Applied Probability Trust (23 November 2012)

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

ANKUSH AGARWAL, SANTANU DEY AND SANDEEP JUNEJA,*

*

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, light-tailed and non-lattice 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 queuing and financial credit risk modeling. It has been extensively studied in 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. Further, 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 asymptotically vanishing relative error estimator for the practically important *expected overshoot* of sums of iid random variables.

* Postal address: STCS, Tata Institute of Fundamental Research, 1 Homi Bhabha Road, Mumbai, India. Email: {dsantanu, ankush} @tcs.tifr.res.in; juneja@tifr.res.in

Keywords: Rare Event Simulation; Importance Sampling;Saddle Point Approximation; Fourier inversion; Large Deviations

2000 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 (iid) light tailed (their moment generating function is finite in a neighborhood of zero) non-lattice (modulus of their characteristic function is strictly less than one) random vectors taking values in $\Re^d$, for $d \geq 1$. In this paper† we consider the problem of efficient simulation estimation of the probability density function of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ at points away from EX_i, and the tail probability $P(\bar{X}_n \in \mathcal{A})$ for sets $\mathcal{A}$ that do not contain EX_i and essentially are affine transformations of the non-negative orthant of $\Re^d$. 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 $\bar{X}_n$ obtained from Fourier inversion of the characteristic function of X_1 (see e.g., [2], [6] and [17]). Furthermore, using Parseval's relation, similar representations for $P(\bar{X}_n \in \mathcal{A})$ are easily developed. To illustrate the broader applicability of the proposed methodology, we also develop similar representation for $E(\bar{X}_n : \bar{X}_n \geq a)$‡ in a single dimension setting ($d = 1$), for $a > EX_i$, and using it 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 practical viewpoint its 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 $P(\bar{X}_n \in \mathcal{A})$ via importance sampling, besides being of independent importance, may

†A very preliminary version of this paper appeared as [8].

‡Authors thank the editor for suggesting this application

also be considered a building block for more complex problems involving many streams of i.i.d. random variables (see e.g., [19], for a queuing application; [13] for applications in credit risk modeling). This problem has been extensively studied in rare event simulation literature (see e.g., [3], [10], [12], [14], [20], [21]). Essentially, the literature exploits the fact that the zero variance importance sampling estimator for $P(\bar{X}_n \in \mathcal{A})$, though unimplementable, has a Markovian representation. This representation may be exploited to come up with provably efficient, implementable approximations (see [1] and [15]).

Sadowsky and Bucklew in [21] (also see [5]) developed exponential twisting based importance sampling algorithms to arrive at unbiased estimators for $P(\bar{X}_n \in \mathcal{A})$ that they proved were asymptotically or weakly efficient (as per the current standard terminology in 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 by [21] were state independent in that each X_{k+1} was generated from a distribution independent of the previously generated $(X_i : i \leq k)$. Blanchet, Leder and Glynn in [3] also considered the problem of estimating $P(\bar{X}_n \in \mathcal{A})$ 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 $E\left[(\bar{X}_n - a) : \bar{X}_n \geq a\right]$ 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 article we exploit the saddle point based representations of the rare event quantities considered. These representations allow us to write the quantity of interest α_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 estimation of β_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 β_n exist, however they require computation of third and higher order derivatives of the log-moment generating function of X_i. This is particularly difficult in higher dimensions. In addition, it is difficult to control the bias in such approximations. As we note later in numerical experiments, these biases can be significant even when 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., [13]).

In our analysis, we note that the integral β_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 α_n has the desirable asymptotically vanishing relative error property. More tangibly, the estimator of the integral β_n has the property that its variance converges to zero as $n \to \infty$. An additional advantage of the proposed approach over existing methodologies for estimating $P(\bar{X}_n \in \mathcal{A})$ 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., [2], [16] and [17]). 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 $P(\bar{X}_n \in \mathcal{A})$. Then, in Section 3, we develop an importance sampling estimator for the density of $\bar{X}_n$ and show that it has asymptotically vanishing relative error. In Section 4, we devise an integral representation for $P(\bar{X}_n \in \mathcal{A})$ 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 asymptotically vanishing relative error estimator for $E(\bar{X}_n : \bar{X}_n \geq a)$ 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 [9], a more elaborate version of this paper.

2. Rare event simulation, a brief review

Let $\alpha_n = E_n Y_n = \int Y_n dP_n$ be a sequence of rare event expectations in the sense that $\alpha_n \to 0$ as $n \to \infty$, for non-negative random variables $(Y_n : n \geq 1)$. Here, E_n is the expectation operator under P_n. For example, when $\alpha_n = P(B_n)$, Y_n corresponds to the indicator of the event B_n.

Naive simulation for estimating α_n requires generating many iid samples of Y_n under P_n. Their average then provides an unbiased estimator of α_n. Central limit theorem based approximations then provide an asymptotically valid confidence interval for α_n (under the assumption that $E_n Y_n^2 < \infty$). Importance sampling involves expressing $\alpha_n = \int Y_n L_n d\tilde{P}_n = \tilde{E}_n[Y_n L_n]$, where $\tilde{P}_n$ is another probability measure such that P_n is absolutely continuous w.r.t. $\tilde{P}_n$, with $L_n = \frac{dP_n}{d\tilde{P}_n}$ denoting the associated Radon-Nikodym derivative, or the likelihood ratio, and $\tilde{E}_n$ is the expectation operator under $\tilde{P}_n$. The importance sampling unbiased estimator $\hat{\alpha}_n$ of α_n is obtained by taking an average of generated iid samples of $Y_n L_n$ under $\tilde{P}_n$. Note that by setting $d\tilde{P}_n = \frac{Y_n}{E_n(Y_n)} dP_n$ the simulation output $Y_n L_n$ is $E_n(Y_n)$ almost surely, signifying that such a $\tilde{P}_n$ provides a zero variance estimator for α_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 its mean. Therefore, the latter is a good measure of efficiency of the estimator. Note that under naive simulation, when $Y_n = I(B_n)$ (For any set D, $I(D)$ denotes its indicator), the standard deviation of each sample of simulation output equals $\sqrt{\alpha_n(1-\alpha_n)}$ so that when divided by α_n, the ratio increases to infinity as $\alpha_n \to 0$. Below we list some criteria that are popular in evaluating the efficacy of the proposed importance sampling estimator (see [1]). Here, $Var(\hat{\alpha}_n)$ denotes the variance of the estimator $\hat{\alpha}_n$ under the appropriate importance sampling measure.

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 $\limsup_{n\to\infty} \frac{\sqrt{Var(\hat{\alpha}_n)}}{\alpha_n^{1-\epsilon}} < \infty$ for all $\epsilon > 0$;

to be *strongly efficient* or have *bounded relative error* if $\limsup_{n\to\infty} \frac{\sqrt{Var(\hat{\alpha}_n)}}{\alpha_n} < \infty$; and to have *asymptotically vanishing relative error* if $\lim_{n\to\infty} \frac{\sqrt{Var(\hat{\alpha}_n)}}{\alpha_n} = 0$.

3. Efficient estimation of probability density function of $\bar{X}_n$

In this section we first develop a saddle point based representation for the probability density function (pdf) of $\bar{X}_n$ in Proposition 3.1 (for proof see e.g., [2], [6], [17] and [9]). We then develop an approximation to the zero variance estimator for this pdf. Our main result is Theorem 3.1, where we prove that the proposed estimator has an asymptotically vanishing relative error.

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

$$M(\theta) := E\left[e^{\theta \cdot X_1}\right] = E[e^{\theta_1 X_1^1 + \theta_2 X_1^2 + \cdots + \theta_d X_1^d}],$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_d)$ and for $x, y \in \Re^d$ 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

$$\varphi(\theta) := E\left[e^{\iota\theta \cdot X_1}\right] = E[e^{\iota(\theta_1 X_1^1 + \theta_2 X_1^2 + \cdots + \theta_d X_1^d)}]$$

where $\iota = \sqrt{-1}$. In this paper we assume that the distribution of X_i is non-lattice, which means that $|\varphi(\theta)| < 1$ for all $\theta \in \Re^d - \{0\}$.

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

$$\Theta := \left\{\theta = (\theta_1, \theta_2, \ldots, \theta_d) \in \Re^d | \Lambda(\theta) < \infty\right\}.$$

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

Denote the Euclidean norm of $x \in \Re^d$ by $|x| := \sqrt{x \cdot x}$. For a square matrix A, $\det(A)$ denotes the determinant of A, while norm of A is denoted by $||A|| := \max_{|x|=1} |Ax|$. Let $\Lambda''(\theta)$ 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 $\Lambda''(\theta)$.

Proposition 3.1. *Suppose $\Lambda''(\theta)$ is strictly positive definite for some $\theta \in \Theta^0$. Furthermore, suppose that $|\varphi|^\gamma$ is integrable for some $\gamma \geq 1$. Then f_n, the density function of $\bar{X}_n$, exists for all $n \geq \gamma$ and its value at any point x_0 is given by:*

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

where $\psi(y, \theta, n) = \exp\left[n \times \eta(y, \theta)\right]$ and

$$\eta(y, \theta) = \frac{1}{2} y^t \Lambda''(\theta) y + \Lambda\left(\theta + \iota y\right) - (\theta + \iota y) \cdot x_0 - \Lambda(\theta) + \theta \cdot x_0. \tag{2}$$

For a given $x_0 \in \Re^d, x_0 \neq EX_1$, suppose that the solution θ^* 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 is well-known (proof can be found, e.g., in [17], [11], [9]):

Proposition 3.2. *Suppose $\Lambda''(\theta^*)$ is strictly positive definite and $|\varphi|^\gamma$ is integrable for some $\gamma \geq 1$. Then,*

$$\int_{v \in \Re^d} \psi(n^{-\frac{1}{2}} A(\theta^*)v, \theta^*, n) \times \phi(v)\, dv = 1 + o\left(\frac{1}{\sqrt{n}}\right). \tag{3}$$

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

$$\int_{v \in \Re^d} \psi(n^{-\frac{1}{2}} A(\theta^*)v, \theta^*, n) \frac{\phi(v)}{g(v)} g(v)\, dv\,,$$

where g is a density supported on $\Re^d$. Now if $V_1, V_2, \ldots, V_N$ are iid with distribution given by the density g, then

$$\hat{f}_n(\bar{x}) := \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \frac{\exp\left[n\left\{\Lambda(\theta^*) - \theta^* \cdot x_0\right\}\right]}{\sqrt{\det(\Lambda''(\theta^*))}} \frac{1}{N} \sum_{i=1}^{N} \frac{\psi(n^{-\frac{1}{2}} A(\theta^*)V_i, \theta^*, n)\phi(V_i)}{g(V_i)} \tag{4}$$

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

$$g(v) \propto \psi(n^{-\frac{1}{2}} A(\theta^*)v, \theta^*, n)\phi(v)\,.$$

We now argue that

$$\psi(n^{-\frac{1}{2}} A(\theta^*)v, \theta^*, n) \sim 1 \tag{5}$$

for all $v = o(n^{\frac{1}{6}})$. We may then select an IS density g that is asymptotically similar to ϕ for $v = o(n^{\frac{1}{6}})$. 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 (5). Note from the definition of $\eta(v,\theta)$, that

$$\eta(0,\theta) = 0, \quad \eta''(0,\theta) = 0 \quad \text{and} \quad \eta'''(v,\theta) = (\iota)^3 \Lambda'''(\theta + \iota v) \tag{6}$$

for all θ, while

$$\eta'(0,\theta^*) = 0 \tag{7}$$

for the saddle point θ^*. Here η' , η'' and η''' are the first, second and third derivatives of η w.r.t. v, with θ held fixed. Note that while η' and η'' are d-dimensional vector and $d \times d$ matrix respectively, $\eta'''(v,\theta)$ is the array of numbers: $((\frac{\partial^3 \eta}{\partial v_i \partial v_j \partial v_k}(v,\theta)))_{1 \le i,j,k \le d}$.

The following notation aids in dealing with such quantities: If $A = (a_{ijk})_{1 \le i,j,k \le d}$ is a $d \times d \times d$ array of numbers and $u = (u_1, u_2, \ldots, u_d)$ is a d-dimensional vector and B is a $d \times d$ matrix then we use the notation $A \odot u = \sum_{1 \le i,j,k \le d} a_{ijk} u_i u_j u_k$ and $A \star B = (c_{ijk})_{1 \le i,j,k \le d}$, where $c_{ijk} = \sum_{m,n,p} a_{mnp} b_{mi} b_{nj} b_{pk}$. It then follows that $A \odot (Bu) = (A \star B) \odot u$. Since, it follows from the three term Taylor series expansion and (6) and (7) above, that $\psi(n^{-\frac{1}{2}} A(\theta^*) v, \theta^*, n)$ equals

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

continuity of Λ''' in the neighborhood of θ^* implies (5).

3.1.2. *Proposed importance sampling 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

$$g(v) = \begin{cases} b \times \phi(v) & \text{when } |v| < a \\ \frac{C}{|v|^\alpha} & \text{when } |v| \ge a\,. \end{cases} \tag{8}$$

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 $IG(\omega, x) = \frac{1}{\Gamma(\omega)} \int_0^x e^{-t} t^{\omega-1}\, dt$ is the incomplete Gamma integral (or the Gamma distribution function, see e.g, [17]), then $C = \frac{(1-p)}{\int_{|v| \ge a} \frac{dv}{|v|^\alpha}} > 0$, 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*

$$\int_{u \in \Re^d} |u|^{\alpha_0} \, |\varphi(u)|^{\gamma} \, du < \infty \, .$$

By 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 [11]): If k-th order derivative of the underlying density is integrable then for any α_0, Assumption 1 holds with $\gamma > \frac{1+\alpha_0}{k}$.

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

$$\varphi_\theta(u) := E_\theta \left[e^{\iota u \cdot (X_1 - x_0)} \right] = e^{-\iota u \cdot x_0} \frac{M(\theta + \iota u)}{M(\theta)} \ ,$$

where E_θ denotes the expectation operator under the distribution F_θ. Let

$$h(x) := 1 - \sup_{|u| \geq x} |\varphi_{\theta^*}(u)|^2. \tag{9}$$

Then $0 \leq h(x) \leq 1$, $h(0) = 0$, $h(x)$ is continuous, non-decreasing and $h(x) \uparrow 1$ as $x \downarrow 0$. Further, since φ is the characteristic function of a non-lattice distribution, $h(x) > 0$ if $x > 0$. We define

$$h_1(y) = \min\{z \,|\, h(z) \geq y\} \text{ for } y \in (0,1).$$

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 β positive, $(1-s_n)^n n^\beta \to 0$; and $\sqrt{n} h_1(s_n) \to \infty$. Taking s_n to be order $n^{-\epsilon}$ for $\epsilon \in (0,1)$ satisfies these three properties (see [9] 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, $|\varphi_{\theta^*}(u)| < \sqrt{1-s_n}$ for all $u \geq \delta_3(n)$.

Let κ_{min} and κ_{max} denote the minimum and maximum eigenvalue of $\Lambda''(\theta^*)$, respectively. Hence $\frac{1}{\kappa_{min}}$ 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 $|A(\theta^*)v| \geq \delta_3(n)$. Also let $\delta_1(n) = \frac{1}{\sqrt{\kappa_{min}}}\delta_2(n) = \sqrt{\frac{\kappa_{max}}{\kappa_{min}}}\delta_3(n)$, so that $|v| < \delta_2(n)$ implies $|A(\theta^*)v| < \delta_1(n)$.

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

$$\lim_{n\to\infty} \frac{(1-s_n)^n n^{\frac{d+\alpha}{2}}}{\left[1 - b_n \times IG\left(\frac{d}{2}, \frac{a_n^2}{2}\right)\right]} = 0. \tag{10}$$

For example, $b_n = 1 - n^{-\xi}$ for any $\xi > 0$ satisfies (10). For each n, let g_n denote the pdf of the form (8) with parameters α, a_n and b_n chosen as above. Let E_n and Var_n denote the expectation and variance, respectively, w.r.t. the density g_n.

Theorem 3.1. *Suppose Assumption 1 holds and $\theta^* \in \Theta^0$. Then,*

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

Consequently, from Proposition 3.2, it follows that

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

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

We will use the following lemma from [11].

Lemma 1. *For any λ, $\beta \in \mathbb{C}$,*

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

Also note that from the definitions of ψ and η it follows that, for any $\theta \in \Theta$,

$$\exp\left\{-\frac{v \cdot v}{2}\right\}\psi(n^{-\frac{1}{2}}A(\theta)v, \theta, n)$$

is a characteristic function. To see this, observe that $\exp\left\{-\frac{v\cdot v}{2}\right\}\psi(n^{-\frac{1}{2}}A(\theta)v, \theta, n)$ equals

$$\left[\exp\left\{-\frac{v \cdot v}{2n} + \eta\left(n^{-\frac{1}{2}}A(\theta)v, \theta\right)\right\}\right]^n = \left(E_\theta\left[e^{\iota n^{-\frac{1}{2}}A(\theta)v\cdot(X_1 - x_0)}\right]\right)^n = \left[\varphi_\theta\left(n^{-\frac{1}{2}}A(\theta)v\right)\right]^n.$$

Some more observations are useful for proving Theorem 3.1. Since η''' is continuous, it follows from the three term Taylor series expansion,

$$\eta(v,\theta) = \eta(0,\theta) + \eta'(0,\theta)v + \frac{1}{2}(v)^T\eta''(0,\theta)v + \frac{1}{6}\eta'''(\tilde{v},\theta)\odot v$$

(where $\tilde{v}$ is between v and the origin) and (6) and (7) above that there exists a sequence $\{\epsilon_n\}$ of positive numbers converging to zero so that

$$|\eta(v,\theta^*) - \frac{1}{3!}\eta'''(0,\theta^*)\odot v| \le \epsilon_n(\kappa_{min})^{\frac{3}{2}}|v|^3 \quad \text{for } |v| < \delta_1(n),$$

or equivalently

$$|\eta(v,\theta^*) - \frac{1}{3!}\Lambda'''(\theta^*)\odot(\iota v)| \le \epsilon_n(\kappa_{min})^{\frac{3}{2}}|v|^3 \quad \text{for } |v| < \delta_1(n)\,. \tag{11}$$

Furthermore, for n sufficiently large,

$$\left|\frac{1}{3!}\Lambda'''(\theta^*)\odot(\iota v)\right| < \frac{1}{8}\kappa_{min}|v|^2 \tag{12}$$

and

$$|\eta(v,\theta^*)| < \frac{1}{8}\kappa_{min}|v|^2 \tag{13}$$

for all $|v| < \delta_1(n)$. We shall assume that n is sufficiently large so that (12) and (13) hold in the remaining analysis.

Proof. (**Theorem 3.1**)

We write

$$\int_{v\in\Re^d} \frac{\psi^2(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n)\phi^2(v)}{g_n(v)}\,dv = I_3 + I_4\,,$$

where I_3 equals

$$\int_{|v|<\sqrt{n}\delta_2(n)} \frac{\psi^2(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n)\phi^2(v)}{g_n(v)}\,dv \quad\text{and } I_4 = \int_{|v|\ge\sqrt{n}\delta_2(n)} \frac{\psi^2(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n)\phi^2(v)}{g_n(v)}\,dv.$$

From (8) we see that I_3 equals

$$\frac{1}{b_n}\int_{|v|<\sqrt{n}\delta_2(n)} \psi^2(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n)\phi(v)\,dv \quad\text{and } I_4 = \frac{1}{C_n}\int_{|v|\ge\sqrt{n}\delta_2(n)} |v|^\alpha\psi^2(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n)\phi^2(v)\,dv.$$

For any $c>0$, put $\Phi_d(c) := \int_{|v|<c}\phi(v)dv\left(= IG\left(\frac{d}{2},\frac{c^2}{2}\right)\right)$. By triangle inequality

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

Since as $n \to \infty$ we have $\Phi_d\left(\sqrt{n}\delta_2(n)\right) \to 1$ and $b_n \to 1$, the second term in RHS converges to zero. Writing $\zeta_3(\theta^*) = \Lambda'''(\theta^*) \star A(\theta^*)$, for the first term we have

$$\begin{aligned}
\left| I_3 - \frac{\Phi_d\left(\sqrt{n}\delta_2(n)\right)}{b_n} \right| &= \frac{1}{b_n}\left| \int_{|v|<\sqrt{n}\delta_2(n)} \left\{ \psi^2(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n) - 1 \right\} \phi(v)\,dv \right| \\
&= \frac{1}{b_n}\left| \int_{|v|<\sqrt{n}\delta_2(n)} \left\{ \psi^2(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n) - 1 - \frac{\zeta_3(\theta^*)}{3\sqrt{n}} \odot (\iota v) \right\} \phi(v)\,dv \right| \\
&\le \frac{1}{b_n}\frac{1}{(2\pi)^{\frac{d}{2}}} \int_{|v|<\sqrt{n}\delta_2(n)} \left| \psi^2(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n) - 1 - \frac{\zeta_3(\theta^*)}{3\sqrt{n}} \odot (\iota v) \right| e^{-\frac{v^2}{2}}\,dv.
\end{aligned}$$

We apply Lemma (1) with $\lambda = 2n \times \eta\left(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*\right)$ and $\beta = n\frac{\Lambda'''(\theta^*)}{3} \odot \left(\iota n^{-\frac{1}{2}}A(\theta^*)v\right)$. Since $\frac{|\beta|^2}{2} = \frac{1}{n}P(v)$, where P is a homogeneous polynomial whose coefficients do not depend on n, and $|v| < \sqrt{n}\delta_2(n)$ implies $|n^{-\frac{1}{2}}A(\theta^*)v| < \delta_1(n)$, we have from (13), (12) and (11), respectively

$$\begin{aligned}
|\lambda| &= 2n\left|\eta\left(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*\right)\right| < 2n\frac{1}{8}\kappa_{min}|n^{-\frac{1}{2}}A(\theta^*)v|^2 \le \frac{1}{8}\kappa_{min}||A(\theta^*)||^2|v|^2 = \frac{|v|^2}{4}, \\
|\beta| &= 2n\left|\frac{1}{3!}\Lambda'''(\theta^*) \odot \left(\iota n^{-\frac{1}{2}}A(\theta^*)v\right)\right| < 2n\frac{1}{8}\kappa_{min}|n^{-\frac{1}{2}}A(\theta^*)v|^2 \le \frac{1}{8}\kappa_{min}||A(\theta^*)||^2|v|^2 = \frac{|v|^2}{4},
\end{aligned}$$

and $|\lambda - \beta|$ satisfies

$$2n\left|\eta\left(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*\right) - \frac{1}{3!}\Lambda'''(\theta^*) \odot \left(\iota n^{-\frac{1}{2}}A(\theta^*)v\right)\right| < 2n\epsilon_n(\kappa_{min})^{\frac{3}{2}}|n^{-\frac{1}{2}}A(\theta^*)v|^3 \le \frac{2\epsilon_n|v|^3}{\sqrt{n}}.$$

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

$$\exp\left\{\frac{|v|^2}{4}\right\} \times \left(\frac{2\epsilon_n|v|^3}{\sqrt{n}} + \frac{1}{n}P(v)\right)\exp\left\{-\frac{|v|^2}{2}\right\} \times = \exp\left\{-\frac{|v|^2}{4}\right\}\left(\frac{2\epsilon_n|v|^3}{\sqrt{n}} + \frac{1}{n}P(v)\right).$$

Therefore we have $I_3 = 1 + o(n^{-\frac{1}{2}})$. Also

$$\begin{aligned}
|I_4| &\le \frac{1}{(2\pi)^d C_n} \int_{|v|>\sqrt{n}\delta_2(n)} |v|^\alpha \left| \exp\left\{-|v|^2\right\} \psi^2(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n) \right| dv \\
&= \frac{1}{(2\pi)^d C_n} \int_{|v|>\sqrt{n}\delta_2(n)} |v|^\alpha \left| \varphi_{\theta^*}\left(n^{-\frac{1}{2}}A(\theta^*)v\right) \right|^{2n} dv \\
&\le \frac{(1-s_n)^{n-\frac{\gamma}{2}}}{(2\pi)^d C_n} \int_{v\in\Re} |v|^\alpha \left| \varphi_{\theta^*}\left(n^{-\frac{1}{2}}A(\theta^*)v\right) \right|^{\gamma} dv \\
&= \frac{(1-s_n)^{n-\frac{\gamma}{2}} n^{\frac{d+\alpha}{2}} \sqrt{|\Lambda''(\theta^*)|}}{(2\pi)^d C_n} \int_{u\in\Re} |A(\theta^*)^{-1}u|^\alpha \left|\varphi_{\theta^*}(u)\right|^\gamma du \\
&\le D_1 \frac{(1-s_n)^{n-\frac{\gamma}{2}} n^{\frac{d+\alpha}{2}}}{C_n} \int_{u\in\Re} |u|^\alpha \left|\varphi_{\theta^*}(u)\right|^\gamma du \\
&\le D_1 \frac{(1-s_n)^{n-\frac{\gamma}{2}} n^{\frac{d+\alpha}{2}} \int_{|v|\ge\sqrt{n}\delta_2(n)} \frac{dv}{|v|^\alpha}}{(1-p_n)} \int_{u\in\Re} |u|^\alpha \left|\varphi_{\theta^*}(u)\right|^\gamma du\,.
\end{aligned}$$

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 $\int_{|v|\geq\sqrt{n}\delta_2(n)} \frac{dv}{|v|^\alpha} \leq \int_{|v|\geq 1} \frac{dv}{|v|^\alpha}$. By choice of b_n we can conclude that $I_4 \to 0$ as $n \to \infty$, proving Theorem 3.1.$\square$

4. Efficient Estimation of Tail Probability

In this section we consider the problem of efficient estimation of $P(\bar{X}_n \in \mathcal{A})$ for sets $\mathcal{A}$ that are affine transformations of the non-negative orthants $\Re^d_+$ along with some minor variations. As in ([4]), dominating point of the set $\mathcal{A}$ plays a crucial role in our analysis. As is well known, a point x_0 is called a dominating point of $\mathcal{A}$ if x_0 uniquely satisfies the following properties: i) x_0 is in the boundary of $\mathcal{A}$; ii) there exists a unique $\theta^* \in \Re^d$ with $\Lambda'(\theta^*) = x_0$; iii) $\mathcal{A} \subseteq \{x|\theta^* \cdot (x - x_0) \geq 0\}$. In the remaining paper, we assume the existence of a dominating point x_0 for $\mathcal{A}$.

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

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

An exact expression for the tail probability is given by:

$$P[\bar{X}_n \in \mathcal{A}] = P[Y_n \in \mathcal{A}_{n,x_0}] = e^{-n\{\theta^* \cdot x_0 - \Lambda(\theta^*)\}} \int_{y \in \mathcal{A}_{n,x_0}} e^{-\sqrt{n}(\theta^* \cdot y)} h_{n,\theta^*,x_0}(y)\, dy \tag{14}$$

where recall that $\theta^* \in \Theta^0$ is a solution to $\Lambda'(\theta) = x_0$, and x_0 is the dominating point of $\mathcal{A}$. Define

$$c(n, \theta^*, x_0) = \int_{y \in \mathcal{A}_{n,x_0}} \exp\{-\sqrt{n}(\theta^* \cdot y)\}\, dy = n^{\frac{d}{2}} \int_{w \in (\mathcal{A} - x_0)} \exp\{-n(\theta^* \cdot w)\}\, dw$$

We need the following assumption:

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

Since x_0 is a dominating point of $\mathcal{A}$, for any $y \in \mathcal{A}_{n,x_0}$, we have $\theta^* \cdot y \geq 0$. Hence, if $\mathcal{A}$ is a set with finite Lebesgue measure then $c(n, \theta^*, x_0)$ is finite. Assumption 2 may hold even when $\mathcal{A}$ has infinite Lebesgue measure, as Example 1 below illustrates.

When Assumption 2 holds, we can rewrite the right hand side of (14) as

$$c(n,\theta^*,x_0)e^{-n\{\theta^*\cdot x_0-\Lambda(\theta^*)\}}\int_{y\in\mathcal{A}_{n,x_0}} r_{n,\theta^*,x_0}(y)h_{n,\theta^*,x_0}(y)\,dy \tag{15}$$

where $r_{n,\theta^*,x_0}(y)$ is a density function that equals $\frac{\exp\{-\sqrt{n}(\theta^*\cdot y)\}}{c(n,\theta^*,x_0)}$ for $y\in\mathcal{A}_{n,x_0}$ 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 $e^{-\iota t\sqrt{n}x_0}\left[\frac{M\left(\theta^*+\frac{\iota t}{\sqrt{n}}\right)}{M(\theta^*)}\right]^n$, by Parseval's relation, (15) is equal to

$$c(n,\theta^*,x_0)e^{-n\{\theta^*\cdot x_0-\Lambda(\theta^*)\}}\left(\frac{1}{2\pi}\right)^d\int_{t\in\Re^d}\rho_{n,\theta^*,x_0}(t)e^{-\iota t\sqrt{n}x_0}\left[\frac{M\left(\theta^*+\frac{\iota t}{\sqrt{n}}\right)}{M(\theta^*)}\right]^n dt. \tag{16}$$

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

$$\frac{c(n,\theta^*,x_0)e^{-n\{\theta^*\cdot x_0-\Lambda(\theta^*)\}}}{\sqrt{\det(\Lambda''(\theta^*))}}\left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\int_{v\in\Re^d}\rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n)\phi(v)\,dv. \tag{17}$$

We need another assumption to facilitate analysis:

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

Proposition 4.1. *Suppose* $\mathcal{A}$ *has a dominating point* x_0*, the associated* $\theta^*\in\Theta^o$ *and* $\Lambda''(\theta^*)$ *is strictly positive definite. Further, Assumptions 2 and 3 hold. Then,*

$$P[\bar{X}_n\in\mathcal{A}]\sim\left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\frac{c(n,\theta^*,x_0)e^{-n\{\theta^*\cdot x_0-\Lambda(\theta^*)\}}}{\sqrt{\det(\Lambda''(\theta^*))}}, \tag{18}$$

or, equivalently by (17)

$$\lim_{n\to\infty}\int_{v\in\Re^d}\rho_{n,\theta^*,x_0}(A(\theta^*)v)\psi(n^{-\frac{1}{2}}A(\theta^*)v,\theta^*,n)\phi(v)\,dv=1. \tag{19}$$

Proof of Proposition 4.1 is similar to that of Proposition 3.2 and is omitted (see [9]).

Let g be any density supported on $\Re^d$. If $V_1,V_2,\ldots,V_N$ are iid with distribution given by density g, then an unbiased estimator for $P[\bar{X}_n\in\mathcal{A}]$ is given by

$$\begin{aligned}\hat{P}[\bar{X}_n\in\mathcal{A}] \quad = \quad & \left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\frac{c(n,\theta^*,x_0)e^{-n\{\theta^*\cdot x_0-\Lambda(\theta^*)\}}}{\sqrt{\det(\Lambda''(\theta^*))}}\\ & \times\frac{1}{N}\sum_{j=1}^N\frac{\rho_{n,\theta^*,x_0}(A(\theta^*)V_j)\psi(n^{-\frac{1}{2}}A(\theta^*)V_j,\theta^*,n)\phi(V_j)}{g(V_j)}.\end{aligned} \tag{20}$$

Note that for above estimator to be useful, one must be able to find closed form expression 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 ρ_{n,θ^*,x_0} and verify Assumptions 2 and 3.

Theorem 4.1. *Under Assumptions 1, 2 and 3,*

$$E_n\left[\frac{\rho^2_{n,\theta^*,x_0}(A(\theta^*)V)\psi^2(n^{-\frac{1}{2}}A(\theta^*)V, \theta^*, n)\phi^2(V)}{g_n^2(V)}\right] = 1 + o(n^{-\frac{1}{2}}) \quad as\ n \to \infty,$$

where g_n is same as Theorem 3.1. Consequently, by Proposition 4.1, it follows that as $n \to \infty$ $Var_n\left[\hat{P}[\bar{X}_n \in \mathcal{A}]\right] \to 0$ and the proposed estimator has asymptotically vanishing relative error.

Proof. The proof follows along the same line as proof of Theorem 3.1. We write

$$\int_{v\in\Re^d} \frac{\rho^2_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n)\phi^2(v)}{g_n(v)}\, dv = I_5 + I_6$$

where

$$\begin{aligned}
I_5 &= \int_{|v|<\delta_2(n)\sqrt{n}} \frac{\rho^2_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n)\phi^2(v)}{g_n(v)}\, dv \\
&= \frac{1}{b_n}\int_{|v|<\delta_2(n)\sqrt{n}} \rho^2_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n)\phi(v)\, dv.
\end{aligned}$$

$$\begin{aligned}
I_6 &= \int_{|v|\ge\delta_2(n)\sqrt{n}} \frac{\rho^2_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n)\phi^2(v)}{g_n(v)}\, dv \\
&= \frac{1}{C_n}\int_{|v|\ge\delta_2(n)\sqrt{n}} \rho^2_{n,\theta^*,x_0}(A(\theta^*)v)|v|^\alpha\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n)\phi^2(v)\, dv.
\end{aligned}$$

Now

$$\begin{aligned}
|I_5 - 1| &= \left|\frac{1}{b_n}\int_{|v|<\delta_2(n)\sqrt{n}} \rho^2_{n,\theta^*,x_0}(A(\theta^*)v)\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n)\phi(v)\, dv - 1\right| \\
&\le \frac{1}{b_n}\left|\int_{|v|<\delta_2(n)\sqrt{n}} \rho^2_{n,\theta^*,x_0}(A(\theta^*)v)\left\{\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n) - 1\right\}\phi(v)\, dv\right| + o(1) \\
&\le \frac{1}{b_n}\left|\int_{|v|<\delta_2(n)\sqrt{n}} \rho^2_{n,\theta^*,x_0}(A(\theta^*)v)\left\{\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n) - 1 - \frac{\zeta_3(\theta^*)}{3\sqrt{n}}\odot(\iota v)\right\}\phi(v)\, dv\right| + o(1) \\
&\le \frac{1}{b_n}\int_{|v|<\delta_2(n)\sqrt{n}} |\rho_{n,\theta^*,x_0}(A(\theta^*)v)|^2\left|\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n) - 1 - \frac{\zeta_3(\theta^*)}{3\sqrt{n}}\odot(\iota v)\right|\phi(v)\, dv + o(1) \\
&\le \frac{1}{b_n}\int_{|v|<\delta_2(n)\sqrt{n}} \left|\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n) - 1 - \frac{\zeta_3(\theta^*)}{3\sqrt{n}}\odot(\iota v)\right|\phi(v)\, dv + o(1).
\end{aligned}$$

Now as in the case of Theorem 3.1 we conclude that $I_5 = 1 + o(n^{-\frac{1}{2}})$. Also, since

$$\begin{aligned} |I_6| &\leq \frac{1}{C_n} \int_{|A(\theta^*)v| \geq \delta_2(n)\sqrt{n}} |v|^\alpha \left|\rho_{n,\theta^*,x_0}(A(\theta^*)v)\right|^2 \left|\psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n)\right| \phi^2(v)\, dv \\ &\leq \frac{1}{(2\pi)^d C_n} \int_{|A(\theta^*)v| \geq \delta_2(n)\sqrt{n}} |v|^\alpha \left|\exp\left\{-v^2\right\} \psi^2(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n)\right|\, dv, \end{aligned}$$

we conclude that $I_6 \to 0$ as $n \to \infty$ proving the theorem.□

4.1. Examples

Example 1. Let $\mathcal{A} = x_0 + \Re_+^d$, where $x_0 = (x_0^1, x_0^2, \ldots, x_0^d)$ is a given point in $\Re^d$. Further suppose that $\forall i = 1, 2, \ldots, d,\ \theta_i^* > 0$. It is easy to see that existence of such a θ^* implies that x_0 is a dominating point for $\mathcal{A}$. It also follows that Assumption 2 holds and $c(n, \theta^*, x_0) = \frac{1}{n^{\frac{d}{2}}\theta_1^*\theta_2^*\cdots\theta_d^*}$. It can easily be verified that $\rho_{n,\theta^*,x_0}(t_1, t_2, \ldots t_d) = \prod_{i=1}^d \left(\frac{1}{1+\frac{\iota t_i}{\sqrt{n}\theta_i^*}}\right)$. Therefore Assumption 3 also holds in this case. By Proposition 4.1, we then have

$$P[\bar{X}_n - x_0 \in \Re_+^d] \sim \frac{e^{n\{\Lambda(\theta^*) - \theta^* \cdot x_0\}}}{(2\pi)^{\frac{d}{2}} n^{\frac{d}{2}} \sqrt{\det(\Lambda''(\theta^*))}\theta_1^*\theta_2^* \cdots \theta_d^*}.$$

By Theorem 4.1, (20) is an unbiased estimator for $P[\bar{X}_n - x_0 \in \Re_+^d]$ and has an asymptotically vanishing relative error.

Example 2. When $\mathcal{A} = x_0 + B\Re_+^d$ and B a nonsingular matrix, the problem can also be reduced to that considered in Example 1 by a simple change of variable. Set $y = B^{-1}z$. Then, it follows that for any θ

$$c(n, \theta, x_0) = \det(B) \int_{z \in \Re_+^d} \exp\{-\sqrt{n}(B^T\theta \cdot z)\}\, dz.$$

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

For $0 \leq d' < d$, let

$$Q_{d'}^+ := \{(x_1, x_2, \ldots, x_d) \in \Re^d |\ x_i \geq 0\ \forall\ 0 \leq i \leq d'\}.$$

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

Example 3. In above examples we have considered sets $\mathcal{A}$ which are unbounded. In this example we show that similar analysis holds when the set $\mathcal{A}$ is bounded. Consider the three increasing regions $(\mathcal{A}_i : i = 1, 2, 3)$, where $\mathcal{A}_3$ corresponds to region $\mathcal{A}$ considered in Example 1, $\mathcal{A}^{(1)}$ is the d-dimensional rectangle given by $\prod_i^d [x_0^i, x_0^i + D_i]$, and $\mathcal{A}^{(2)}$ is such that $\mathcal{A}^{(1)} \subset \mathcal{A}^{(2)} \subset \mathcal{A}^{(3)}$. Then, x_0 is the common dominating point for all the three sets. Again suppose that $\forall i = 1, 2, \ldots, d,\ \theta_i^* > 0$. Suppressing dependence on x_0 and θ^*, for $i = 1, 2$, let

$$c_n^{(i)} := \int_{y \in \sqrt{n}(\mathcal{A}^{(i)} - x_0)} \exp\{-\sqrt{n}(\theta^* \cdot y)\}\, dy$$

and

$$\rho_n^{(i)}(t) := \frac{1}{c_n^{(i)}} \int_{y \in \sqrt{n}(\mathcal{A}^{(i)} - x_0)} \exp\{-\iota t \cdot y - \sqrt{n}(\theta^* \cdot y)\}\, dy.$$

Then

$$c_n^{(1)} = \frac{(1 - e^{-n\theta_1^* D_1})(1 - e^{-n\theta_2^* D_2}) \cdots (1 - e^{-n\theta_d^* D_d})}{n^{\frac{d}{2}} \theta_1^* \theta_2^* \cdots \theta_d^*}$$

and

$$\rho_n^{(1)}(t_1, t_2, \ldots t_d) = \prod_{i=1}^{d} \left(\frac{1}{1 + \frac{\iota t_i}{\sqrt{n}\theta_i^*}} \times \frac{1 - e^{-n\theta_i^* D_i (1 + \frac{\iota t_i}{\sqrt{n}\theta_i^*})}}{1 - e^{-n\theta_i^* D_i}} \right).$$

Therefore, it follows that Assumption 3 holds for $\mathcal{A}^{(1)}$. Also note that,

$$\begin{aligned}
|\rho_n^{(2)}(t) - 1| &\le \frac{1}{c_n^{(2)}} \int_{y \in \sqrt{n}(\mathcal{A}^{(2)} - x_0)} \exp\{-\sqrt{n}(\theta^* \cdot y)\} \left| e^{-\iota t \cdot y} - 1 \right| \, dy \\
&\le \frac{1}{n^{\frac{d}{2}} c_n^{(1)}} \int_{z \in n(\mathcal{A}^{(2)} - x_0)} \exp\{-\theta^* \cdot z\} \left| e^{-\frac{\iota t \cdot z}{\sqrt{n}}} - 1 \right| \, dz \\
&\le \frac{1}{n^{\frac{d}{2}} c_n^{(1)}} \int_{z \in \Re_+^d} \exp\{-\theta^* \cdot z\} \left| e^{-\frac{\iota t \cdot z}{\sqrt{n}}} - 1 \right| \, dz.
\end{aligned}$$

Since the last integral converges to zero, it follows that Assumption 3 holds for $\mathcal{A}^{(2)}$. Similar analysis carries over if these sets are transformed using a non-singular matrix B under the conditions as in Example 2.

In Example 1 we assumed that $\forall i = 1, 2, \ldots, d,\ \theta_i^* > 0$. In many setting, this may not be true but the problem can be easily transformed to be amenable to the proposed algorithms. This is discussed further in [9].

4.2. Estimating expected overshoot

The methodology developed previously to estimate the tail probability $P(\bar{X}_n \in \mathcal{A})$ can be extended to estimate $E[\bar{X}_n^{\alpha}\,|\bar{X}_n \in \mathcal{A}]$ for $\alpha \in (\mathbb{Z}_+ - \{0\})^d$. We illustrate this in a single dimension setting ($d = 1$) for $\alpha = 1$, and $\mathcal{A} = (x_0, \infty)$ for $x_0 > EX_i$.

Let $S_n = \sum_{i=1}^n X_i$. In finance and in insurance one is often interested in estimating $E[(S_n - nx_0)|S_n > nx_0]$, which is known as the expected overshoot or the peak over threshold. As we have an efficient estimator for $P(\bar{X}_n > x_0)$, the problem of efficiently estimating $E[S_n|S_n > nx_0]$ is equivalent to that of efficiently estimating $E[(S_n - nx_0)I(S_n > nx_0)]$. Note that $E[((S_n - nx_0)I(S_n > nx_0)] = \sqrt{n}E[Y_n I(Y_n > 0)]$, where $Y_n = \sqrt{n}(\bar{X}_n - x_0)$. Using (14) we get

$$E[Y_n I(Y_n > 0)] = e^{-n\{\theta^*\cdot x_0 - \Lambda(\theta^*)\}} \int_0^{\infty} y\, e^{-\sqrt{n}(\theta^*\cdot y)} h_{n,\theta^*,x_0}(y)\, dy, \qquad (21)$$

where recall that $\theta^* \in \Theta$ is a solution to $\Lambda'(\theta) = x_0$ and $h_{n,\theta^*,x_0}(y)$ is the density of Y_n when each X_i has distribution F_{θ^*}. Define

$$\tilde{c}(n,\theta^*) = \int_0^{\infty} y\, \exp\{-\sqrt{n}(\theta^*\cdot y)\}\, dy = (n\,\theta^{*2})^{-1}$$

Hence, $\forall n$, $\tilde{c}(n,\theta^*) < \infty$. The right hand side of (21) may be re-expressed as

$$\tilde{c}(n,\theta^*) e^{-n\{\theta^*\cdot x_0 - \Lambda(\theta^*)\}} \int_0^{\infty} \tilde{r}_{n,\theta^*}(y) h_{n,\theta^*,x_0}(y)\, dy \qquad (22)$$

where the density function $\tilde{r}_{n,\theta^*}(y) = \frac{y\,\exp\{-\sqrt{n}(\theta^*\cdot y)\}}{\tilde{c}(n,\theta^*)}$ for $y > 0$, and zero otherwise. Let $\tilde{\rho}_{n,\theta^*}(t)$ denote the complex conjugate of the characteristic function of $\tilde{r}_{n,\theta^*}(y)$. By simple calculations, it follows that $\tilde{\rho}_{n,\theta^*}(t) = \frac{1}{1 - \frac{t^2}{n\theta^{*2}} - \frac{2\iota\, t}{\sqrt{n}\theta^*}}$ and $\lim_{n\to\infty} \tilde{\rho}_{n,\theta^*}(t) = 1$. Then, repeating the analysis for the tail probability, analogously to (17), we see that (22) equals

$$\frac{\tilde{c}(n,\theta^*) e^{-n\{\theta^*\cdot x_0 - \Lambda(\theta^*)\}}}{\sqrt{2\pi\,\Lambda''(\theta^*)}} \int_0^{\infty} \tilde{\rho}_{n,\theta^*}(A(\theta^*)v)\psi(n^{-\frac{1}{2}}A(\theta^*)v, \theta^*, n)\phi(v)\, dv.$$

As in Proposition 4.1, we can see that

$$E[(S_n - nx_0)I(S_n > nx_0)] \sim \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \frac{\tilde{c}(n,\theta^*)e^{-n\{\theta^*\cdot x_0 - \Lambda(\theta^*)\}}}{\sqrt{\det(\Lambda''(\theta^*))}} = \left(\frac{1}{2\pi n}\right)^{\frac{1}{2}} \frac{e^{-n\{\theta^*\cdot x_0 - \Lambda(\theta^*)\}}}{\theta^{*2}\sqrt{\det(\Lambda''(\theta^*))}},$$

so that $\frac{E[(S_n - nx_0)I(S_n > nx_0)]}{P[S_n > nx_0]} \sim \frac{1}{\theta^*}$.

Using analysis identical to that in Theorem 4.1, it follows that the resulting unbiased estimator of $E[(S_n - nx_0)I(S_n > nx_0)]$ (when density g_n is used) has an asymptotically vanishing relative error.

5. Numerical Experiments

5.1. Estimation of probability density function of $\bar{X}_n$

We first use the proposed method to estimate the probability density function of $\bar{X}_n$ for the case where sequence of random variables $(X_i : i \geq 1)$ are 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 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 [9] for a discussion on how these parameters may be selected) based on N generated samples. Table 1 shows the comparison of our method with the conditional Monte Carlo (CMC) method proposed in Asmussen and Glynn (2008) (pg. 145-146) for estimating the density function of $\bar{X}_n$ 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. Figure 1 shows this comparison graphically 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.

x	True value	SP-IS estimate	Sample variance	CMC estimate	Sample variance
1.0	2.179	2.185	0.431	2.360	31.387
1.5	0.085	0.087	4.946×10^{-4}	0.067	0.478
2.0	1.094×10^{-4}	1.105×10^{-4}	1.066×10^{-9}	7.342×10^{-7}	3.341×10^{-1}

Table 1: True density function and its estimates using the proposed (SP-IS) method and the conditional Monte Carlo (CMC) for an average of 30 independent exponentially distributed mean = 1 random variables. For $x = 1.0$ and 1.5, the number of generated samples $N = 1000$ in both the methods, and for $x = 2.0$, $N = 10,000$.

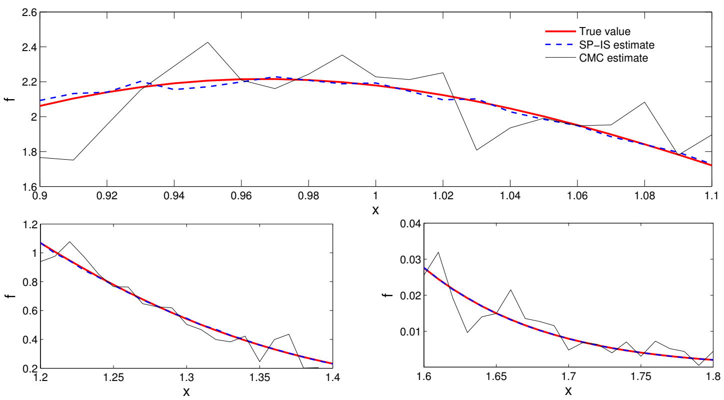

FIGURE 1: True density function and its estimates using the proposed (SP-IS) method and the conditional Monte Carlo (CMC) for an average of 30 independent exponentially distributed mean $= 1$ random variables. The plot illustrates the performance of the two methods over wide range of x values. In both simulations $N = 1,000$ at each point.

5.2. Comparison with state dependent exponential twisting

We compare the efficiency of SP-IS method for estimating the tail probability $P(\bar{X}_n \in \mathcal{A})$ with the optimal state dependent exponential twisting method proposed by [3] (referred to as BGL method). They restrict their analysis to convex sets $\mathcal{A}$ with twice continuously differentiable boundary whereas SP-IS method is applicable to sets that are affine transformations of the non-negative orthants $\Re^d_+$. The two methods agree in the single dimension and hence we compare them on a single dimension example (see [9] for a numerical comparison of the SP-IS method with the one proposed by Sadowsky and Bucklew (1990) in the multi-dimension setting).

For a sequence of random variables $(X_i : i \geq 1)$ that are independent and identically exponentially distributed with mean 1, $P(\bar{X}_n \geq 1.5)$ is estimated for different values of n. Table 2 reports the estimates based on different N generated samples. In this experiment, $a_n = 2, \alpha = 2$ and $p_n = 0.9$ for SP-IS method. BGL method is implemented as per [3] 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 mean of the distribution is $\frac{nx_0 - \sum_{i=1}^k X_i}{n-k}$. The exponential twisting is dynamically updated until the generated $\sum_{i=1}^k X_i \geq nx_0$ at which point we stop the importance sampling

n	N	True value (exact asymptotic c_n)	BGL	CoV	SP-IS	CoV	VR	CT	
								BGL	SP-IS
	10^3		9.276×10^{-4}	1.41	9.055×10^{-4}	0.32	20.38		
50	10^4	9.039×10^{-4}	9.127×10^{-4}	1.41	9.036×10^{-4}	0.32	19.77	7.5	0.9
	10^5	(9.992×10^{-4})	9.036×10^{-4}	1.41	9.038×10^{-4}	0.32	19.13		
	10^3		5.936×10^{-6}	1.44	5.932×10^{-6}	0.28	25.84		
100	10^4	5.924×10^{-6}	5.913×10^{-6}	1.45	5.923×10^{-6}	0.29	24.54	15.4	0.9
	10^5	(6.261×10^{-6})	5.928×10^{-6}	1.44	5.921×10^{-6}	0.29	24.20		
	10^3		3.355×10^{-10}	1.48	3.378×10^{-10}	0.28	25.83		
200	10^4	3.371×10^{-10}	3.381×10^{-10}	1.46	3.368×10^{-10}	0.28	26.17	32.0	0.9
	10^5	(3.473×10^{-10})	3.370×10^{-10}	1.46	3.374×10^{-10}	0.28	26.92		

Table 2: SP-IS method has a decreasing coefficient of variation (CoV) and it provides increasing variance reduction (VR) over BGL method. Computation time per sample (CT), reported in micro seconds, increases with n for BGL method whereas it remains constant for SP-IS method.

and sample rest of $n-k$ values with the original distribution. In the other case, if distance to the boundary $nx_0 - \sum_{i=1}^{k} X_i$ is sufficiently large relative to remaining time horizon $n-k$ $\left(\frac{nx_0 - \sum_{i=1}^{k} X_i}{n-k} \geq 2\,x_0\right)$, then we generate the next $n-k$ samples with exponentially twisted distribution with mean $\frac{nx_0 - \sum_{i=1}^{k} X_i}{n-k}$. In this example, the true value of tail probability for different values of n is calculated using approximation of gamma density function available in MATLAB. Variance reduction achieved by SP-IS method over BGL method is reported. This increases with increasing n. In addition, we note that the computation time per sample for BGL method increases with n whereas it remains constant for the SP-IS method. Table 2 shows that the exact asymptotic c_n can differ significantly from the estimated value of the probability. As shown in [9], 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 iid 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 expected overshoot associated with a sum of iid random variables taking a 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 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 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 Assumptions 2 or 3 in our analysis. Then, our IS estimators may not have asymptotically vanishing relative error but may have bounded relative error. This is illustrated through an example in [9].

References

[1] Asmussen, S. and Glynn, P. (2008). *Stochastic Simulation: Algorithms and Analysis.* Springer Verlag. New York, NY, USA.

[2] Butler, R. W. (2007). *Saddlepoint Approximation with Applications.* Cambridge University Press. Cambridge.

[3] Blanchet, J. Leder, D. and Glynn, P. (2008). Strongly efficient algorithms for light-tailed random walks: An old folk song sung to a faster new tune... *MCQMC 2008.Editor:Pierre LEcuyer and Art Owen* Springer 227–248.

[4] Bucklew, J. (2004). *An Introduction to Rare Event Simulation.* Springer Series in Statistics.

[5] Bucklew, J. A. Ney, P. and Sadowsky, J. S. (1990). Monte Carlo Simulation and Large Deviations Theory for Uniformly Recurrent Markov Chains *Journal of Applied Probability*, Vol. 27, No. 1 , 44-59.

[6] DANIELS, H. E. (1954). Saddlepoint Approximation in Statistics. *Annals of Mathematical Statistics.* **25** No.4, 631–650.

[7] DEMBO, A. AND ZEITOUNI, O. (1998). *Large Deviation Techniques and Applications. 2nd ed.* Springer. New York.

[8] DEY, S. AND JUNEJA, S. (2011). Efficient Estimation of Density and Probability of Large Deviations of Sum of IID Random Variables. *Proceedings of the 2011 Winter Simulation Conference.* IEEE 3800–3811.

[9] DEY, S. JUNEJA, S. AND AGARWAL, A. (2012). Efficient Simulation of Density and Probability of Large Deviations of Sum of Random Vectors using Saddle Point Representations. *Arxiv 1203.0817.*

[10] DIEKER, A. B. AND MANDJES, M. (2005). On Asymptotically Efficient Simulation of Large Deviations Probability. *Advances in Applied Probability.* **37** No.2, 539–552.

[11] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications Vol.2.* John Wiley and Sons.

[12] GLASSERMAN, P AND JUNEJA, S. (2008). Uniformly Efficient Importance Sampling for the Tail distribution of Sums of Random Variables *Mathematics of Operation Research* **33** No.1, 36–50.

[13] GLASSERMAN, P AND LI, J. (2005). Importance Sampling for Portfolio Credit Risk *Management Science* **51** 1643–1656.

[14] GLASSERMAN, P AND WANG, Y. (1997). Counterexamples in Importance Sampling for Large Deviation Probabilities *The Annals of Applied Probability* **7** No. 3, 731–746.

[15] JUNEJA, S. AND SHAHABUDDIN, P. (2006). Rare Event Simulation Techniques *Handbooks in Operation Research and Management Science* **13** Simulation.Elsevier North-Holland, Amsterdam, 291–350.

[16] LUGNNANI, R. AND RICE, S. (1980). Saddle Point Approximation for Distribution of the Sum of Independent Random Variables. *Advances in Applied Probability* **12** No.2, 475–490.

[17] JENSEN, J. L. (1995). *Saddlepoint Approximations.* Oxford University Press. Oxford.

[18] NEY, P. (1983). Dominating Points and the Asymptotics of Large Deviations for Random Walk on $\mathbb{R}^d$ *Annals of Probability* **11** No.1, 158–167.

[19] PAREKH, S. AND WALRAND, J. (1989). A Quick Simulation Method for Excessive Backlogs in Networks of Queue. *IEEE Transactions on Automatic Control.* **34** No.1, 54–66.

[20] SADOWSKY, J. S. (1996). On Monte Carlo Estimation of Large Deviation Probabilities *The Annals of Applied Probability.* **6**, 399–422.

[21] Sadowsky, J. S. and Bucklew, J. A. (1990). On Large Deviation Theory and Asymptotically Efficient Monte Carlo Simulation Estimation *IEEE Trans. Inform. Theory* **36** No.1, 579–588.