\mu. \tag{13}$$

The objective function of this finite-dimensional constrained optimization problem is strictly concave and its constraints linear. Therefore, the use of a Lagrange multiplier converts the problem to one of root finding of a one-dimensional monotone function.

In the next subsection we will prove that a function defined based on (12) satisfies an HJB equation. We will use this fact to prove the convergence of (9) to (12).

4.2. The limit HJB equation

Let us generalize the problem in (11) so that the problem starts from any initial point $x \leq \mu/2$ at any time $t \in [0, 1]$:

$$V(x, t) := \sup_{\theta} \int_t^1 [\tilde{r}(s)\theta(s) + H(\theta(s))] \, ds. \tag{14}$$

Here the sup is over all measurable $\theta(\cdot) \geq 0$ such that $x + \int_t^1 \theta(s) \, ds \leq \mu/2$. The sup in (11) equals $V(0, 0)$. Generalizing (12), for $A_i \leq t < A_{i+1}$, we have

$$V(x, t) = \sup\left\{ (A_{i+1} - t)(\theta_i \log(s_i - 1) + H(\theta_i)) + \sum_{j=i+1}^{\sigma} a_j(\theta_j \log(s_j - 1) + H(\theta_j)) \right\}, \tag{15}$$

where the sup is subject to

$$\theta_j \in (0, 1), \qquad x + \theta_i(A_{i+1} - t) + \sum_{j=i+1}^{\sigma} a_j \theta_j \leq \frac{\mu}{2}. \tag{16}$$

Let us now write V more explicitly. Firstly, the absolute maximizer of (15) without the constraints in (16) is

$$\theta_j^* = \frac{s_j - 1}{s_j}. \tag{17}$$

If the θ_j^* satisfy (16), i.e. if

$$x + \theta_i^*(A_{i+1} - t) + \sum_{j=i+1}^{\sigma} a_j \theta_j^* \leq \frac{\mu}{2}, \tag{18}$$

then V equals

$$V(x, t) = (A_{i+1} - t)\left[\frac{s_i - 1}{s_i} \log(s_i - 1) + H\left(\frac{s_i - 1}{s_i} \right) \right]$$
$$+ \sum_{j=i+1}^{\sigma} a_j\left[\frac{s_j - 1}{s_j} \log(s_j - 1) + H\left(\frac{s_j - 1}{s_j} \right) \right].$$

If the absolute maximizers (17) do not satisfy (16) then we can use a Lagrange multiplier λ to solve (15):

$$\log(s_j - 1) + \log \frac{(1 - \theta_j)}{\theta_j} = \lambda, \qquad j \geq i.$$

Then

$$\theta_j^*(\lambda) = \frac{s_j - 1}{e^\lambda + s_j - 1}. \tag{19}$$

For these to give a solution to (15), they must satisfy (16) with equality:

$$(A_{i+1} - t)\frac{s_i - 1}{e^\lambda + s_i - 1} + \sum_{j=i+1}^{\sigma} a_j\frac{s_j - 1}{e^\lambda + s_j - 1} = \frac{\mu}{2} - x. \tag{20}$$

For $\lambda = 0$, the left-hand side of (20) is by assumption greater than $\mu/2 - x$ and, for $\lambda = \infty$, it is 0. Because it is monotone in λ, there exists a unique $\lambda^*(t, x)$ for which (20) is satisfied. By the implicit function theorem, $\lambda^*(t, x)$ is twice differentiable in both t and x with bounded derivatives for $t \neq A_j$. For $t = A_j$, λ has left and right derivatives in t, and an ordinary derivative in x. Because the function that is optimized in (15) is strictly concave, the stationary point given by λ^* is actually a global maximizer. Define

$$\tilde{V}(t, \lambda) := (A_{i+1} - t)\left[\frac{s_i - 1}{e^\lambda + s_i - 1}\log(s_i - 1) + H\left(\frac{s_i - 1}{e^\lambda + s_i - 1}\right)\right]$$
$$+ \sum_{j=i+1}^{\sigma} a_j\left[\frac{s_j - 1}{e^\lambda + s_j - 1}\log(s_j - 1) + H\left(\frac{s_j - 1}{e^\lambda + s_j - 1}\right)\right].$$

In light of the above computations, $V(x, t)$ of (15) can be written more explicitly as

$$V(x, t) = \begin{cases} \tilde{V}(t, 0) & \text{if (18) holds,} \\ \tilde{V}(t, \lambda^*(x, t)) & \text{otherwise.} \end{cases}$$

We obtain the following proposition by calculus and implicit differentiation.

Proposition 2. *The function V is twice differentiable except for $t = A_i$ where it has directional derivative $V_t(x, t)$ which is defined as $V_t(x, t) = \lim_{h \searrow 0}(V(x, t + h) - V(x, t))/h$. Higher order t partial derivatives similarly exist. In particular, for any t and x, we have*

$$V(x + \delta, t + h) = V(x, t) + \delta V_x(x, t) + h V_t(x, t) + c(x, t)(\delta^2 + h^2),$$

where $\sup_{x,t} |c(x, t)| = C < \infty$.

Now we state the HJB equation satisfied by V.

Theorem 1. *The following dynamic programming equation holds:*

$$0 = \sup_{\theta \in [0,1]} \{\tilde{r}(t)\theta + H(\theta) + V_x(x, t)\theta + V_t(x, t)\} \tag{21}$$

for $(x, t) \in [0, \mu/2] \times [0, 1)$.

Proof. Take $(x, t) \in [0, \mu/2] \times [0, 1)$, a small $\delta > 0$, and $\theta \in [0, 1]$. Equation (14) implies that

$$V(x, t) \geq \int_t^{t+\delta} [\tilde{r}(s)\theta + H(\theta)]\,ds + V(x + \theta\delta, t + \delta),$$
$$V(x, t) - V(x + \theta\delta, t + \delta) \geq [\log(s_i - 1) + H(\theta)]\delta.$$

Because V_t and V_x exist, dividing both sides of the last display by δ and letting $\delta \to 0$ gives

$$-V_t - \theta V_x \geq \log(s_i - 1) + H(\theta).$$

Because this is true for all $\theta \in [0, 1]$, we have

$$0 \geq \sup_{\theta \in [0,1]} \{\tilde{r}(t)\theta + H(\theta) + V_x(x, t)\theta + V_t(x, t)\}.$$

We replace '$\geq$' with '$=$' by taking θ to be the optimal control $\theta^*(\lambda^*(x, t))$.

4.3. Convergence analysis

In this subsection we formally connect the sequence of stochastic optimal control problems in (10) to the limit control problem (11) and its solution developed in the previous subsection. Figure 1 gives the level curves of $V(x, t)$ and $V_{60}(\lfloor nx \rfloor, \lfloor nt \rfloor)$, where

$$V_n(x, i) = \frac{1}{n} \log \mathbb{E}\left[\mathbf{1}_{\{x + S_n \leq \mu n/2\}} \prod_{j=i}^{n} 2r(X_j, j)\right]$$

for $a_1 = a_2 = a_3 = \frac{1}{3}$, $s_1 = 2$, $s_2 = 30$, $s_3 = 100$, and $\mu = 0.1$. This figure suggests that $V_n(nx, nt) \to V(x, t)$ for all values of (x, t). Our main convergence theorem, which we state and prove next, concerns the special case when $(x, t) = (0, 0)$.

Theorem 2. *The large deviations limit in (9) equals $V(0, 0)$, i.e.*

$$\lim_n \frac{1}{n} \log 2^n \mathbb{E}\left[\mathbf{1}_{\{S_n \leq t_n/2\}} \prod_{j=1}^{n} r(X_j, j)\right] = \sup\left\{\sum_{i=1}^{\sigma} a_i (\theta_i \log(s_i - 1) + H(\theta_i))\right\}, \qquad (22)$$

where the sup *is over*

$$\theta_i \in (0, 1), \qquad \langle a, \theta \rangle \leq \tfrac{1}{2}\mu. \qquad (23)$$

Proof. The proof will be a verification argument using V and the HJB equation (21). By Proposition 1, there exists $\bar{p}^n(\cdot \mid \cdot, \cdot)$ such that

$$\log \mathbb{E}\left[\mathbf{1}_{\{S_n \leq t/2\}} \prod_{j=1}^{n} 2r(X_j, j)\right] = \bar{\mathbb{E}}\left[\sum_{j=1}^{n} \log r(X_j, j) - \log \bar{p}^n(X_j \mid j, S_{j-1})\right] + \varepsilon(n),$$

where $\varepsilon(n) \to 0$ and $\bar{\mathbb{E}}$ is the expectation with respect to $\bar{p}^n(\cdot \mid \cdot, \cdot)$. We have

$$V(0, 0) = \bar{\mathbb{E}}\left[V(0, 0) - V\left(\frac{S_n}{n}, 1\right)\right] = \sum_{j=0}^{n-1} \bar{\mathbb{E}}\left[V\left(\frac{S_j}{n}, \frac{j}{n}\right) - V\left(\frac{S_{j+1}}{n}, \frac{j+1}{n}\right)\right].$$
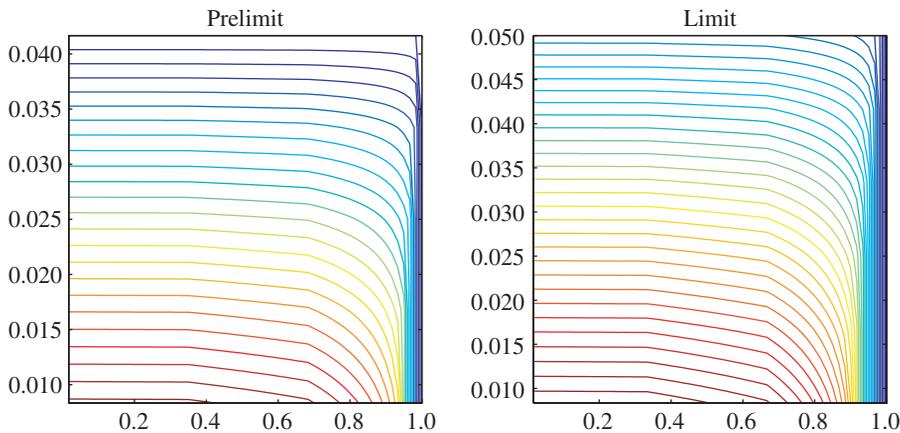

FIGURE 1: The level curves of $V(x, t)$ and V_{60}.

By Proposition 2, this equals

$$\frac{C(n)}{n} + \sum_{j=1}^{n} \bar{E}\left[-V_x\left(\frac{S_j}{n}, \frac{j}{n}\right)\frac{X_j}{n} - V_t\left(\frac{S_j}{n}, \frac{j}{n}\right)\frac{1}{n} \right],$$

where $\sup_n |C(n)| < \infty$. We can condition the last expectation on S_j to rewrite it as

$$\frac{C(n)}{n} + \frac{1}{n}\sum_{j=1}^{n} \bar{E}\left[-V_x\left(\frac{S_j}{n}, \frac{j}{n}\right)\bar{p}^n\left(1 \;\middle|\; \frac{S_j}{n}, \frac{j}{n}\right) - V_t\left(\frac{S_j}{n}, \frac{j}{n}\right) \right].$$

By Theorem 1, this last sum is greater than

$$\frac{C(n)}{n} + \sum_{j=1}^{n} \bar{E}\left[\tilde{r}\left(\frac{j}{n}\right)\bar{p}^n\left(1 \;\middle|\; \frac{S_j}{n}, \frac{j}{n}\right) + H(\bar{p}^n(1 \mid S_j, j)) \right],$$

which equals

$$\frac{C(n)}{n} + \frac{1}{n}\sum_{j=1}^{n} \bar{E}[\log r(X_j, j) - \log(\bar{p}^n(X_j \mid S_j, j))].$$

Letting n go to ∞ yields

$$V(0, 0) \geq \limsup \frac{1}{n}\bar{E}^n\left[\sum_{j=1}^{n} \log r(X_j, j) - \log \bar{p}^n(X_j \mid S_j, j) \right]$$

$$\geq \limsup \frac{1}{n}\log E\left[\mathbf{1}_{\{S_n \leq t_n/2\}} \prod_{j=1}^{n} 2r(X_j, j) \right] - \varepsilon.$$

For the reverse inequality, we first note that the result of the optimization in (22) is continuous in the strength parameter μ which appears in constraint (23). Let θ_i^* be the optimizers of (22) when the μ in (23) is replaced with $\mu - 4\varepsilon$, where $\varepsilon > 0$ is a small constant. Let

$$\bar{p}^*(1 \mid x, j) = \theta_i^* \quad \text{if } A_i \leq j/n \leq A_{i+1}, \tag{24}$$

and let $\bar{P}^*$ be the measure on the path space of (S, X) corresponding to $\bar{p}^*$. By definition, X_j is i.i.d. for $A_i \leq j/n < A_{i+1}$. Therefore, the law of large numbers is applicable and gives

$$\bar{P}^*\left(\frac{S_n}{n} > \frac{\mu}{2} - \varepsilon \right) \to 0. \tag{25}$$

Let $p_n^* = \bar{P}^*(S_n \leq \mu n/2)$, and define

$$\bar{P}^{*,c} = \frac{1}{p_n^*}\mathbf{1}_{\{S_n \leq \mu n/2\}}\bar{P}^*. \tag{26}$$

Under $\bar{P}^{*,c}$, (X_n, S_n) is a Markov chain whose transition probabilities are determined by

$$\bar{p}^{*,c}(1 \mid x, j) = \bar{p}^*(1 \mid x, j)\frac{\bar{P}^*(S_n \leq \mu n/2 \mid S_{n-j} = x + 1)}{\bar{P}^*(S_n \leq \mu n/2 \mid S_{n-j-1} = x)}. \tag{27}$$

We have $\bar{P}^{*,c}(S_n \leq \mu n/2) = 1$ and, therefore, by Proposition 1,

$$\frac{1}{n} \log 2^n \mathrm{E}\left[\mathbf{1}_{\{S_n \leq t/2\}} \prod_{j=1}^n r(X_j, j)\right] \geq \frac{1}{n} \bar{\mathrm{E}}^{*,c}\left[\sum_{j=1}^n \log r(X_j, j) - \log \bar{p}^{*,c}(X_j \mid S_{j-1}, j)\right].$$

By (26) and (27), this equals

$$\frac{1}{p_n^*} \frac{1}{n} \bar{\mathrm{E}}^*\left[\mathbf{1}_{\{S_n \leq \mu n/2\}} \sum_{j=1}^n \log r(X_j, j) - \log \bar{p}^*(X_j \mid S_{j-1}, j)\right] + \frac{1}{n} \log p_n^*. \tag{28}$$

Note that θ_i^* and r are all positive and bounded. Therefore, there exists a positive C such that

$$(28) \geq \frac{1}{p_n^*} \frac{1}{n} \bar{\mathrm{E}}^*\left[\sum_{j=1}^n \log r(X_j, j) - \log \bar{p}^*(X_j \mid S_{j-1}, j)\right] + \frac{1}{n} \log p_n^* - \frac{C(1 - p_n^*)}{p_n^*}$$

$$= \frac{1}{p_n^*} \frac{1}{n} \sum_{i=1}^{\sigma} l_i(\theta_i^* \log(s_i - 1) + H(\theta_i^*)) + \frac{1}{n} \log p_n^* - \frac{C(1 - p_n^*)}{p_n^*}.$$

Equation (25) implies that $p_n^* \to 1$. This, the last sequence of inequalities, and the definition of l_i give

$$\liminf \frac{1}{n} \log 2^n \mathrm{E}\left[\mathbf{1}_{\{S_n \leq t/2\}} \prod_{j=1}^n r(X_j, j)\right] \geq \sum_{i=1}^{\sigma} a_i(\theta_i^* \log(s_i - 1) + H(\theta_i^*)) = V(0, 0) - \delta,$$

where δ is a small number that goes to 0 with ε. This inequality concludes the proof of this theorem.

5. Importance sampling

The expectation representation (6) of the Rao bound brings to mind the possibility of estimating it using simulation. The aforementioned expectation is over the set $\{S_n \leq t/2\}$. If we take $t = \mu n$ with $\mu \leq 1$, as n goes to ∞, the probability of S_n being less than $t/2$ goes to 0 exponentially. This makes a direct simulation infeasible. However, as this section explains, it is possible to construct fast and accurate IS algorithms to compute the Rao bound.

5.1. IS review

Take a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and a measurable integrable function $f : \Omega \to \mathbb{R}$. Suppose that $\hat{\mathrm{P}}$ is a probability measure on $(\Omega, \mathcal{F})$ with respect to which P is absolutely continuous. We have the basic identity

$$\mathrm{E}[f] = \int_\Omega f(\omega)\, \mathrm{d}\mathrm{P}(\omega) = \int_\Omega f(\omega) \frac{\mathrm{d}\mathrm{P}}{\mathrm{d}\hat{\mathrm{P}}}(\omega)\, \mathrm{d}\hat{\mathrm{P}}(\omega) = \hat{\mathrm{E}}\left[f \frac{\mathrm{d}\mathrm{P}}{\mathrm{d}\hat{\mathrm{P}}}\right], \tag{29}$$

where $\mathrm{d}\mathrm{P}/\mathrm{d}\hat{\mathrm{P}}$ is the Radon–Nikodým derivative of P with respect to $\hat{\mathrm{P}}$, and E and $\hat{\mathrm{E}}$ are the expectations with respect to P and $\hat{\mathrm{P}}$, respectively. Identity (29) suggests the following simulation algorithm to compute $\mathrm{E}[f]$. Simulate i.i.d. copies $\omega_1, \omega_2, \ldots, \omega_N$ of ω from $\hat{\mathrm{P}}$ and use the following to estimate $\mathrm{E}[f]$: $\hat{s}_N = (1/N) \sum_{i=1}^N \hat{f}(i)$, $\hat{f}(i) := f(\omega_i)\, \mathrm{d}\mathrm{P}(\omega_i)/\mathrm{d}\hat{\mathrm{P}}$. By the law of large numbers, $\hat{s}_N \to \hat{\mathrm{E}}[f\, \mathrm{d}\mathrm{P}/\mathrm{d}\hat{\mathrm{P}}]$, which, by (29), equals $\mathrm{E}[f]$. Furthermore,

the linearity of the expectation and (29) imply that $\hat{E}[\hat{s}_N] = \hat{E}[\hat{f}(1)] = E[f]$. Therefore, $\hat{s}_N$ is an unbiased estimator of $E[f]$ and it converges to this value as $N \to \infty$. This method of estimating $E[f]$ is called importance sampling (IS). IS is a well-known method for estimating small probabilities; a partial list of articles and books on the subject is [9], [15], [21], [23], [24], [30], [32], and [35].

Because $\hat{s}_N$ is unbiased, its variance depends on $\hat{P}$ only through its second moment, which equals $\hat{E}[\hat{f}^2(1)]/N$. Here N is the number of samples used in the estimation and is taken to be a constant. Therefore, to find a good IS estimator, we try to solve the optimization problem

$$\inf_{\hat{P}} \hat{E}[\hat{f}^2(1)] = \inf_{\hat{P}} \hat{E}\left[\left(f \frac{dP}{d\hat{P}}\right)^2\right] = \inf_{\hat{P}} E\left[f^2 \frac{dP}{d\hat{P}}\right], \tag{30}$$

where the inf is over all $\hat{P}$ with respect to which $\mathbf{1}_{\{f \neq 0\}}$ dP is absolutely continuous.

5.1.1. *Asymptotic analysis.* Suppose now that instead of a single random variable there is a sequence $\{f_n\}$ of random variables satisfying

$$\lim \frac{1}{n} \log E[f_n] =: \gamma > 0. \tag{31}$$

Jensen's inequality and the unbiasedness of $\hat{f}_n(1)$ imply that

$$\liminf_n \frac{1}{n} \log \hat{E}[\hat{f}_n^2(1)] \geq \liminf_n \frac{2}{n} \log \hat{E}[\hat{f}_n(1)] \equiv 2\gamma.$$

A sequence of IS estimators to estimate $E[f_n]$ is said to be *asymptotically optimal* if the corresponding lower bound is achieved, i.e. if

$$\limsup_n \frac{1}{n} \log \hat{E}[\hat{f}_n^2(1)] = \limsup_n \frac{1}{n} \log E\left[f_n^2 \frac{dP}{d\hat{P}_n^*}\right] \leq 2\gamma. \tag{32}$$

5.2. The IS problem for the Rao bound

To estimate the expectation representation (6) of the Rao bound using IS, we set f_n in (31) to

$$f_n = \mathbf{1}_{\{S_n \leq t/2\}} \prod_{j=1}^n 2r(X_j, j),$$

where S_j is the symmetric random walk with increments X_j defined earlier. Theorem 2 says that the γ in (31) for this sequence is $V(0, 0) > 0$. In order to simulate X and S using IS, we specify a sampling distribution $\hat{p}(v \mid x, i)$, $v \in \{0, 1\}$ and $x \in \mathbb{Z}_i$, and execute the following steps: set $S_0 = 0$; at step i of the simulation, sample the increment X_i from the distribution $\hat{p}(\cdot \mid S_i, i)$ and set $S_{i+1} = X_i + S_i$. Note that the distribution of the increment X_i is allowed to depend on the current position of the random walk S. Let $\hat{P}$ denote the probability measure on the sample paths of S_n defined by $\hat{p}(\cdot \mid x, i)$. The Radon–Nikodým derivative $dP/d\hat{P}$ equals $\prod_{i=1}^n 0.5/\hat{p}(X_j^k \mid S_j^k, j)$. Then, the IS estimator of $E[f_n]$ using K sample paths is

$$\frac{1}{K} \sum_{k=1}^K \hat{f}_n(k), \qquad \hat{f}_n(k) := \mathbf{1}_{\{S_n^k \leq t/2\}} \prod_{j=1}^n \frac{r(X_j^k, j)}{\hat{p}(X_j^k \mid S_j^k, j)}, \tag{33}$$

where S^k denotes the kth independent sample path used in the simulation. The increments $\{X^k\}$ are i.i.d. copies of the increment process X sampled from $\hat{p}$. Then, by Theorem 2, the

optimality condition (32) for the IS estimator (33) is

$$\limsup \frac{1}{n} \log \mathrm{E}\left[\mathbf{1}_{\{S_n \leq t/2\}} \prod_{j=1}^{n} \frac{2r(X_j, j)^2}{\hat{p}(X_j \mid S_j, j)} \right] \leq 2V(0, 0). \qquad (34)$$

5.3. Connection between IS and LDs

The connection between IS and LDs has been known, at least heuristically, for a long time; see, for example, [30] in the context of queueing networks. A more rigorous and clear connection has been established recently in [9], [10], [15], [32], and [33].

For the present case, the IS optimization problem (30) becomes

$$\inf_{\hat{p}} \log \mathrm{E}\left[\mathbf{1}_{\{S_n \leq t/2\}} \prod_{j=1}^{n} \frac{2r(X_j, j)^2}{\hat{p}(X_j \mid S_j, j)} \right].$$

This equals

$$\inf_{\hat{p}} \sup_{\{\bar{p}: \, \bar{\mathrm{P}}(S_n \leq t_n/2)\}} \bar{\mathrm{E}}\left[\sum_{j=1}^{n} 2 \log r(X_j, j) - \log \hat{p}(X_j \mid S_j, j) - \log \bar{p}(X_j \mid S_j, j) \right],$$

by a direct generalization of Proposition 1 to the present case. This expression is convex in $\hat{p}$ and concave in $\bar{p}$ and, therefore, the order of the inf and sup can be switched without affecting the result. Once this is done, the optimization in $\hat{p}$ gives the optimizer $\hat{p}^* = \bar{p}$ and the problem reduces to

$$\sup_{\{\bar{p}: \, \bar{\mathrm{P}}(S_n \leq t_n/2)\}} \bar{\mathrm{E}}\left[\sum_{j=1}^{n} 2 \log r(X_j, j) - 2 \log \bar{p}(X_j \mid j, S_j) \right],$$

and this is the same problem as in representation (10) except for a factor of 2. Thus, it is seen that the IS and LD problems are represented by the same optimization problem. The difference between them is that the LD is about the minimum value, whereas IS is about the optimizer.

5.4. An asymptotically optimal IS sampling measure based on LD analysis

There are many asymptotically optimal IS sampling measures to estimate (6). For example, $\bar{p}^*$ of (24). The problem with this change of measure is that it requires the solution of (20) at every step of the random walk S_n. For large n, this is inefficient. A much preferable situation is a fixed change of measure, i.e. a change of measure $\hat{p}$ that does not depend on t and x. In the estimation of the Rao bound, we expect such a change of measure to exist for two reasons: (i) the underlying process is i.i.d. and one-dimensional and (ii) the probability of interest concerns the exit from a region with only one boundary point. For more on these points, we refer the reader to [9], [22], [30], [32], and [35]. Let us now construct a fixed change of measure for our problem.

Let θ_i^* to be the unique minimizers of (12), and define

$$\hat{p}^*(1 \mid x, j) = \theta_i^* \quad \text{if } A_i \leq j/n < A_{i+1}. \qquad (35)$$

The transition probability $\hat{p}^*$ is almost fixed in the sense that it depends only on the step number j and not on the position x of the random walk S_n. The dependence on j is only through the block that j is a member of: each block of the orthogonal array has its own fixed

jump probability θ_i^*; for the steps corresponding to the ith block, we use this fixed probability to sample the increments of S_n.

Before we state and prove our theorem which asserts that an IS estimation based on (35) is asymptotically optimal, we would like to set up several things that will be needed in the proof. Let us begin with the computation of (35). We simply use (16) with $t = 0$ and $x = 0$, and (19). Then

$$\theta_i^* = \frac{s_i - 1}{e^{\lambda^*} + s_i - 1},\tag{36}$$

where λ^* is the unique solution of

$$\sum_{i=1}^{\sigma} a_j \frac{s_j - 1}{e^{\lambda} + s_j - 1} = \frac{\mu}{2}.\tag{37}$$

Therefore, we can compute the IS change of measure $\hat{p}^*$ of (35) by solving the one-dimensional problem (37) to identify λ^* before the simulation begins. Throughout the simulation, no further computation will be necessary to calculate $\hat{p}^*$. This is a great advantage over an IS simulation based on (24) which would require the solution of (37) at every step of the simulated random walk S_j.

5.4.1. *Subsolutions.* A function V that satisfies

$$\sup_{\theta \in [0,1]} \{\tilde{r}(t)\theta + H(\theta) + V_x(x,t)\theta + V_t(x,t)\} \geq 0$$

is called a *subsolution* to the partial differential equation (21). In the next paragraph we will construct a subsolution to (21) and the proof of asymptotic optimality will be a control theoretic verification argument based on this subsolution. This technique is from the 'subsolution approach' to IS which was first developed in the context of queueing networks in [15] and [32]. For a more general development, see [11] and [12]. The paper that precedes these articles and which introduced many of the ideas that underlie the subsolution approach is [9]. Other articles using the approach include [14], [33], and [34].

Usually, the subsolution approach is useful for constructing good IS algorithms. This is the case in most of the aforementioned references. In the present case, we already have a simple algorithm and we will use the approach to prove that our algorithm is optimal. For the subsolution, let us call it W, we set the partial derivative $W_x(x,t) = -2\lambda^*$ for all (x,t) and choose the partial derivative W_t so that W solves (21):

$$W_t(t,x) := 2\lambda^*\theta_i^* - 2\log(s_i - 1)\theta_i^* - 2H(\theta_i^*), \qquad A_i \leq t < A_{i+1}.\tag{38}$$

These define W up to an additive constant. This is sufficient for our needs since only the increments and partial derivatives of W appear in a verification argument. By its construction, W is piecewise affine, continuous, and in fact a solution (and, hence, a subsolution) to (21).

Remark 1. Note that W is a solution to (21) and, as already noted in Theorem 1, so is V defined in (15). Evidently, $W \neq V$. This is a common situation in optimal control, that is, an HJB equation may have many solutions. What makes V unique is that it is the maximal solution to (21). For more on this, we refer the reader to [19].

In addition to being a solution to (21), here are two properties of W that play a key role in the optimality proof.

Lemma 1. *We have $W_x < 0$ and $W(\mu/2, 1) - W(0, 0) = -2V(0, 0)$.*

Proof. Let $g(\lambda)$ denote the left-hand side of (37). The function g is decreasing in λ, with $\lim_{\lambda \to \infty} g(\lambda) = 0$ and $\lim_{\lambda \to -\infty} g(\lambda) = 1$. Furthermore, $(s_i - 1)/s_i \geq 0.5$, because each s_i is an integer greater than 1. Then $g(0) \geq 0.5 > \mu/2$. It follows that $\lambda^* > 0$. By definition, $W_x = -2\lambda^* < 0$, proving the first part of the lemma.

By their definition, the θ_i^* satisfy $\sum_{i=1}^{\sigma} \theta_i^* a_i = \mu/2$; see (36) and (37). Let $x_j = \sum_{i=1}^{j} \theta_i^* a_i$. We can write $W(\mu/2, 1) - W(0, 0)$ as the following telescoping sum:

$$W\left(\frac{\mu}{2}, 1\right) - W(0, 0) = \sum_{i=1}^{\sigma} [W(x_i, A_i) - W(x_{i-1}, A_{i-1})].$$

By definition, W is affine for $t \in (A_{i-1}, A_i)$, with partial derivatives $W_x = -2\lambda^*$ and W_t given in (38); therefore, this last sum equals

$$\sum_{i=1} [W_x(x_{i-1}, A_{i-1})(x_i - x_{i-1}) + W_t(x_{i-1}, A_{i-1})(A_i - A_{i-1})]$$

$$= \sum_{i=1}^{\sigma} [-2\lambda^* \theta_i^* a_i + 2\lambda^* \theta_i^* a_i - 2\log(s_i - 1)\theta_i^* a_i - H(\theta_i^*)a_i]$$

$$= -\sum_{i=1}^{\sigma} a_i (2\log(s_i - 1)\theta_i^* + 2H(\theta_i^*)). \tag{39}$$

The θ_i^* are the unique optimizers of (12); therefore,

$$(39) = -2 \sup\left\{\sum_{i=1}^{\sigma} a_i (\log(s_i - 1)\theta_i + H(\theta_i))\right\},$$

where the sup is subject to (13). The last quantity equals $-2V(0, 0)$ and this concludes the proof of the second part of the lemma.

It follows directly from the definitions of W_t and θ_i^* that

$$W_t + \log\left(e^{W_x}(s_i - 1)^2 \frac{1}{\theta_i^*} + \frac{1}{1 - \theta_i^*}\right) = 0. \tag{40}$$

Let X_i be a Bernoulli random variable with $P(X_i = 1) = 0.5$. For integers x and $A_{i-1}n \leq j < A_i n$, we can represent the previous display probabilistically as

$$E\left[e^{W_x X_i + W_t(x/n, j/n)}(s_i - 1)^{2X_i} \frac{2}{\hat{p}^*(X_i \mid j, x)}\right] = 1. \tag{41}$$

Remark 2. The way it is presented above, (40) seems unmotivated. We should think of it as a multiplicative representation of (21). We can derive (40) directly from (21) first representing the optimization problem in that display as a trivial game and then using a representation result similar to (10). For a similar argument, see [32, Lemma 2.5.2].

Theorem 3. *The IS estimator based on $\hat{p}^*$ of (35) is asymptotically optimal.*

The following proof follows the same steps as the optimality proof given in [15]. It is simpler because there is a fixed time horizon n, so no truncation of time is needed.

Proof of Theorem 3. To ease notation, let $\Delta W_i = W((S_{j+1}, (j+1))/n) - W((S_j, j)/n)$. Define

$$M_k = \prod_{i=1}^{k-1} e^{n\Delta W_i} r(X_i, i)^2 \frac{2}{\hat{p}^*(X_i \mid S_i, i)}.$$

It follows from (41) that M_k is a martingale and that

$$E\left[\mathbf{1}_{\{S_n \leq t_n/2\}} \prod_{i=1}^{n-1} e^{n\Delta W_i} r(X_i, i)^2 \frac{2}{\hat{p}^*(X_i \mid S_i, i)}\right] = 1.$$

We saw in Lemma 1 that $W_x < 0$; therefore,

$$\sum_{i=1}^{n-1} n\Delta W_i = n\left(W\left(\frac{S_n}{n}, 1\right) - W(0, 0)\right) > n\left(W\left(\frac{\mu}{2}, 1\right) - W(0, 0)\right) = -2nV(0, 0)$$

on $\{S_n < \mu n/2\}$. The last two displays imply that

$$e^{2nV(0,0)} \geq E\left[\mathbf{1}_{\{S_n \leq t_n/2\}} \prod_{i=1}^{n-1} r(X_i, i)^2 \frac{2}{\hat{p}^*(X_i \mid S_i, i)}\right].$$

Taking the log of both sides, dividing by n, and letting n go to ∞ proves that (34) holds for the change of measure $\hat{p}^*(\cdot \mid \cdot, \cdot)$, i.e. the IS change algorithm defined by this change of measure is asymptotically optimal, which is what we wanted to prove.

5.5. A numerical study of the coefficient of variation

A natural measure of the performance of an estimator E of an expectation α is the ratio $\sqrt{\text{var}(E)}/\alpha$, which is called the coefficient of variation of the estimator. If the estimator is unbiased, as is the case for the estimators in the present paper, $\sqrt{E[E^2]}/\alpha$ is another performance measure that is equivalent to the coefficient of variation. With a slight abuse of notation, we will refer to this last ratio as the coefficient of variation.

The asymptotic optimality of the IS estimator of the previous section implies that the coefficient of variation of that estimator grows subexponentially in row length. We now present numerical evidence that this ratio in fact grows like $n^{1/4}$.

For integers $0 \leq x \leq t/2$ and $0 \leq k \leq n$, define

$$S(x, k) = E\left[\mathbf{1}_{\{x+S_{n-j} \leq t/2\}} \prod_{j=k+1}^{n} \frac{2r(X_j, j)^2}{\hat{p}^*(X_j \mid S_j, j)}\right].$$

Note that the second moment of the IS estimator (33) using $K = 1$ samples and using the change of measure (35) is $S(0, 0)$. Similarly to the recursive formula (7) for the Rao bound, we have the following recursion for $S(x, k)$:

$$S(x, k) = S(x+1, k+1) \frac{r^2(x, k)}{\hat{p}^*(1 \mid x, k)} + S(x, k+1) \frac{1}{\hat{p}^*(0 \mid x, k)}. \tag{42}$$

Figure 2 depicts the coefficient of variation $\sqrt{S(0, 0)}/M(0, 0)$ of the IS estimator based on (35) as a function of the row length n for the parameter values

$$\sigma = 8, \qquad s = (15, 15, 10, 7, 5, 8, 12, 13),$$

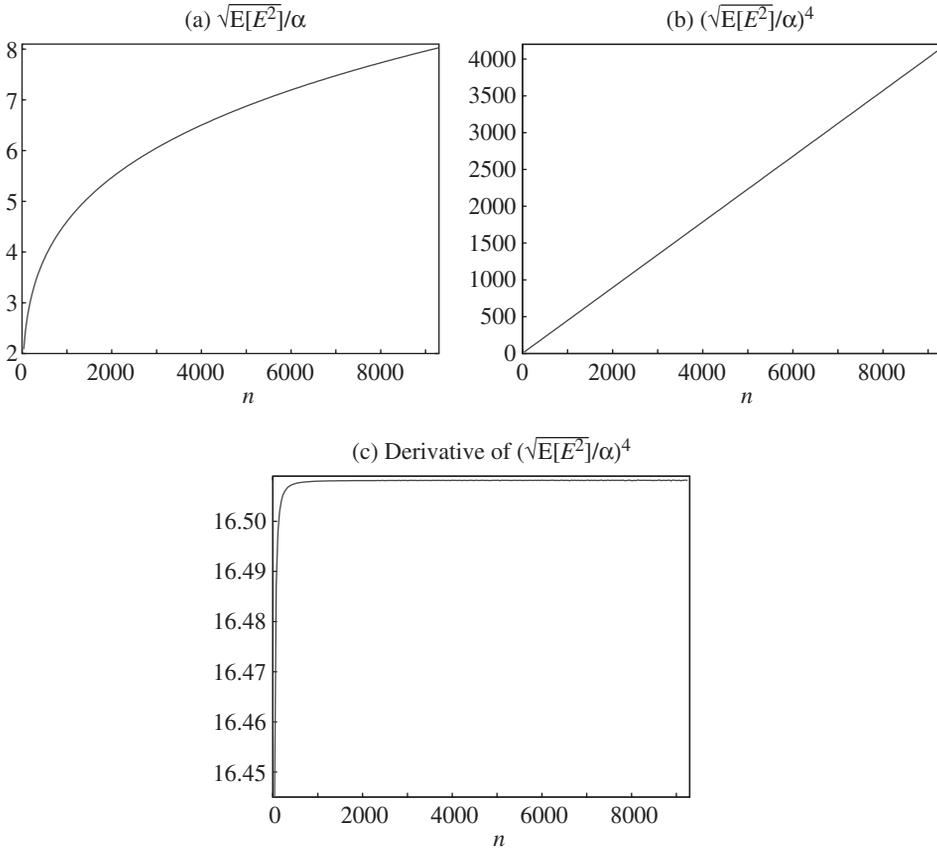$$a = \tfrac{1}{37}(5, 5, 5, 5, 5, 4, 4, 4), \qquad 37 \leq n \leq 9287.$$

FIGURE 2: The coefficient of variation.

Figure 2(a) depicts the coefficient of variation $\sqrt{S(0,0)}/M(0,0)$ itself, Figure 2(b) depicts its fourth power, and Figure 2(c) depicts the discrete derivative of the fourth power (the discrete derivative of a sequence of real numbers $\{x_n\}$ is $\{x_{n+1} - x_n\}$). The values depicted were computed with PARI/GP (see http://pari.math.u-bordeaux.fr/) using (7) and (42).

These computations suggest that the coefficient of variation of the IS estimator (33) with $\hat{p} = \hat{p}^*$ grows like $n^{1/4}$ in the row length n when $K = 1$, i.e. when a single sample is used. Assuming that this growth rate is accurate, the coefficient of variation of the same estimator as a function of the row length n will stay bounded if a $K = \sqrt{n}$ number of independent samples are used in the estimation. With $\sqrt{n}$ many samples, the IS estimation requires $O(n^{3/2})$ operations, which is an improvement over the complexity $O(n^2)$ of the deterministic algorithm defined by (7).

5.5.1. *Comments on a rigorous derivation of the growth rate of the relative error.* Any proof that the coefficient of variation of the IS estimator defined by (35) grows like $O(n^{1/4})$ will need more precise asymptotics than provided by the LD analysis that the present work undertakes, which computes exponential decay and growth rates. Therefore, a key ingredient of a proof of this subpolynomial growth rate will involve a theorem analogous to that of Bahadur and Rao's

(see [7, Theorem 3.7.4]), which describes the exact asymptotics of $P(S_n/n > \alpha)$, where S_n is a sum of a real-valued i.i.d. sequence. It appears possible to extend this result to cover the problem in the present work. However, this extension does not seem trivial and the methods involved are different from those used so far and, therefore, we have not attempted it in the present work.

Assuming that a Bahadur–Rao-type result is available for the present problem, there are several ways to proceed to obtain the $O(n^{1/4})$ growth rate. The first is a new method developed in [2] which is related to the subsolution method that we have used in our asymptotic analysis. In that paper, the authors constructed Lyapunov functions from related subsolutions to obtain refined asymptotics for the second moment of the IS estimator. When combined with the Bahadur–Rao theorem, these asymptotics imply polynomial growth rates of the relative error of the IS estimator. The problems studied in [2] have boundaries that make several exit directions likely. This leads to subsolutions whose gradient is a weighted average of several vectors, and Blanchet *et al.* [2] constructed its Lyapunov functions using these weights. On the other hand, the expectation (6) studied in the present work concerns an exit problem with a single point of exit and, therefore, no weights appear in the subsolution W given in (38). After an initial effort, it did not appear immediately clear to us how to define a precise Lyapunov function from W and we hope to look at this problem in a future work.

A second approach is to follow the argument in [3, Section 2], which proves the same growth rate for a static estimator of the probability $P(S_n/n > \alpha)$. This argument is a direct computation based on Bahadur–Rao exact asymptotics, LD asymptotics, and the simple static nature of the IS estimator. It appears that this approach would yield the desired growth rate once a Bahadur–Rao-type result becomes available.

6. The Gilbert–Varshamov bound

The results derived for the Rao bound (3) in Sections 3, 4, and 5 can be derived for the Gilbert–Varshamov (GV) bound (4). The analysis and the results are essentially the same, the main difference is that μ replaces $\mu/2$ in (23) and other similar places.

The key quantity in (4) is

$$\sum_{i=0}^{t-1} \sum_{u_1, u_2, \ldots, u_\sigma, \sum u_m = i} s_\sigma \binom{l_\sigma - 1}{u_\sigma - 1} (s_\sigma - 1)^{u_\sigma - 1} \prod_{m=1}^{\sigma - 1} \binom{l_m}{u_m} (s_m - 1)^{u_m}. \tag{43}$$

Let S_n, X_i, and r be defined as in Section 3. The expectation representation of (43) is

$$s_\sigma E\left[\mathbf{1}_{\{S_{n-1} \leq t-1\}} \prod_{i=1}^{n-1} 2r(X_i, i) \right]. \tag{44}$$

This is exactly the same as (6), except for the following differences.

1. The expectation is over a random walk that takes $n - 1$ steps, rather than n.

2. There is a s_σ factor in front.

3. The expectation is over those trajectories such that $S_{n-1} \leq t - 1$ rather than $S_n \leq t/2$.

As in Section 4, the asymptotic analysis of (44) will involve a $(1/n)$ log scaling. Under this

scaling, the asymptotics of (44) are the same as that of

$$\mathrm{E}\left[\mathbf{1}_{\{S_n \leq t\}} \prod_{i=1}^{n} 2r(X_i, i)\right]. \tag{45}$$

Let l_i, a_i, t_i, and μ be as in (8). Theorem 2 implies that

$$\lim_{n} \frac{1}{n} \log \mathrm{E}\left[\mathbf{1}_{\{S_n \leq t_n\}} \prod_{j=1}^{n} 2r(X_j, j)\right] = \sup\left\{\sum_{i=1}^{\sigma} a_i(\theta_i \log(s_i - 1) + H(\theta_i))\right\}, \tag{46}$$

where the sup is over

$$\theta_i \in (0, 1), \qquad \langle a, \theta \rangle \leq \mu. \tag{47}$$

If $\mu \in (0.5, 1)$, $\{S_n \leq \mu n\}$ is not a rare event and there is no need for IS to simulate (45) effectively; we can use straightforward Monte Carlo for this purpose. Otherwise, Theorem 3 implies that the minimizers of (46) define an asymptotically optimal IS change of measure to estimate (45).

7. Numerical examples

We used the Octave numerical computation environment [16] for the numerical computations in this section.

7.1. The Rao bound

Example 1. Consider the following parameter values for an orthogonal array: $\sigma = 4$, alphabet sizes $s = [13, 10, 7, 5]$, block lengths $l = [20, 20, 20, 20]$, and $t = 4$. Then $n = 80$, the scaled strength parameter $\mu = 0.05$, and length parameters $a = [0.25, 0.25, 0.25, 0.25]$.

For this example, the exact Rao bound can be computed in two ways: using either the original formula (3) or the recursive algorithm (7). Both of these algorithms very quickly yield the value 190 051.

We solve (12) with the above parameter values to get the LD decay rate $V(0, 0) = 0.1681$. Then the LD estimate of the Rao bound is $e^{V(0,0)n} = e^{13.44} = 689\,760$, which is about three times larger than the actual bound found above. This type of inaccuracy is expected since an LD analysis only identifies the exponential growth rate.

We know from Section 5 that if the optimizers of (12) are used as an IS change of measure in (33), the resulting IS algorithm is asymptotically optimal. The optimizers of (12) for the above parameter values is $\theta^* = (0.0383, 0.0290, 0.0195, 0.0131)$. In Table 1 we present five estimation results using this algorithm with $K = 2000$ sample paths. The standard error column

TABLE 1.

Estimation	Estimate $\hat{s}_K$	Standard error	95% confidence intervals	Scaling
1	1.94	0.06	[1.82, 2.06]	$\times 10^5$
2	1.82	0.06	[1.70, 1.94]	$\times 10^5$
3	1.83	0.06	[1.71, 1.95]	$\times 10^5$
4	1.82	0.06	[1.70, 1.94]	$\times 10^5$
5	1.92	0.06	[1.80, 2.04]	$\times 10^5$

TABLE 2.

Estimation	Estimate $\hat{s}_K$	Standard error	95 % confidence intervals	Scaling
1	2.49	0.14	[2.21, 2.77]	$\times 10^{38}$
2	2.58	0.14	[2.30, 2.86]	$\times 10^{38}$
3	2.43	0.14	[2.15, 2.71]	$\times 10^{38}$
4	2.35	0.14	[2.07, 2.63]	$\times 10^{38}$
5	2.55	0.14	[2.27, 2.83]	$\times 10^{38}$

presents the estimated standard deviation $\hat{\sigma}(\hat{s}_K)$. The informal 95% confidence intervals are $[\hat{s}_K - 2\hat{\sigma}(\hat{s}_K), \hat{s}_K + 2\hat{\sigma}(\hat{s}_K)]$.

The results of Table 1 suggest that the asymptotically optimal IS scheme derived in Section 5 also performs well in practice. All of the estimates are close to the actual value, the formal confidence intervals are tight and they all happen to contain the exact Rao bound.

Example 2. Now consider $\sigma = 40$, alphabet sizes $s_i = 20 + i$, block lengths $l_i = 20$, $i = 1, 2, \ldots, 40$, and strength parameter $t = 20$. Then $n = 800$, $\mu = 0.025$, and $a_i = 0.025$.

For this example, the complexity analysis (5) in Section 2 indicates that the direct computation of (3) would require about 10^{41} operations, which of course is not possible to perform in any reasonable amount of time. The recursive algorithm (7) yields 2.57×10^{38} in a second or less.

The LD decay rate $V(0, 0)$ turns out to be 0.113 for this problem and the corresponding LD estimate of (3) is $e^{V(0,0)n} = e^{90.4} = 1.82 \times 10^{38}$, which is, at the scale of 10^{38}, close to the exact value.

The optimizers of (12) is a forty-dimensional vector and is inconvenient to list explicitly. The IS estimate based on (3) using these optimizers and $K = 1000$ samples are given in Table 2.

As in the first example, practical performance of the IS estimator is very good here. All the estimates are close to the exact value, the confidence intervals are tight and they all happen to contain the exact value. The run time for each estimation is around a second.

7.2. The GV bound

Let us continue with the previous parameter values. The computation for this bound is the same as the Rao bound. In the example below, we calculate the expectation (44) rather than the actual quantity (43), which is a multiple of the expectation. We can use our recursive algorithm (7) to compute the exact GV bound (4) to be 3.13×10^{71}. The LD growth rate $V(0, 0) = 0.2088$ and the LD estimate of the GV bound is $e^{V(0,0)800} = 2.85 \times 10^{71}$. The IS results are given in Table 3. Once again, the simulation results are accurate and reliable.

TABLE 3.

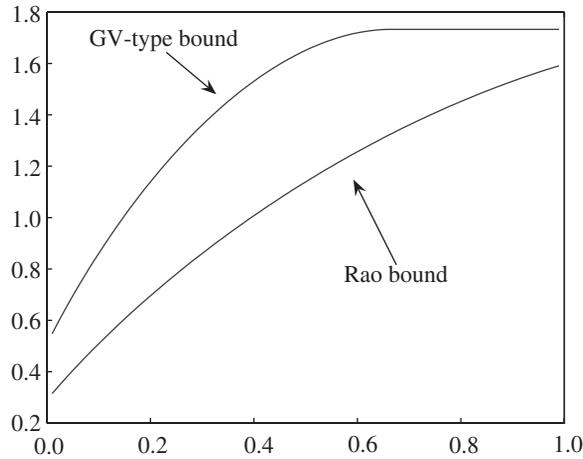| Estimation	Estimate $\hat{s}_K$	Standard error	95 % confidence intervals	Scaling
1	3.53	0.23	[3.07, 3.99]	$\times 10^{71}$
2	3.47	0.23	[3.01, 3.93]	$\times 10^{71}$
3	3.42	0.23	[2.96, 3.88]	$\times 10^{71}$
4	3.28	0.22	[2.84, 3.72]	$\times 10^{71}$
5	2.83	0.20	[2.44, 3.23]	$\times 10^{71}$

FIGURE 3: A comparison of the Rao and GV bounds.

7.3. A comparison

Set $\sigma = 4$, $s_1 = 2$, $s_2 = 4$, $s_3 = 8$, $s_4 = 16$, and $a_i = 0.25$. Figure 3 depicts the graphs of the Rao and the GV asymptotic bounds for these parameter values as a function of the strength parameter $\mu \in [0, 1]$.

The large gap between them is due to the difference of a factor of 2 between the constraints of the Rao and the GV bounds. The GV bound is flat for larger values of μ, because, for these values of μ, the unique global maximizer of (46) satisfies constraint (47).

References

[1] BLANCHET, J. H. (2007). Efficient importance sampling for binary contingency tables. Preprint. Available at http://arxiv.org/abs/0908.0999v1.

[2] BLANCHET, J. H., GLYNN, P. AND LEDER, K. (2009). On Lyapunov inequalities and subsolutions for efficient importance sampling. Preprint.

[3] BLANCHET, J. H., LEDER, K. AND GLYNN, P. W. (2008). Efficient simulation of light-tailed sums: an old folk song sung to a faster new tune. In *Monte Carlo and Quasi-Monte Carlo Methods 2008*, eds P. L. Ecuyer and A. B. Owen, Springer, Berlin, pp. 227–248.

[4] BLITZSTEIN, J. AND DIACONIS, P. (2006). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. Preprint.

[5] CHEN, Y., DINWOODIE, I. H. AND SULLIVANT, S. (2006). Sequential importance sampling for multiway tables. *Ann. Statist* **34,** 523–545.

[6] CHEN, Y., DIACONIS, P., HOLMES, S. P. AND LIU, J. S. (2005). Sequential Monte Carlo methods for statistical analysis of tables. *J. Amer. Statist. Assoc.* **100,** 109–120.

[7] DEMBO, A. AND ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications* (Appl. Math. **38**), 2nd edn. Springer, New York.

[8] DUPUIS, P. AND ELLIS, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley, New York.

[9] DUPUIS, P. AND WANG, H. (2004). Importance sampling, large deviations, and differential games. *Stoch. Stoch. Reports* **76,** 481–508.

[10] DUPUIS, P. AND WANG, H. (2005). Adaptive importance sampling for uniformly recurrent Markov chains. *Ann. Appl. Prob.* **15,** 1–38.

[11] DUPUIS, P. AND WANG, H. (2005). Subsolutions of an Isaacs equation and efficient schemes for importance sampling: convergence analysis. Preprint. Available at http://www.dam.brown.edu/people/huiwang.

[12] DUPUIS, P. AND WANG, H. (2005). Subsolutions of an Isaacs equation and efficient schemes for importance sampling: examples and numerics. Preprint. Available at http://www.dam.brown.edu/lcds/publications.

[13] DUPUIS, P., ISHII, H. AND SONER, H. M. (1990). A viscosity solution approach to the asymptotic analysis of queueing systems. *Ann. Prob.* **18,** 226–255.

[14] DUPUIS, P., LEDER, K. AND WANG, H. (2007). Large deviations and importance sampling for a tandem network with slow-down. *Queueing Systems* **57,** 71–83.

[15] DUPUIS, P., SEZER, A. D. AND WANG, H. (2007). Dynamic importance sampling for queueing networks. *Ann. Appl. Prob.* **17,** 1306–1346.

[16] EATON, J. W. (2002). *GNU Octave Manual*. Network Theory Limited.

[17] FENG, K., XU, L. AND HICKERNELL, F. J. (2006). Linear error-block codes. *Finite Fields Appl.* **12,** 638–652.

[18] FLEMING, W. H. (1977/78). Exit probabilities and optimal stochastic control. *Appl. Math. Optimization* **4,** 329–346.

[19] FLEMING, W. H. AND SONER, H. M. (1992). *Controlled Markov Processes and Viscosity Solutions*. Springer, New York.

[20] FLEMING, W. H. AND TSAI, C. P. (1981). Optimal exit probabilities and differential games. *Appl. Math. Optimization* **7,** 253–282.

[21] GLASSERMAN, P. (2004). Monte Carlo Methods in Financial Engineering. (Appl. Math. **53**). Springer, New York.

[22] GLASSERMAN, P. AND KOU, S.-G. (1995). Analysis of an importance sampling estimator for tandem queues. *ACM Trans. Model and Comput. Simul.* **5,** 22–42.

[23] GOERTZEL, G. (1949). Quota sampling and importance functions in stochastic solution of particle problems. Tech. Rep. 434, Oak Ridge National Laboratory.

[24] HAMMERSLEY, J. M. AND HANDSCOMB, D. C. (1964). *Monte Carlo Methods*. Methuen & Co, London.

[25] HEDAYAT, A. S., SLOANE, N. J. A. AND STUFKEN, J. (1999). *Orthogonal Arrays*. Springer, New York.

[26] LING, S. AND ÖZBUDAK, F. (2007). Constructions and bounds on linear error-block codes. *Des. Codes Cryptogr.* **45,** 297–316.

[27] MACWILLIAMS, F. J. AND SLOANE, N. J. A. (1977). *The Theory of Error-Correcting Codes. I*. North-Holland, Amsterdam.

[28] MACWILLIAMS, F. J. AND SLOANE, N. J. A. (1977). *The Theory of Error-Correcting Codes. II*. North-Holland, Amsterdam.

[29] NIEDERREITER, H. AND ÖZBUDAK, F. (2007). Improved asymptotic bounds for codes using distinguished divisors of global function fields. *SIAM J. Discrete Math.* **21,** 865–899.

[30] PAREKH, S. AND WALRAND, J. (1989). A quick simulation method for excessive backlogs in networks of queues. *IEEE Trans. Automatic Control* **34,** 54–66.

[31] RADHAKRISHNA RAO, C. (1947). Factorial experiments derivable from combinatorial arrangements of arrays. *Suppl. J. R. Statist. Soc.* **9,** 128–139.

[32] SEZER, A. D. (2005). Dynamic importance sampling for queueing networks. Doctoral Thesis, Division of Applied Mathematics, Brown University.

[33] SEZER, A. D. (2008). Asymptotically optimal importance sampling for Jackson networks with a tree topology. Preprint. Available at http://arxiv.org/abs/0708.3260v3.

[34] SEZER, A. D. (2009). Importance sampling for a Markov modulated queuing network. *Stoch. Process. Appl.* **119,** 491–517.

[35] SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4,** 673–684.