EFFICIENT SIMULATION OF LARGE DEVIATION EVENTS FOR SUMS OF RANDOM VECTORS USING SADDLE-POINT REPRESENTATIONS

ANKUSH AGARWAL,∗ **
SANTANU DEY ** *** AND
SANDEEP JUNEJA,∗ **** Tata Institute of Fundamental Research

Abstract

We consider the problem of efficient simulation estimation of the density function at the tails, and the probability of large deviations for a sum of independent, identically distributed (i.i.d.), light-tailed, and nonlattice random vectors. The latter problem besides being of independent interest, also forms a building block for more complex rare event problems that arise, for instance, in queueing and financial credit risk modeling. It has been extensively studied in the literature where state-independent, exponential-twisting-based importance sampling has been shown to be asymptotically efficient and a more nuanced state-dependent exponential twisting has been shown to have a stronger bounded relative error property. We exploit the saddle-point-based representations that exist for these rare quantities, which rely on inverting the characteristic functions of the underlying random vectors. These representations reduce the rare event estimation problem to evaluating certain integrals, which may via importance sampling be represented as expectations. Furthermore, it is easy to identify and approximate the zero-variance importance sampling distribution to estimate these integrals. We identify such importance sampling measures and show that they possess the asymptotically vanishing relative error property that is stronger than the bounded relative error property. To illustrate the broader applicability of the proposed methodology, we extend it to develop an asymptotically vanishing relative error estimator for the practically important expected overshoot of sums of i.i.d. random variables.

Keywords: Rare event simulation; importance sampling; saddle-point approximation; Fourier inversion; large deviations

2010 Mathematics Subject Classification: Primary 65C05; 60E10; 60F10
Secondary 65C50; 65T99

1. Introduction

Let \((X_i: i \geq 1)\) denote a sequence of independent, identically distributed (i.i.d.), light-tailed (their moment generating function is finite in a neighborhood of 0), nonlattice (modulus of their characteristic function is strictly less than 1) random vectors taking values in \(\mathbb{R}^d\) for \(d \geq 1\). In this paper (see also the preliminary version of this paper [7]) we consider the problem of efficient simulation estimation of the probability density function of \(\bar{X}_n = n^{-1} \sum_{i=1}^n X_i\) at points away from \(E X_i\), and the tail probability \(P(\bar{X}_n \in A)\) for sets \(A\) that do not contain

Received 6 October 2011; revision received 18 October 2012.

* Postal address: STCS, Tata Institute of Fundamental Research, 1 Homi Bhabha Road, Mumbai, India.
** Email address: anush@tcs.tifr.res.in
*** Email address: dsantanu@tcs.tifr.res.in
**** Email address: juneja@tifr.res.in
EX_1 and are essentially affine transformations of the nonnegative orthant of \( \mathbb{R}^d \). We develop an efficient simulation estimation methodology for these rare quantities that exploits the well-known saddle-point representations for the probability density function of \( \tilde{X}_n \) obtained from Fourier inversion of the characteristic function of \( X_1 \) (see, e.g. [5, pp. 75–93], [6], and [14, pp. 26–27]). Furthermore, using Parseval’s relation, similar representations for \( \mathbb{P}(\tilde{X}_n \in \mathcal{A}) \) are easily developed. To illustrate the broader applicability of the proposed methodology, we also develop a similar representation for \( \mathbb{E}(\tilde{X}_n : \tilde{X}_n \geq a) \) (the authors thank the editor for suggesting this application) in a single-dimension setting (\( d = 1 \)) for \( a > \mathbb{E}X_1 \), and use it to develop an efficient simulation methodology for this quantity as well.

The problem of efficient simulation estimation of the tail probability density function has not been studied in the literature, although, from a practical viewpoint, it is clear that the shape of such density functions provides a great deal of insight into the tail behavior of the sums of random variables. Another potential application maybe in the maximum likelihood framework for parameter estimation where closed-form expressions for density functions of observed outputs are not available, but simulation-based estimators provide an accurate proxy. The problem of efficiently estimating \( \mathbb{P}(\tilde{X}_n \in \mathcal{A}) \) via importance sampling, besides being of independent importance, may also be considered a building block for more complex problems involving many streams of i.i.d. random variables (see, e.g. [17] for a queueing application and [12] for applications in credit risk modeling). This problem has been extensively studied in the rare event simulation literature (see, e.g. [2], [9], [11], [13], [18], and [19]). Essentially, the literature exploits the fact that the zero-variance importance sampling estimator for \( \mathbb{P}(\tilde{X}_n \in \mathcal{A}) \), though unimplementable, has a Markovian representation. This representation may be exploited to come up with provably efficient, implementable approximations (see [1, pp. 163–171] and [15]).

Sadowsky and Bucklew [19] (see also [4]) developed exponential-twisting-based importance sampling algorithms to arrive at unbiased estimators for \( \mathbb{P}(\tilde{X}_n \in \mathcal{A}) \) that they proved were asymptotically or weakly efficient (as per the current standard terminology in the rare event simulation literature); see, e.g. [1] and [15] for an introduction to rare event simulation. Popular efficiency criteria for rare event estimators are also discussed later in Section 2.1. The importance sampling algorithms proposed in [19] were state independent in that each \( \tilde{X}_{k+1} \) was generated from a distribution independent of the previously generated \( (X_i : i \leq k) \). Blanchet et al. [2] also considered the problem of estimating \( \mathbb{P}(\tilde{X}_n \in \mathcal{A}) \) where they introduced state-dependent, exponential-twisting-based importance sampling distributions (the distribution of generated \( X_{k+1} \) depended on the previously generated \( (X_i : i \leq k) \)). They showed that, when done correctly, such an algorithm is strongly efficient, or, equivalently, has the bounded relative error property.

The problem of efficient estimation of the expected overshoot \( \mathbb{E}((\tilde{X}_n - a) : \tilde{X}_n \geq a) \) is of considerable importance in finance and insurance settings. To the best of our knowledge, this is the first paper that directly tackles this estimation problem.

As mentioned earlier, in this paper we exploit the saddle-point-based representations of the rare event quantities considered. These representations allow us to write the quantity of interest \( \alpha_n \) as a product \( c_n \times \beta_n \), where \( c_n \sim \alpha_n \) (that is, \( c_n/\alpha_n \to 1 \) as \( n \to \infty \)) and is known in closed form. So the problem of interest is the estimation of \( \beta_n \), which is an integral of a known function. Note that \( \beta_n \to 1 \) as \( n \to \infty \). In the literature, asymptotic expansions for \( \beta_n \) exist; however, they require computation of third- and higher-order derivatives of the log-moment generating function of \( X_1 \). This is particularly difficult in higher dimensions. In addition, it is difficult to control the bias in such approximations. As we note later in the numerical experiments, these biases can be significant even when the probabilities are as small as of order \( 10^{-9} \). In the insurance and financial industry, simulation, with its associated variance reduction techniques,
is the preferred method for tail risk measurement even when asymptotic approximations are available (since these approximations are typically poor in the range of practical interest; see, e.g. [12]).

In our analysis we note that the integral \( \beta_n \) can be expressed as an expectation of a random variable using importance sampling. Furthermore, the zero-variance estimator for this expectation is easily ascertained. We approximate this estimator by an implementable importance sampling distribution and prove that the resulting unbiased estimator of \( \alpha_n \) has the desirable asymptotically vanishing relative error property. More tangibly, the estimator of the integral \( \beta_n \) has the property that its variance converges to 0 as \( n \to \infty \). An additional advantage of the proposed approach over existing methodologies for estimating \( \mathbb{P}(X_n \in \mathcal{A}) \) and related rare quantities is that while these methods require \( O(n) \) computational effort to generate each sample output, our approach per sample requires small and fixed effort independent of \( n \).

The use of saddle-point methods to compute tail probabilities has a long and rich history (see, e.g. [5], [14], and [16]). To the best of our knowledge, the proposed methodology is the first attempt to combine the expanding literature on rare event simulation with the classical theory of saddle-point approximations.

The rest of the paper is organized as follows. In Section 2 we briefly review the popular performance evaluation measures used in rare event simulation, and the existing literature on estimating \( \mathbb{P}(X_n \in \mathcal{A}) \). Then in Section 3 we develop an importance sampling estimator for the density of \( X_n \) and show that it has asymptotically vanishing relative error. In Section 4 we devise an integral representation for \( \mathbb{P}(X_n \in \mathcal{A}) \) and develop an importance sampling estimator for it, and again prove that it has asymptotically vanishing relative error. In this section we also discuss how this methodology can be adapted similarly to develop an asymptotically vanishing relative error estimator for \( \mathbb{E}(X_n : X_n \geq a) \) in a single-dimension setting. In Section 5 we report the results of a few numerical experiments to support our analysis. We end with a brief conclusion in Section 6. For brevity, proofs similar to relevant known results, routine technicalities, figures, and some numerical experiments are omitted. These can be found in [8], a more elaborate version of this paper.

### 2. Rare event simulation, a brief review

Let \( \alpha_n = \mathbb{E}_n Y_n = \int Y_n d\mathbb{P}_n \) be a sequence of rare event expectations in the sense that \( \alpha_n \to 0 \) as \( n \to \infty \) for nonnegative random variables \( Y_n : n \geq 1 \). Here \( \mathbb{E}_n \) is the expectation operator under \( \mathbb{P}_n \). For example, when \( \alpha_n = \mathbb{P}(B_n) \), \( Y_n \) corresponds to the indicator of the event \( B_n \).

Naive simulation for estimating \( \alpha_n \) requires generating many i.i.d. samples of \( Y_n \) under \( \mathbb{P}_n \). Their average then provides an unbiased estimator of \( \alpha_n \). Central limit theorem based approximations then provide an asymptotically valid confidence interval for \( \alpha_n \) (under the assumption that \( \mathbb{E}_n Y_n^2 < \infty \)). Importance sampling involves expressing \( \alpha_n = \int Y_n L_n d\mathbb{P}_n = \mathbb{E}_n(Y_n L_n) \), where \( \mathbb{P}_n \) is another probability measure such that \( \mathbb{P}_n \) is absolutely continuous with respect to \( \mathbb{P}_n \), with \( L_n = d\mathbb{P}_n/d\mathbb{P}_n \) denoting the associated Radon–Nikodym derivative, or the likelihood ratio, and \( \mathbb{E}_n \) is the expectation operator under \( \mathbb{P}_n \). The importance sampling unbiased estimator \( \hat{\alpha}_n \) of \( \alpha_n \) is obtained by taking an average of generated i.i.d. samples of \( Y_n L_n \) under \( \mathbb{P}_n \). Note that by setting \( d\mathbb{P}_n = Y_n d\mathbb{P}_n/\mathbb{E}_n(Y_n) \), the simulation output \( Y_n L_n \) is \( \mathbb{E}_n(Y_n) \) almost surely, signifying that such a \( \mathbb{P}_n \) provides a zero-variance estimator for \( \alpha_n \).

#### 2.1. Popular performance measures

Note that the relative width of the confidence interval obtained using the central limit theorem approximation is proportional to the ratio of the standard deviation of the estimator divided by
Throughout this paper, we assume that $0 \in \Lambda_1$ (see [1, p. 159]). Here $\text{var}(\hat{\alpha}_n)$ denotes the variance of the proposed importance sampling estimator (see [1, p. 159]). Here $\text{var}(\hat{\alpha}_n)$ denotes the variance of the proposed importance sampling estimator (see [1, p. 159]). Here $\text{var}(\hat{\alpha}_n)$ denotes the variance of the proposed importance sampling estimator (see [1, p. 159]). Here $\text{var}(\hat{\alpha}_n)$ denotes the variance of the proposed importance sampling estimator (see [1, p. 159]).

A given sequence of estimators $(\hat{\alpha}_n : n \geq 1)$ for quantities $(\alpha_n : n \geq 1)$ is said to be weakly efficient or asymptotically efficient if $\lim \sup_{n \to \infty} \sqrt{\text{var}(\hat{\alpha}_n) / \alpha_n^{1-\varepsilon}} < \infty$ for all $\varepsilon > 0$; to be strongly efficient or have bounded relative error if $\lim \sup_{n \to \infty} \sqrt{\text{var}(\hat{\alpha}_n) / \alpha_n} < \infty$; and to have asymptotically vanishing relative error if $\lim_{n \to \infty} \sqrt{\text{var}(\hat{\alpha}_n) / \alpha_n} = 0$.

3. Efficient estimation of the probability density function of $X_n$

In this section we first develop a saddle-point-based representation for the probability density function (PDF) of $X_n$ in Proposition 1 (for a proof, see, e.g. [5, pp. 89–90], [6], [8], and [14, pp. 12–14]). We then develop an approximation to the zero-variance estimator for this PDF. Our main result is Theorem 1, where we prove that the proposed estimator has an asymptotically vanishing relative error.

Some notation is needed in our analysis. Recall that $(X_i : i \geq 1)$ denotes a sequence of i.i.d. light-tailed random vectors taking values in $\mathbb{R}^d$. Let $(X_i^1, \ldots, X_i^d)$ denote the components of $X_i$, each taking value in $\mathbb{R}$. Let $F(\cdot)$ denote the distribution function of $X_i$. Denote the moment generating function of $F$ by $M(\cdot)$, so that

$M(\theta) := \mathbb{E}(e^{\theta \cdot X_i}) = \mathbb{E}(e^{\theta_1 X_i^1 + \theta_2 X_i^2 + \cdots + \theta_d X_i^d})$,

where $\theta = (\theta_1, \theta_2, \ldots, \theta_d)$ and, for $x, y \in \mathbb{R}^d$, the Euclidean inner product between them is denoted by $x \cdot y := x_1 y_1 + x_2 y_2 + \cdots + x_d y_d$. The characteristic function (CF) of $X_i$ is given by

$\phi(\theta) := \mathbb{E}(e^{i \theta \cdot X_i}) = \mathbb{E}(e^{i \theta_1 X_i^1 + i \theta_2 X_i^2 + \cdots + i \theta_d X_i^d})$,

where $i = \sqrt{-1}$. In this paper we assume that the distribution of $X_i$ is nonlattice, which means that $|\phi(\theta)| < 1$ for all $\theta \in \mathbb{R}^d \setminus \{0\}$.

Let $\Lambda(\theta) := \ln M(\theta)$ denote the cumulant generating function (CGF) of $X_i$. We define $\Theta$ to be the effective domain of $\Lambda(\theta)$, that is,

$\Theta := \{\theta = (\theta_1, \theta_2, \ldots, \theta_d) \in \mathbb{R}^d \mid \Lambda(\theta) < \infty\}$.

Throughout this paper, we assume that $0 \in \Theta^0$, the interior of $\Theta$.

Denote the Euclidean norm of $x \in \mathbb{R}^d$ by $|x| := \sqrt{x \cdot x}$. For a square matrix $A$, $\det(A)$ denotes the determinant of $A$, while the norm of $A$ is denoted by $\|A\| := \max_{|x|=1} |Ax|$. Let $A''(\theta)$ denote the Hessian of $\Lambda(\theta)$ for $\theta \in \Theta^0$. Whenever this is strictly positive definite, let $A(\theta)$ be the inverse of the unique square root of $A''(\theta)$.

**Proposition 1.** Suppose that $A''(\theta)$ is strictly positive definite for some $\theta \in \Theta^0$. Furthermore, suppose that $|\psi(\cdot)|^\gamma$ is integrable for some $\gamma \geq 1$. Then $f_n$, the density function of $X_n$, exists for all $n \geq \gamma$ and its value at any point $x_0$ is given by

$$f_n(x_0) = \left(\frac{n}{2\pi}\right)^{d/2} \frac{\exp[n(A(\theta) - \theta \cdot x_0)]}{\sqrt{\det(A''(\theta))}} \int_{v \in \mathbb{R}^d} \psi\left(n^{-1/2} A(\theta)v, \theta, n\right) \times \phi(v) \, dv,$$

(1)
We now argue that $\psi(y, \theta, n) = \exp[n \times \eta(y, \theta)]$ and
\[
\eta(y, \theta) = \frac{1}{2} y^2 \Lambda''(\theta) y + \Lambda(\theta + iy) - (\theta + iy) \cdot x_0 - \Lambda(\theta) + \theta \cdot x_0.
\]

For a given $x_0 \in \mathbb{R}^d$, $x_0 \neq \mathbb{E}X_1$, suppose that the solution $\theta^*$ to the equation $\Lambda'(\theta) = x_0$ exists and $\theta^* \in \Theta^0$. Then the expansion of the integral in (1) is available. For example, the following result is well known (its proof can be found in, e.g. [8], [10, pp. 533–534], and [14, pp. 17–21]).

**Proposition 2.** Suppose that $\Lambda''(\theta^*)$ is strictly positive definite and that $|\varphi|^{\gamma}$ is integrable for some $\gamma \geq 1$. Then
\[
\int_{v \in \mathbb{R}^d} \psi(n^{-1/2}A(\theta^*)v, \theta^*, n) \times \phi(v) \, dv = 1 + o\left(\frac{1}{\sqrt{n}}\right).
\]

### 3.1. Monte Carlo estimation

The integral in (1) may be estimated via Monte Carlo simulation. In particular, this integral may be re-expressed as
\[
\int_{v \in \mathbb{R}^d} \psi(n^{-1/2}A(\theta^*)v, \theta^*, n) \phi(v) \, dv,
\]
where $g$ is a density supported on $\mathbb{R}^d$. Now if $V_1, V_2, \ldots, V_N$ are i.i.d. with distribution given by the density $g$, then
\[
f_n(\tilde{x}) := \left(\frac{n}{2\pi}\right)^{d/2} \frac{\exp[n(A(\theta^*) - \theta^* \cdot x_0)]}{\sqrt{\det(A''(\theta^*))}} \frac{1}{N} \sum_{i=1}^{N} \psi(n^{-1/2}A(\theta^*)V_i, \theta^*, n)\phi(V_i)
\]
is an unbiased estimator for $f_n(x_0)$.

#### 3.1.1. Approximating the zero-variance estimator

Note that to get a zero-variance estimator for the above integral we need
\[
g(v) \propto \psi(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi(v).
\]

We now argue that
\[
\psi(n^{-1/2}A(\theta^*)v, \theta^*, n) \sim 1
\]
for all $v = o(n^{1/6})$. We may then select an importance sampling (IS) density $g$ that is asymptotically similar to $\phi$ for $v = o(n^{1/6})$. In the further tails, we allow $g$ to have fatter power law tails. This ensures that large values of $V$ in the simulation do not contribute substantially to the variance. Further analysis is needed to see (2). Note from the definition of $\eta(v, \theta)$ that
\[
\eta(0, \theta) = 0, \quad \eta''(0, \theta) = 0, \quad \text{and} \quad \eta'''(v, \theta) = (i)^3 \Lambda'''(\theta + iv)
\]
for all $\theta$, while
\[
\eta(0, \theta^*) = 0
\]
for the saddle point $\theta^*$. Here $\eta'$, $\eta''$, and $\eta'''$ are the first, second, and third derivatives of $\eta$ with respect to $v$, with $\theta$ held fixed. Note that while $\eta'$ and $\eta''$ are a $d$-dimensional vector and a $d \times d$ matrix, respectively, $\eta'''(v, \theta)$ is the array of numbers $(\frac{\partial^3 \eta(v, \theta)}{\partial v_i \partial v_j \partial v_k})_{1 \leq i,j,k \leq d}$.

The following notation aids in dealing with such quantities. If $A = (a_{i,j,k})_{1 \leq i,j,k \leq d}$ is a $d \times d \times d$ array of numbers, $u = (u_1, u_2, \ldots, u_d)$ is a $d$-dimensional vector, and $B$ is a $d \times d$ matrix,
then we use the notation \( A \odot u = \sum_{1 \leq i, j, k} a_{ijk} u_i u_j u_k \) and \( A \ast B = (c_{ijk})_{1 \leq i, j, k} \), where \( c_{ijk} = \sum_{m, n, p} a_{mnp} b_{mi} b_{nj} b_{pk} \). It then follows that \( A \odot (Bu) = (A \ast B) \odot u \). Since it follows from the three-term Taylor series expansion, (3), and (4) that \( \psi(n^{-1/2}A(\theta^*)v, \theta^*, n) \) equals

\[
\exp[n\eta(n^{-1/2}A(\theta^*)v, \theta^*)] = \exp\left[\frac{1}{6\sqrt{n}} \Lambda'''(\theta^* + in^{-1/2}A(\theta^*)\tilde{v}) \odot (iA(\theta^*)v)\right],
\]

where \( \tilde{v} \) is between \( v \) and the origin. Hence, continuity of \( \Lambda''' \) in the neighborhood of \( \theta^* \) implies (2).

3.1.2. Proposed IS density. We now define the form of the IS density \( g \). We first show its parametric structure and then specify the parameters that achieve asymptotically vanishing relative error.

For \( a \in (0, \infty) \), \( b \in (0, \infty) \), and \( \alpha \in (1, \infty) \), set

\[
g(v) = \begin{cases} 
    b \times \phi(v) & \text{when } |v| < a, \\
    \frac{C}{|v|^\alpha} & \text{when } |v| \geq a.
\end{cases}
\] (5)

Note that if we put

\[
p := \int_{|v| < a} g(v) \, dv = b \int_{|v| < a} \phi(v) \, dv = b \times IG\left(\frac{d}{2}, \frac{a^2}{2}\right),
\]

where \( IG(\omega, x) = (1/\Gamma(\omega)) \int_0^x e^{-r^{\omega-1}} \, dr \) is the incomplete gamma integral (or the gamma distribution function, see, e.g. [14, p. 38]), then \( C = (1 - p)/\int_{|v| \geq a} |v|^{-\alpha} \, dv > 0 \), provided \( p < 1 \).

The following assumption is important for coming up with the parameters of the proposed IS density.

**Assumption 1.** There exist \( \alpha_0 > 1 \) and \( \gamma \geq 1 \) such that

\[
\int_{u \in \mathbb{R}^d} |u|^{\alpha_0} |\varphi(u)|^\gamma \, du < \infty.
\]

By the Riemann–Lebesgue lemma, if the probability distribution of \( X_1 \) is given by a density function then \( |\varphi(u)| \to 0 \) as \( |u| \to \infty \). Assumption 1 is easily seen to hold when \( |\varphi(u)| \) decays as a power law as \( |u| \to \infty \). This is true, for example, for gamma-distributed random variables. More generally, this holds when the underlying density has integrable higher derivatives (see [10, p. 514]): if the \( k \)-th order derivative of the underlying density is integrable then, for any \( \alpha_0 \), Assumption 1 holds with \( \gamma > (1 + \alpha_0)/k \).

To specify the parameters of the IS density, we need further analysis. Define

\[
\phi_\theta(u) := \mathbb{E}_\theta (e^{iu(X_1 - x_0)}) = e^{-iu \cdot x_0} \frac{M(\theta + iu)}{M(\theta)},
\]

where \( \mathbb{E}_\theta \) denotes the expectation operator under the distribution \( F_\theta \). Let

\[
h(x) := 1 - \sup_{|u| \leq x} |\phi_\theta(u)|^2.
\]

Then \( 0 \leq h(x) \leq 1 \), \( h(0) = 0 \), \( h(x) \) is continuous and nondecreasing, and \( h(x) \uparrow 1 \) as \( x \downarrow 0 \). Furthermore, since \( \psi \) is the characteristic function of a nonlattice distribution, \( h(x) > 0 \) if \( x > 0 \). We define

\[
h_1(y) = \min\{z \mid h(z) \geq y\} \quad \text{for } y \in (0, 1).
\]

Then, for any \( y \in (0, 1) \), we have \( h(h_1(y)) \geq y \) and \( h_1(z) \downarrow 0 \) as \( z \downarrow 0 \).
Let \( \{s_n\}_{n=1}^{\infty} \) be any sequence such that, as \( n \to \infty \), \( s_n \downarrow 0 \); for any positive \( \beta \), \((1 - s_n)^{n \beta} \to 0 \); and \( \sqrt{n} h_n(s_n) \to \infty \). Taking \( s_n \) to be of order \( n^{-\varepsilon} \) for \( \varepsilon \in (0, 1) \) satisfies these three properties (see [8] for this and for further discussion on how \( s_n \) may be selected in practice). Set \( \delta_3(n) : = h_1(s_n) \). Then it follows that if \( x \geq \delta_3(n) \) then \( h(x) \geq s_n \). Equivalently, \( |\psi_{n}(u)| \leq \sqrt{1 - s_n} \) for all \( u \geq \delta_3(n) \).

Let \( \kappa_{\min} \) and \( \kappa_{\max} \) denote the minimum and maximum eigenvalues of \( \Lambda''(\theta^*) \), respectively. Hence, \( 1/\kappa_{\min} \) is the maximum eigenvalue of \( \Lambda''(\theta^*)^{-1} = A(\theta^*) A(\theta^*) \). Therefore, we have

\[
\frac{1}{\kappa_{\min}} = \|A(\theta^*)\|^2.
\]

Next, put \( \delta_2(n) = \sqrt{\kappa_{\max}} \delta_3(n) \). Then \( \sqrt{n} \delta_2(n) \to \infty \) and \( |v| \geq \delta_2(n) \) implies that \( |A(\theta^*)v| \geq \delta_2(n) \). Also, let \( \delta_1(n) = \delta_2(n)/\sqrt{\kappa_{\max}} = \delta_3(n)/\sqrt{\kappa_{\max}/\kappa_{\min}} \), so that \( |v| < \delta_2(n) \) implies that \( |A(\theta^*)v| < \delta_1(n) \).

Now we are in position to specify the parameters for the proposed IS density. Let \( \alpha = \alpha_0 \) and \( \kappa_{\min} = \sqrt{n} \delta_2(n) \). Let \( p_n = b_n \times IG(d/2, a_n^2/2) \). For \( g \) to be a valid density function, we need \( p_n < 1 \). Since \( IG(d/2, a_n^2/2) \to 1 \), select \( b_n \) to be a sequence of positive real numbers that converge to 1 in such a way that \( b_n < 1/IG(d/2, a_n^2/2) \) and

\[
\lim_{n \to \infty} (1 - s_n)^{n(d + \alpha)/2} \times IG(d/2, a_n^2/2) = 0.
\]

For example, \( b_n = 1 - n^{-\xi} \) for any \( \xi > 0 \) satisfies (6). For each \( n \), let \( g_n \) denote the PDF of the form (5) with parameters \( \alpha, \kappa_{\min} \), and \( b_n \) chosen as above. Let \( \mathbb{E}_n \) and \( \text{var}_n \) denote the expectation and variance, respectively, with respect to the density \( g_n \).

**Theorem 1.** Suppose that Assumption 1 holds and that \( \theta^* \in \Theta_0 \). Then

\[
\mathbb{E}_n \left( \frac{\psi^2(n^{-1/2} A(\theta^*) V, \theta^*, n) \phi^2(V)}{g_n(V)} \right) = \int_{v \in \mathbb{R}^{d}} \frac{\psi^2(n^{-1/2} A(\theta^*) v, \theta^*, n) \phi^2(v)}{g_n(v)} dv = 1 + o(n^{-1/2}).
\]

Consequently, from Proposition 2, it follows that

\[
\text{var}_n \left( \frac{\psi(n^{-1/2} A(\theta^*) V_i, \theta^*, n) \phi(V_i)}{g_n(V_i)} \right) \to 0 \quad \text{as} \quad n \to \infty,
\]

so that the proposed estimators for \( (f_n(x_0) : n \geq 1) \) have an asymptotically vanishing relative error.

We will use the following result from Equation (2.8) of [10, pp. 534].

**Lemma 1.** For any \( \lambda, \beta \in \mathbb{C} \),

\[
| \exp(\lambda) - 1 - \beta | \leq \left( |\lambda - \beta| + \frac{|\beta|^2}{2} \right) \exp(\omega) \quad \text{for all} \quad \omega \geq \max\{|\lambda|, |\beta|\}.
\]

Also, note that, from the definitions of \( \psi \) and \( \eta_i \), it follows that, for any \( \theta \in \Theta \),

\[
\exp \left[ - \frac{v \cdot v}{2} \right] \psi(n^{-1/2} A(\theta) v, \theta, n)
\]
is a characteristic function. To see this, observe that
\[
\exp\left(-\frac{1}{2} v \cdot v \right) \psi(n^{-1/2} A(\theta) v) = \left[ \mathbb{E}_0(\exp(-n^{-1/2} A(\theta) v \cdot (X_1 - x_0))) \right]^n
\]
\[
= [\psi(n^{-1/2} A(\theta) v)]^n.
\]

Some more observations are useful for proving Theorem 1. Since \( \eta''' \) is continuous, it follows, from the three-term Taylor series expansion,
\[
\eta(v, \theta) = \eta(0, \theta) + \frac{1}{2!} \eta'(0, \theta) v + \frac{1}{3!} \eta''(0, \theta) v + \frac{1}{6} \eta'''(\tilde{v}, \theta) \odot v
\]
(where \( \tilde{v} \) is between \( v \) and the origin), (3), and (4), that there exists a sequence \( \{\epsilon_n\} \) of positive numbers converging to 0 so that
\[
\left| \eta(v, \theta^*) - \frac{1}{3!} \eta'''(0, \theta^*) \odot v \right| \leq \epsilon_n (\kappa_{\min})^{3/2} |v|^3 \quad \text{for} \quad |v| < \delta_1(n),
\]

or, equivalently,
\[
\left| \eta(v, \theta^*) - \frac{1}{3!} \Lambda'''(\theta^*) \odot (iv) \right| \leq \epsilon_n (\kappa_{\min})^{3/2} |v|^3 \quad \text{for} \quad |v| < \delta_1(n).
\]

Furthermore, for sufficiently large \( n \),
\[
\left| \frac{1}{3!} \Lambda'''(\theta^*) \odot (iv) \right| < \frac{1}{8} \kappa_{\min} |v|^2
\]
and
\[
|\eta(v, \theta^*)| < \frac{1}{8} \kappa_{\min} |v|^2
\]
for all \( |v| < \delta_1(n) \). We will assume that \( n \) is sufficiently large so that (8) and (9) hold in the remaining analysis.

**Proof of Theorem 1.** We write
\[
\int_{v \in \mathbb{R}^d} \frac{\psi^2(n^{-1/2} A(\theta^*) v) \phi^2(v)}{g_n(v)} \, dv = I_3 + I_4,
\]
where
\[
I_3 = \int_{|v| < \sqrt{\delta_2(n)}} \frac{\psi^2(n^{-1/2} A(\theta^*) v) \phi^2(v)}{g_n(v)} \, dv
\]
and
\[
I_4 = \int_{|v| \geq \sqrt{\delta_2(n)}} \frac{\psi^2(n^{-1/2} A(\theta^*) v) \phi^2(v)}{g_n(v)} \, dv.
\]
From (5) we see that
\[
I_3 = \frac{1}{b_n} \int_{|v| < \sqrt{\delta_2(n)}} |v|^2 \psi^2(n^{-1/2} A(\theta^*) v) \phi(v) \, dv
\]
and
\[
I_4 = \frac{1}{C_n} \int_{|v| \geq \sqrt{\delta_2(n)}} |v|^2 \psi^2(n^{-1/2} A(\theta^*) v) \phi(v) \, dv.
\]
For any $c > 0$, put $\Phi_d(c) := \int_{|v| < c} \phi(v) \, dv (= IG(d/2, c^2/2))$. By the triangle inequality,

$$|I_3 - 1| \leq \left| I_3 - \frac{\Phi_d(\sqrt{n} \delta_2(n))}{b_n} \right| + \left| \frac{\Phi_d(\sqrt{n} \delta_2(n))}{b_n} - 1 \right|.$$ 

Since, as $n \to \infty$, we have $\Phi_d(\sqrt{n} \delta_2(n)) \to 1$ and $b_n \to 1$, the second term on the right-hand side converges to 0. Writing $\xi_3(\theta^*) = \Lambda''(\theta^*) \ast A(\theta^*)$, for the first term, we have

$$\left| I_3 - \frac{\Phi_d(\sqrt{n} \delta_2(n))}{b_n} \right| = \frac{1}{b_n} \int_{|v| < \sqrt{n} \delta_2(n)} \left[ \psi^2(n^{-1/2} A(\theta^*) v, \theta^*, n) - 1 \right] \phi(v) \, dv \leq \frac{1}{b_n} \int_{|v| < \sqrt{n} \delta_2(n)} \left[ \psi^2(n^{-1/2} A(\theta^*) v, \theta^*, n) - 1 - \frac{\xi_3(\theta^*)}{3\sqrt{n}} \right] \phi(v) \, dv \leq \frac{1}{b_n} \frac{1}{(2\pi)^{d/2}} \int_{|v| < \sqrt{n} \delta_2(n)} \left| \psi^2(n^{-1/2} A(\theta^*) v, \theta^*, n) - 1 - \frac{\xi_3(\theta^*)}{3\sqrt{n}} \right| e^{-v^2/2} \, dv.$$

We apply Lemma 1 with $\lambda = 2n|\eta(n^{-1/2} A(\theta^*) v, \theta^*)|$ and $\beta = \frac{1}{3} \Lambda''(\theta^*) \ast (in^{-1/2} A(\theta^*) v)$. Since $|\beta|^2/2 = n^{-1} P(v)$, where $P$ is a homogeneous polynomial whose coefficients do not depend on $n$, and $|v| < \sqrt{n} \delta_2(n)$ implies that $|n^{-1/2} A(\theta^*) v| < \delta_1(n)$, we obtain, from (9), (8), and (7), respectively,

$$|\lambda| = 2n|\eta(n^{-1/2} A(\theta^*) v, \theta^*)| < 2n \frac{1}{3} \kappa_{\min} |n^{-1/2} A(\theta^*) v|^2 \leq \frac{1}{3} \kappa_{\min} \| A(\theta^*) \|^2 |v|^2 = \frac{|v|^2}{4},$$

$$|\beta| = 2n \left| \frac{1}{3} \Lambda''(\theta^*) \ast (in^{-1/2} A(\theta^*) v) \right| < 2n \frac{1}{3} \kappa_{\min} |n^{-1/2} A(\theta^*) v|^2 \leq \frac{1}{3} \kappa_{\min} \| A(\theta^*) \|^2 |v|^2 = \frac{|v|^2}{4},$$

and $|\lambda - \beta|$ satisfies

$$2n \left| \eta(n^{-1/2} A(\theta^*) v, \theta^*) - \frac{1}{3} \Lambda''(\theta^*) \ast (in^{-1/2} A(\theta^*) v) \right| < 2n \kappa_{\min} \frac{1}{3} \| n^{-1/2} A(\theta^*) v \|^3 
\leq \frac{2\kappa_{\min} |v|^3}{\sqrt{n}}.$$ 

From Lemma 1, it now follows that the integrand in the last integral is dominated by

$$\exp \left[ -\frac{|v|^2}{2} \right] \left( \frac{2\kappa_{\min} |v|^3}{\sqrt{n}} + \frac{1}{n} P(v) \right) \exp \left[ -\frac{|v|^2}{2} \right] = \exp \left[ -\frac{|v|^2}{4} \right] \left( \frac{2\kappa_{\min} |v|^3}{\sqrt{n}} + \frac{1}{n} P(v) \right).$$
Therefore, we have $I_3 = 1 + o(n^{-1/2})$. Also,

$$|I_4| \leq \frac{1}{(2\pi)^d C_n} \int_{|v| > \sqrt{A} \delta_2(n)} |v|^d |\exp[-|u|^2]|^{(n-1/2)A(\theta^*)v, \theta^*, n)}| \, dv$$

$$= \frac{1}{(2\pi)^d C_n} \int_{|v| > \sqrt{A} \delta_2(n)} |v|^d |\psi_{\theta^*}(n^{-1/2}A(\theta^*)v)|^{2n} \, dv$$

$$\leq (1 - s_n)^{-\gamma/2} \int_{\mathbb{R}} |u|^d |\psi_{\theta^*}(n^{-1/2}A(\theta^*)u)|^{\gamma'} \, du$$

$$= (1 - s_n)^{-\gamma/2} C_n \int_{\mathbb{R}} |u|^d |\psi_{\theta^*}(u)|^{\gamma'} \, du$$

$$\leq D_1 (1 - s_n)^{-\gamma/2} C_n \int_{\mathbb{R}} |u|^d |\psi_{\theta^*}(u)|^{\gamma'} \, du$$

$$\leq D_1 (1 - s_n)^{-\gamma/2} C_n \int_{|v| > \sqrt{A} \delta_2(n)} |v|^{-\gamma} \, dv$$

$$\leq D_1 (1 - s_n)^{-\gamma/2} C_n \int_{\mathbb{R}} |u|^d |\psi_{\theta^*}(u)|^{\gamma'} \, du,$$

where $D_1$ is a constant independent of $n$. By Assumption 1, the above integral over $u$ is finite. For large $n$, we also have $\int_{|v| > \sqrt{A} \delta_2(n)} |v|^{-\gamma} \, dv \leq \int_{|v| > 1} |v|^{-\gamma} \, dv$. By our choice of $b_n$ we can conclude that $I_4 \to 0$ as $n \to \infty$, completing the proof.

### 4. Efficient estimation of tail probability

In this section we consider the problem of efficient estimation of $P(\bar{X}_n \in A)$ for sets $A$ that are affine transformations of the nonnegative orthants $\mathbb{R}_+^d$ along with some minor variations. As in [3, pp. 83–89], the dominating point of the set $A$ plays a crucial role in our analysis. As is well known, a point $x_0$ is called a dominating point of $A$ if $x_0$ uniquely satisfies the following properties: (i) $x_0$ is in the boundary of $A$; (ii) there exists a unique $x^* \in \mathbb{R}^d$ with $A(x^*) = x_0$; and (iii) $A \subseteq \{x \mid x^* \cdot (x - x_0) \geq 0\}$. In the remainder of the paper we assume the existence of a dominating point $x_0$ for $A$.

Our estimation relies on a saddle-point representation of $P(\bar{X}_n \in A)$ obtained using Parseval’s relation. Let $Y_n := \sqrt{n}(\bar{X}_n - x_0)$ and $A_{n,x_0} := \sqrt{n}(A - x_0)$, where $x_0 = (x_0^1, x_0^2, \ldots, x_0^d)$ is an arbitrarily chosen point in $\mathbb{R}^d$. Let $h_{n,\theta^*, x_0}(y)$ be the density function of $Y_n$ when each $X_i$ has distribution function $F_\theta$ obtained by exponentially twisting $F$ by $\theta$. That is,

$$dF_\theta(x) = e^{\theta^* \cdot x} M(\theta)^{-1} dF(x) = e^{\theta^* \cdot x - \Lambda(\theta)} dF(x).$$

An exact expression for the tail probability is given by

$$P(\bar{X}_n \in A) = P(Y_n \in A_{n,x_0}) = e^{-n\theta^* \cdot x_0 - \Lambda(\theta^*)} \int_{y \in A_{n,x_0}} e^{-\sqrt{n}\Lambda(\theta^*)} h_{n,\theta^*, x_0}(y) \, dy, \quad (10)$$

where recall that $\theta^* \in \Theta^{\theta}$ is a solution to $\Lambda'(\theta) = x_0$, and $x_0$ is the dominating point of $A$.

Define

$$c(n, \theta^*, x_0) = \int_{y \in A_{n,x_0}} e^{-\sqrt{n}\Lambda(\theta^*)} \, dy = n^{d/2} \int_{w \in (A - x_0)} e^{-n\theta^* \cdot w} \, dw.$$

We need the following assumption.

**Assumption 2.** For all $n$, $c(n, \theta^*, x_0) < \infty$. 

Downloaded from https://www.cambridge.org/core. IP address: 54.70.40.11, on 26 May 2019 at 05:15:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1239/jap/1378401231
Since $x_0$ is a dominating point of $A$, for any $y \in A_{n,x_0}$, we have $\theta^* \cdot y \geq 0$. Hence, if $A$ is a set with finite Lebesgue measure then $c(n, \theta^*, x_0)$ is finite. Assumption 2 may hold even when $A$ has infinite Lebesgue measure, as Example 1 below illustrates.

When Assumption 2 holds, we can rewrite the right-hand side of (10) as

$$c(n, \theta^*, x_0)e^{-n(\theta^* \cdot x_0 - \Lambda(\theta^*))} \int_{y \in A_{n,x_0}} r_{n,\theta^*,x_0}(y) \rho_{n,\theta^*,x_0}(y) \, dy,$$

where $r_{n,\theta^*,x_0}(y)$ is a density function that equals $e^{-\sqrt{n}(\theta^* \cdot y)}/c(n, \theta^*, x_0)$ for $y \in A_{n,x_0}$ and 0 otherwise.

Let $\rho_{n,\theta^*,x_0}(t)$ denote the complex conjugate of the characteristic function of $r_{n,\theta^*,x_0}(y)$. Since the characteristic function of $h(n, \theta^*, x_0)$ equals $e^{-\sqrt{n}t}M(\theta^* + it/\sqrt{n})M(\theta^*)^{-1}$, by Parseval’s relation, (11) is equal to

$$c(n, \theta^*, x_0)e^{-n(\theta^* \cdot x_0 - \Lambda(\theta^*))} \left(\frac{1}{2\pi}\right)^d \int_{t \in \mathbb{R}^d} \rho_{n,\theta^*,x_0}(t)e^{-it\sqrt{n}\varphi(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi(v)} \, dt.$$

This in turn, by the change of variable $t = A(\theta^*)v$ and rearrangement of terms, equals

$$\frac{c(n, \theta^*, x_0)e^{-n(\theta^* \cdot x_0 - \Lambda(\theta^*))}}{\sqrt{\det(\Lambda''(\theta^*))}} \left(\frac{1}{2\pi}\right)^{d/2} \times \int_{v \in \mathbb{R}^d} \rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi(v) \, dv. \tag{12}$$

We need another assumption to facilitate the analysis.

**Assumption 3.** For all $t \in \mathbb{R}^d$, $\lim_{n \to \infty} \rho_{n,\theta^*,x_0}(t) = 1$.

**Proposition 3.** Suppose that $A$ has a dominating point $x_0$, that the associated $\theta^* \in \Theta^0$, and that $\Lambda''(\theta^*)$ is strictly positive definite. Furthermore, suppose that Assumptions 2 and 3 hold. Then

$$\mathbb{P}(\bar{X}_n \in A) \sim \left(\frac{1}{2\pi}\right)^{d/2} \frac{c(n, \theta^*, x_0)e^{-n(\theta^* \cdot x_0 - \Lambda(\theta^*))}}{\sqrt{\det(\Lambda''(\theta^*))}},$$

or, equivalently, by (12),

$$\lim_{n \to \infty} \int_{v \in \mathbb{R}^d} \rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi(v) \, dv = 1.$$

The proof of Proposition 3 is similar to that of Proposition 2 and is thus omitted (see [8]). Let $g$ be any density supported on $\mathbb{R}^d$. If $V_1, V_2, \ldots, V_N$ are i.i.d. with distribution given by density $g$, then an unbiased estimator for $\mathbb{P}(\bar{X}_n \in A)$ is given by

$$\hat{P}(\bar{X}_n \in A) = \left(\frac{1}{2\pi}\right)^{d/2} \frac{c(n, \theta^*, x_0)e^{-n(\theta^* \cdot x_0 - \Lambda(\theta^*))}}{\sqrt{\det(\Lambda''(\theta^*))}} \times \frac{1}{N} \sum_{j=1}^N \rho_{n,\theta^*,x_0}(A(\theta^*)V_j)\psi(n^{-1/2}A(\theta^*)V_j, \theta^*, n)\phi(V_j). \tag{13}$$

Note that, for the above estimator to be useful, we must be able to find closed-form expressions for $c(n, \theta^*, x_0)$ and $\rho_{n,\theta^*,x_0}(t)$ or these should be cheaply computable. In Section 4.1 we consider some examples where we explicitly compute $c(n, \theta^*, x_0)$ and $\rho_{n,\theta^*,x_0}$ and verify Assumptions 2 and 3.
Theorem 2. Under Assumptions 1, 2, and 3,
\[ \mathbb{E}_n \left( \frac{\rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi^2(v)}{g_n^2(v)} \right) = 1 + o(n^{-1/2}) \quad \text{as } n \to \infty, \]
where \( g_n \) is same as Theorem 1. Consequently, by Proposition 3, it follows that, as \( n \to \infty \), \( \text{var}(\hat{\theta} \mid \mathcal{X}_n \in \mathcal{A}) \to 0 \) and the proposed estimator has asymptotically vanishing relative error.

Proof. The proof follows along the same lines as the proof of Theorem 1. We write
\[ \int_{v \in \mathbb{R}^d} \frac{\rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi^2(v)}{g_n(v)} \, dv = I_5 + I_6, \]
where
\[ I_5 = \int_{|v| < \delta_2(n)\sqrt{n}} \frac{\rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi^2(v)}{g_n(v)} \, dv \]
\[ = \frac{1}{b_n} \int_{|v| < \delta_2(n)\sqrt{n}} \rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi(v) \, dv, \]
\[ I_6 = \int_{|v| \geq \delta_2(n)\sqrt{n}} \rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi^2(v) \, dv, \]
Now
\[ |I_5 - 1| \leq \frac{1}{b_n} \int_{|v| < \delta_2(n)\sqrt{n}} \rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi(v) \, dv - 1 \]
\[ \leq \frac{1}{b_n} \int_{|v| < \delta_2(n)\sqrt{n}} \rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-1/2}A(\theta^*)v, \theta^*, n) - 1 \phi(v) \, dv \]
\[ \leq \frac{1}{b_n} \int_{|v| < \delta_2(n)\sqrt{n}} \rho_{n,\theta^*,x_0}(A(\theta^*)v)\left\{ \psi^2(n^{-1/2}A(\theta^*)v, \theta^*, n) - 1 - \frac{\zeta_1(\theta^*)}{3\sqrt{n}} \right\} \]
\[ \times \phi(v) \, dv + o(1) \]
\[ \leq \frac{1}{b_n} \int_{|v| < \delta_2(n)\sqrt{n}} \frac{\rho_{n,\theta^*,x_0}(A(\theta^*)v)}{v} \phi^2(n^{-1/2}A(\theta^*)v, \theta^*, n) - 1 - \frac{\zeta_1(\theta^*)}{3\sqrt{n}} \phi(v) \, dv + o(1) \]
Now, as in the case of Theorem 1, we conclude that \( I_5 = 1 + o(n^{-1/2}) \). Also, since
\[ |I_6| \leq \frac{1}{c_n} \int_{|v| \geq \delta_2(n)\sqrt{n}} [v]^{\frac{1}{2}} \rho_{n,\theta^*,x_0}(A(\theta^*)v) \psi^2(n^{-1/2}A(\theta^*)v, \theta^*, n)\phi^2(v) \, dv \]
\[ \leq \frac{1}{2\pi^d c_n} \int_{|v| \geq \delta_2(n)\sqrt{n}} [v]^{\frac{1}{2}} \exp(-v^2) \psi^2(n^{-1/2}A(\theta^*)v, \theta^*, n) \, dv, \]
we conclude that \( I_6 \to 0 \) as \( n \to \infty \), completing the proof.
4.1. Examples

Example 1. Let \( A = x_0 + \mathbb{R}_+^d \), where \( x_0 = (x_0^1, x_0^2, \ldots, x_0^d) \) is a given point in \( \mathbb{R}^d \). Furthermore, suppose that, for all \( i = 1, 2, \ldots, d, \theta_i^* > 0 \). It is easy to see that existence of such a \( \theta^* \) implies that \( x_0 \) is a dominating point for \( A \). It also follows that Assumption 2 holds and

\[
c(n, \theta^*, x_0) = 1/n^{d/2} \theta_1^* \theta_2^* \cdots \theta_d^*.
\]

It can easily be verified that

\[
\rho_{n, \theta^*, x_0}(t_1, t_2, \ldots, t_d) = \prod_{i=1}^d \left( 1 + \frac{1}{1 + n t_i / \sqrt{n \theta_i^*}} \right).
\]

Therefore, Assumption 3 also holds in this case. By Proposition 3, we then have

\[
\mathbb{P}(\bar{X}_n - x_0 \in \mathbb{R}_+^d) \sim \frac{e^{n(\Lambda(\theta^*) - \theta^* x_0)}}{(2\pi)^{d/2} n^{d/2} \sqrt{\det(A(\theta^*))} \theta_1^* \theta_2^* \cdots \theta_d^*}.
\]

Example 2. When \( A = x_0 + B \mathbb{R}_+^d \) and \( B \) is a nonsingular matrix, the problem can also be reduced to that considered in Example 1 by a simple change of variable. Set \( y = B^{-1} x \). Then it follows that, for any \( \theta^* \),

\[
c(n, \theta, x_0) = \det(B) \int_{\mathbb{R}_+^d} \exp[-\sqrt{n}(B^\top \theta \cdot z)] \, dz.
\]

Now if we assume that all the \( d \) components of \( B^\top \theta^* \) are positive then, as in Example 1, both Assumptions 2 and 3 hold.

For \( 0 \leq d' < d \), let

\[
Q_{d'}^+ := \{(x_1, x_2, \ldots, x_d) \in \mathbb{R}_+^d \mid x_i \geq 0 \text{ for all } 0 \leq i \leq d'\}.
\]

Similar analysis holds when \( A = x_0 + B Q_{d'}^+ \) and \( B \) is a nonsingular matrix. Then, a simple change of variable \( y = B^{-1} x \) reduces the problem to a lower dimension problem as in Example 1 with \( d \) replaced by \( d' \).

Example 3. In Examples 1 and 2 we considered sets \( A \) which are unbounded. In this example we show that a similar analysis holds when the set \( A \) is bounded. Consider the three increasing regions \( A_i : i = 1, 2, 3 \), where \( A_3 \) corresponds to region \( A \) considered in Example 1, \( A^{(1)} \) is the \( d \)-dimensional rectangle given by \( \prod_{i=1}^d [x_0^i, x_0^i + D_i] \), and \( A^{(2)} \) is such that \( A^{(1)} \subset A^{(2)} \subset A^{(3)} \). Then \( x_0 \) is the common dominating point for all the three sets. Again, suppose that, for all \( i = 1, 2, \ldots, d, \theta_i^* > 0 \). Suppressing the dependence on \( x_0 \) and \( \theta^* \), for \( i = 1, 2 \), let

\[
c_n^{(i)}(y) := \int_{y \in \mathbb{R}(A^{(i)} - x_0)} e^{-\sqrt{n}(\theta^* \cdot y)} \, dy
\]

and

\[
\rho_n^{(i)}(t) := \frac{1}{c_n^{(i)}} \int_{y \in \mathbb{R}(A^{(i)} - x_0)} e^{-t \cdot y - \sqrt{n}(\theta^* \cdot y)} \, dy.
\]

Then

\[
c_n^{(1)} = \frac{(1 - e^{-n \theta_1^* D_1})(1 - e^{-n \theta_2^* D_2}) \cdots (1 - e^{-n \theta_d^* D_d})}{n^{d/2} \theta_1^* \theta_2^* \cdots \theta_d^*}
\]
and
\[ \rho_n^{(1)}(t_1, t_2, \ldots, t_d) = \prod_{i=1}^{d} \left( \frac{1}{1 + \frac{it_i}{\sqrt{n} \theta_i}} \frac{1 - e^{-\theta_i^* D_i (1 + it_i/\sqrt{n} \theta_i)}}{1 - e^{-\theta_i^* D_i}} \right). \]

Therefore, it follows that Assumption 3 holds for \( A^{(1)} \). Also, note that
\[ |\rho_n^{(2)}(t) - 1| \leq \frac{1}{\epsilon_n} \int_{y \in \sqrt{n}(A^{(2)} - x_0)} e^{-\sqrt{n}(\theta^* \cdot y)} |e^{-it \cdot y} - 1| \, dy \]
\[ \leq \frac{1}{n^{d/2} \epsilon_n} \int_{z \in \sqrt{n}(A^{(2)} - x_0)} e^{-\theta^* \cdot z} |e^{-it \cdot z/\sqrt{n}} - 1| \, dz \]
\[ \leq \frac{1}{n^{d/2} \epsilon_n} \int_{z \in \mathbb{R}^d} e^{-\theta^* \cdot z} |e^{-it \cdot z/\sqrt{n}} - 1| \, dz. \]

Since the last integral converges to 0, it follows that Assumption 3 holds for \( A^{(2)} \). A similar analysis carries over if these sets are transformed using a nonsingular matrix \( B \) under the conditions in Example 2.

In Example 1 we assumed that, for all \( i = 1, 2, \ldots, d, \theta_i^* > 0 \). In many settings, this may not be true, but the problem can be easily transformed to be amenable to the proposed algorithms. This is discussed further in [8].

4.2. Estimating the expected overshoot

The methodology developed previously to estimate the tail probability \( F(X_n) \) can be extended to estimate \( E(X_n^\alpha) \) for \( \alpha \in (\mathbb{Z}_+ - \{0\})^d \). We illustrate this in a single-dimension setting (\( d = 1 \)) for \( \alpha = 1 \), and \( A = (x_0, \infty) \) for \( x_n > E X_n \).

Let \( S_n = \sum_{i=1}^{n} X_i \). In finance and in insurance one is often interested in estimating \( E((S_n - n x_0) | S_n > n x_0) \), which is known as the expected overshoot or the peak over threshold. As we have an efficient estimator for \( P(X_n > x_0) \), the problem of efficiently estimating \( E(S_n | S_n > n x_0) \) is equivalent to that of efficiently estimating \( E((S_n - n x_0) 1_{(S_n > n x_0)}) \). Note that \( E((S_n - n x_0) 1_{(S_n > n x_0)}) = \sqrt{n} E(Y_n 1_{(Y_n > 0)}) \), where \( Y_n = \sqrt{n}(X_n - x_0) \). Using (10), we get
\[ E(Y_n 1_{(Y_n > 0)}) = e^{-n(\theta^* \cdot x_0 - \Lambda(\theta^*))} \int_{0}^{\infty} y e^{-\sqrt{n}(\theta^* \cdot y)} h_{n, \theta^*}(y) \, dy, \] \( (14) \)
where recall that \( \theta^* \in \Theta \) is a solution to \( \Lambda'(\theta) = x_0 \) and \( h_{n, \theta^*}(y) \) is the density of \( Y_n \) when each \( X_i \) has distribution \( F_{\theta^*} \). Define
\[ \tilde{c}(n, \theta^*) = \int_{0}^{\infty} y e^{-\sqrt{n}(\theta^* \cdot y)} \, dy = (n\theta^*)^{-\frac{1}{2}}. \]

Hence, for all \( n, \tilde{c}(n, \theta^*) < \infty \). The right-hand side of (14) may be re-expressed as
\[ \tilde{c}(n, \theta^*) e^{-n(\theta^* \cdot x_0 - \Lambda(\theta^*))} \int_{0}^{\infty} \tilde{\rho}_{n, \theta^*}(y) h_{n, \theta^*}(y) \, dy, \] \( (15) \)
where the density function \( \tilde{\rho}_{n, \theta^*}(y) = y e^{-\sqrt{n}(\theta^* \cdot y)/\tilde{c}(n, \theta^*)} \) for \( y > 0 \), and 0 otherwise. Let \( \tilde{\rho}_{n, \theta^*}(t) \) denote the complex conjugate of the characteristic function of \( \tilde{\rho}_{n, \theta^*}(y) \). By simple calculations, it follows that \( \tilde{\rho}_{n, \theta^*}(t) = 1/1 - t^2/n\theta^* - 2it/\sqrt{n}\theta^* \) and \( \lim_{n \to \infty} \tilde{\rho}_{n, \theta^*}(t) = 1 \).
Then, repeating the analysis for the tail probability, analogously to (12), we see that (15) equals
\[
\tilde{c}(n, \theta^*) e^{-n(\theta^* x_0 - \Lambda(\theta^*))} \int_0^\infty \tilde{\rho}_n, \theta^*(A(\theta^*) v) \psi(n^{-1/2} A(\theta^*) v, \theta^*, n) \phi(v) \, dv.
\]

As in Proposition 3, we can see that
\[
E((S_n - nx_0) \mathbb{I}_{\{S_n > nx_0\}}) \sim \left( \frac{n}{2\pi} \right)^{1/2} \tilde{c}(n, \theta^*) e^{-n(\theta^* x_0 - \Lambda(\theta^*))} \sqrt{\det(\Lambda''(\theta^*))} = \left( \frac{n}{2\pi} \right)^{1/2} \frac{1}{\theta^*} e^{-n(\theta^* x_0 - \Lambda(\theta^*))} \theta^* \sqrt{\det(\Lambda''(\theta^*))},
\]
so that
\[
E((S_n - nx_0) \mathbb{I}_{\{S_n > nx_0\}}) / P(S_n > nx_0) \sim \frac{1}{\theta^*}.
\]

Using analysis identical to that in Theorem 2, it follows that the resulting unbiased estimator of
\[
E((S_n - nx_0) \mathbb{I}_{\{S_n > nx_0\}}) (\text{when density } g_n \text{ is used}) \text{ has an asymptotically vanishing relative error.}
\]

5. Numerical experiments

5.1. Estimation of the probability density function of \( \tilde{X}_n \)

We first use the proposed method to estimate the PDF of \( \tilde{X}_n \) for the case where the sequence of random variables \((X_i : i \geq 1)\) is independent and identically exponentially distributed with mean 1. Then the sum has a known gamma density function facilitating comparison of the estimated value to the true value. The density function estimates using the proposed method (referred to as the SP-IS method) are evaluated for \( n = 30, a_n = 2, \alpha = 2, \) and \( p_n = 0.9 \) (the algorithm performance was observed to be relatively insensitive to small perturbations in these values; see [8] for a discussion on how these parameters may be selected) based on \( N \) generated samples. In Table 1 we present a comparison of our method with the conditional Monte Carlo (CMC) method proposed in Asmussen and Glynn (2008, pp. 145–146) for estimating the density function of \( \tilde{X}_n \) at a few values. As discussed in Asmussen and Glynn (2008), the CMC estimates are given by an average of \( N \) independent samples of \( nf(x - S_{n-1}) \), where \( S_{n-1} \) is generated by sampling \((X_1, \ldots, X_{n-1})\) using their original density function \( f \). In Figure 1 we present a graphical comparison over a wider range of density function values. As may be expected, the proposed method provides an estimator with much smaller variance compared to the CMC method.

<table>
<thead>
<tr>
<th>( x )</th>
<th>True value</th>
<th>SP-IS method</th>
<th>CMC method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Sample variance</td>
<td>Estimate</td>
</tr>
<tr>
<td>1.0</td>
<td>2.179</td>
<td>2.185</td>
<td>0.431</td>
</tr>
<tr>
<td>0.985</td>
<td>0.087</td>
<td>4.946 \times 10^{-4}</td>
<td>9.105 \times 10^{-4}</td>
</tr>
<tr>
<td>1.094 \times 10^{-4}</td>
<td>1.105 \times 10^{-4}</td>
<td>1.066 \times 10^{-9}</td>
<td>7.342 \times 10^{-7}</td>
</tr>
</tbody>
</table>
5.2. Comparison with state-dependent exponential twisting

We compare the efficiency of the SP-IS method for estimating the tail probability $\mathbb{P}(\bar{X}_n \in A)$ with the optimal state-dependent exponential twisting method proposed by Blanchet et al. [2] (referred to as the BGL method). They restricted their analysis to convex sets $A$ with twice continuously differentiable boundary, whereas the SP-IS method is applicable to sets that are affine transformations of the nonnegative orthants $\mathbb{R}^d_+$. The two methods agree in the single dimension, and so we compare them in a single-dimension example (see [8] for a numerical comparison of the SP-IS method with the one proposed by Sadowsky and Bucklew [19] in the multidimension setting).

For a sequence of random variables $(X_i: i \geq 1)$ that are independent and identically exponentially distributed with mean $1$, $\mathbb{P}(\bar{X}_n \geq 1.5)$ is estimated for different values of $n$. In Table 2 we present the estimates based on different $N$ generated samples. In this experiment, $\alpha_n = 2$, $\alpha = 2$, and $p_n = 0.9$ for the SP-IS method. The BGL method is implemented as per [2] as follows. First $X_1$ is generated using an exponentially twisted distribution with mean $x_0 = 1.5$. At each next step, the exponential twisting coefficient in the distribution used to generate $X_{k+1}$ is recomputed such that the mean of the distribution is $(n x_0 - \sum_{i=1}^{k} X_i)/(n - k)$. The exponential twisting is dynamically updated until the generated $\sum_{i=1}^{k} X_i \geq n x_0$ at which point we stop the importance sampling and sample the remaining $n - k$ values with the original distribution. In the other case, if the distance to the boundary $n x_0 - \sum_{i=1}^{k} X_i$ is sufficiently large relative to the remaining time horizon $n - k ((n x_0 - \sum_{i=1}^{k} X_i)/(n - k) \geq 2 x_0)$, then we generate the next $n - k$ samples with exponentially twisted distribution with mean $(n x_0 - \sum_{i=1}^{k} X_i)/(n - k)$. In this example, the true value of the tail probability for different values of $n$ is calculated using an approximation of the gamma density function available.
Table 2: The SP-IS method has a decreasing coefficient of variation (CoV) and it provides an increasing variance reduction (VR) over the BGL method. The computation time per sample (CT), reported in micro seconds, increases with $n$ for the BGL method, whereas it remains constant for the SP-IS method.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$N$</th>
<th>True value (exact asymptotic, $c_n$)</th>
<th>BGL CoV</th>
<th>SP-IS CoV</th>
<th>VR BGL</th>
<th>VR SP-IS</th>
<th>CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>$10^3$</td>
<td>$9.276 \times 10^{-4}$</td>
<td>1.41</td>
<td>$9.055 \times 10^{-4}$</td>
<td>0.32</td>
<td>20.38</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td>$10^4$</td>
<td>$9.039 \times 10^{-4}$</td>
<td>1.41</td>
<td>$9.036 \times 10^{-4}$</td>
<td>0.32</td>
<td>19.77</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td>$10^5$</td>
<td>$(9.992 \times 10^{-4})$</td>
<td>1.41</td>
<td>$9.038 \times 10^{-4}$</td>
<td>0.32</td>
<td>19.13</td>
<td>7.5</td>
</tr>
<tr>
<td>100</td>
<td>$10^3$</td>
<td>$5.936 \times 10^{-6}$</td>
<td>1.44</td>
<td>$5.932 \times 10^{-6}$</td>
<td>0.28</td>
<td>25.84</td>
<td>15.4</td>
</tr>
<tr>
<td></td>
<td>$10^4$</td>
<td>$5.924 \times 10^{-6}$</td>
<td>1.45</td>
<td>$5.923 \times 10^{-6}$</td>
<td>0.29</td>
<td>24.54</td>
<td>15.4</td>
</tr>
<tr>
<td></td>
<td>$10^5$</td>
<td>$(6.261 \times 10^{-6})$</td>
<td>1.44</td>
<td>$5.921 \times 10^{-6}$</td>
<td>0.29</td>
<td>24.20</td>
<td>15.4</td>
</tr>
<tr>
<td>200</td>
<td>$10^3$</td>
<td>$3.355 \times 10^{-10}$</td>
<td>1.48</td>
<td>$3.378 \times 10^{-10}$</td>
<td>0.28</td>
<td>25.83</td>
<td>32.0</td>
</tr>
<tr>
<td></td>
<td>$10^4$</td>
<td>$3.371 \times 10^{-10}$</td>
<td>1.46</td>
<td>$3.368 \times 10^{-10}$</td>
<td>0.28</td>
<td>26.17</td>
<td>32.0</td>
</tr>
<tr>
<td></td>
<td>$10^5$</td>
<td>$(3.473 \times 10^{-10})$</td>
<td>1.46</td>
<td>$3.374 \times 10^{-10}$</td>
<td>0.28</td>
<td>26.92</td>
<td>32.0</td>
</tr>
</tbody>
</table>

in MATLAB®. Variance reduction achieved by the SP-IS method over the BGL method is reported. This increases with increasing $n$. In addition, we note that the computation time per sample for the BGL method increases with $n$, whereas it remains constant for the SP-IS method. The results in Table 2 show that the exact asymptotic $c_n$ can differ significantly from the estimated value of the probability. As shown in [8], this difference can be far more significant in multi-dimension settings, thus emphasizing the need for simulation, despite the existence of asymptotics for the rare quantities considered.

6. Conclusions and direction for further research

In this paper we considered the rare event problem of efficient estimation of the density function of the average of i.i.d. light-tailed random vectors evaluated away from their mean, and the tail probability that this average takes a large deviation. In a single-dimension setting we also considered the estimation problem of the expected overshoot associated with a sum of i.i.d. random variables taking large deviations. We used the well-known saddle-point representations for these performance measures and applied importance sampling to develop provably efficient unbiased estimation algorithms that significantly improve upon the performance of the existing algorithms in the literature and are simple to implement.

Our key contribution was combining rare event simulation with the classical theory of saddle-point-based approximations for tail events. We hope that this approach spurs research towards efficient estimation of a much richer class of rare event problems where saddle-point approximations are well known or are easily developed.

Another direction that is important for further research involves relaxing Assumption 2 or 3 in our analysis. Then our important sampling estimators may not have asymptotically vanishing relative error, but may have bounded relative error. This is illustrated through an example in [8].

References


