

**EFFICIENT ESTIMATION OF DENSITY AND PROBABILITY
OF LARGE DEVIATIONS OF SUM OF IID RANDOM VARIABLES**

Santanu Dey

Sandeep Juneja

School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai, 400 005, INDIA

School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai, 400 005, INDIA

ABSTRACT

We consider the problem of efficient simulation estimation of the density function at the tails, and the probability of large deviations for an average of independent, identically distributed light-tailed random variables. The latter problem 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 variables. We note that 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 approximating importance sampling measures and argue that they possess the asymptotically vanishing relative error property.

1 INTRODUCTION

Let $(X_i : i \geq 1)$ denote a sequence of independent, identically distributed (iid) light tailed random variables (their moment generating function is finite in a neighborhood of zero) taking values in \mathfrak{X} . 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 $P(\bar{X}_n \geq \bar{x})$ for $\bar{x} > EX_i$.

The problem of efficient simulation estimation of the tail probability density function has not been studied in the literature. The problem of efficiently estimating $P(\bar{X}_n \geq \bar{x})$ 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, for e.g., Parekh and Walrand (1989), for a queueing application; Glasserman and Li (2005) for applications in credit risk modeling). This problem has been extensively studied in rare event simulation literature (see, for e.g., Sadowsky and Bucklew (1990), Sadowsky (1996), Glasserman and Wang (1997), Blanchet et al. (2009), Dieker and Mandjes (2005), Glasserman and Juneja (2008)).

In this article, we build upon the well known saddle point based representations for the probability density function of \bar{X}_n (see, e.g., Jensen (1995), Butler (2007)). Furthermore, using Parseval's relation, similar representations for $P(\bar{X}_n \geq \bar{x})$. 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, $c_n/\alpha_n \rightarrow 1$ as $n \rightarrow \infty$) 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 $\beta_n \rightarrow 1$ as $n \rightarrow \infty$. In the literature, asymptotic expansions exist for β_n , 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. In Dey and Juneja (2011) we further extend our analysis to allow X_i 's to take values in a multi-dimension Euclidean space.

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 $n \rightarrow \infty$.

The organization of this paper is as follows: In Section 2 we briefly review the popular performance evaluation measures used in rare event simulation, and the existing literature on estimating $P(\bar{X}_n \geq \bar{x})$. 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 $P(\bar{X}_n \geq \bar{x})$ and develop an importance sampling estimator for it and again argue that it has asymptotically vanishing relative error. We end with a small numerical experiment in Section 5.

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 $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$, 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 .

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

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 $\alpha_n \rightarrow 0$.

Below we list some criteria that are popular in evaluating the efficacy of the proposed importance sampling estimator (see Asmussen and Glynn (2008)). 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

$$\limsup_{n \rightarrow \infty} \frac{\sqrt{Var(\hat{\alpha}_n)}}{\alpha_n^{1-\varepsilon}} < \infty$$

for all $\varepsilon > 0$;

- to be *strongly efficient*, or to have *bounded relative error* if

$$\limsup_{n \rightarrow \infty} \frac{\sqrt{Var(\hat{\alpha}_n)}}{\alpha_n} < \infty;$$

- to have asymptotically vanishing relative error if

$$\lim_{n \rightarrow \infty} \frac{\sqrt{\text{Var}(\hat{\alpha}_n)}}{\alpha_n} = 0.$$

2.2 Literature Review

Recall that $(X_i : i \geq 1)$ denote a sequence of independent, identically distributed light tailed random vectors taking values in \mathfrak{R} . Let $F(\cdot)$ denote the distribution function of X_i . Denote its moment generating function by $M(\cdot)$, so that

$$M(\theta) := E \left[e^{\theta X_1} \right].$$

The characteristic function (CF) of X_i is given by

$$\varphi(\theta) := E \left[e^{i\theta X_1} \right]$$

where $\theta \in \mathfrak{R}$ and $i = \sqrt{-1}$.

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

$$\Theta := \{ \theta \in \mathfrak{R} | \Lambda(\theta) < \infty \}.$$

Throughout this article we assume that $0 \in \Theta^\circ$, the interior of Θ , so that $(X_i : i \geq 1)$ are light tailed.

Now consider the problem of estimating $P(\bar{X}_n \geq \bar{x})$, for $\bar{x} > EX_i$. Let θ^* solve

$$\Lambda'(\theta) = \bar{x}.$$

(Here, Λ' denotes the gradient of Λ). Throughout the paper we further assume that such a $\theta^* \in \Theta^\circ$ exists.

For $\theta \in \Theta^\circ$, let

$$dF_\theta(x) = \exp(\theta x - \Lambda(\theta)) dF(x)$$

denote the exponentially twisted distribution associated with F when the twisting factor equals θ .

For estimating $P(\bar{X}_n \geq \bar{x})$, Sadowsky and Bucklew (1990) propose an importance sampling distribution under which each X_i is iid with the new distribution F_{θ^*} . Then, they prove that under this importance sampling distribution, the resulting estimator of $P(\bar{X}_n \geq \bar{x})$ is weakly efficient. See Asmussen and Glynn (2008) and Juneja and Shahabuddin (2006) for a sense in which this distribution approximates the zero variance estimator for $P(\bar{X}_n \geq \bar{x})$. Since, $\Lambda'(\theta^*) = \bar{x}$, it is easy to see that under the exponentially twisted distribution F_{θ^*} , each X_i has mean \bar{x} .

As mentioned in the introduction, Blanchet et al. (2009) consider a variant importance sampling measure where the distribution of X_j depends on the generated (X_1, \dots, X_{j-1}) . Modulo some technical conditions, they choose an exponentially twisted distribution to generate X_j so that its mean under the new distribution equals $\frac{1}{n-j+1}(n\bar{x} - \sum_{i=1}^{j-1} X_i)$. They prove that the resulting estimator is strongly efficient.

Later in Section 5, we compare the performance of the proposed algorithm to the one based on exponential twisting proposed by Sadowsky and Bucklew (1990).

3 EFFICIENT ESTIMATION OF PROBABILITY DENSITY FUNCTION OF \bar{X}_n

In this section we first develop a well known saddlepoint based representation for the probability density function (pdf) of \bar{X}_n in Proposition 1 (see, e.g., Jensen (1995), Butler (2007)). We then develop an approximation to the zero variance estimator for this pdf. Our main result is Theorem 1, which states that the proposed estimator has asymptotically vanishing relative error.

Some notation is needed in our analysis. Let $\Lambda''(\theta)$ denote the second derivative of $\Lambda(\theta)$ for $\theta \in \Theta^\circ$.

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

$$f_n(\bar{x}) = \sqrt{\frac{n}{2\pi}} \frac{\exp[n\{\Lambda(\theta) - \theta\bar{x}\}]}{\sqrt{\Lambda''(\theta)}} \int_{v \in \mathfrak{R}} \psi\left(n^{-\frac{1}{2}}\Lambda''(\theta)^{-\frac{1}{2}}v, \theta, n\right) \times \phi(v) dv, \tag{1}$$

where

$$\psi(y, \theta, n) = \exp[n \times \eta(y, \theta)]$$

and

$$\eta(y, \theta) = \frac{1}{2}\Lambda''(\theta)y^2 + \Lambda(\theta + iy) - (\theta + iy)\bar{x} - \Lambda(\theta) + \theta\bar{x}.$$

The proof of Proposition (1) follows from Cauchy’s theorem for analytic functions of complex variables. See for example Daniels (1954).

Expansion of the integral in (1) is available (see Jensen (1995)). For example, the following is well-known (recall that θ^* solves the equation $\Lambda'(\theta) = \bar{x}$):

Proposition 2 Under conditions outlined in Proposition 1,

$$\int_{v \in \mathfrak{R}} \psi\left(n^{-\frac{1}{2}}\Lambda''(\theta^*)^{-\frac{1}{2}}v, \theta^*, n\right) \times \phi(v) dv = 1 + o\left(\frac{1}{\sqrt{n}}\right).$$

A proof of Proposition 2 can be found in Jensen (1995). We also present a proof after we state Theorem 1, as this proof requires analysis that may be used with minor variations to prove Theorem 1.

3.1 Monte Carlo Estimation

We propose to estimate the integral in (1) by Monte-Carlo simulation. Specifically, we write it as

$$\int_{v \in \mathfrak{R}} \psi\left(n^{-\frac{1}{2}}\Lambda''(\theta^*)^{-\frac{1}{2}}v, \theta^*, n\right) \frac{\phi(v)}{g(v)} g(v) dv$$

where $g(v)$ is a density supported on \mathfrak{R} . Now if V_1, V_2, \dots, V_N are iid with distribution given by the density g , then the estimator given by

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

is an unbiased estimator for $f_n(\bar{x})$.

3.1.1 Approximating the Zero Variance Estimator

Note that to get a zero variance estimator for this estimator we need

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

We now argue that

$$\psi\left(n^{-\frac{1}{2}}\Lambda''(\theta^*)^{-\frac{1}{2}}v, \theta^*, n\right) \sim 1 \tag{2}$$

for all $v = o(n^{\frac{1}{6}})$. We may then select an IS 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 to ensure that large values of V in the simulation do not contribute substantially to the variance. To see (2), we need further analysis.

Note that from definition of $\eta(v, \theta)$ it easily follows that

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

for all θ , while

$$\eta'(0, \theta^*) = 0 \tag{4}$$

for a saddle point θ^* (i.e., $\Lambda'(\theta^*) = \bar{x}$). Here η' , η'' and η''' are the first, second and third derivatives of η w.r.t. the second argument v , with the first argument θ held fixed.

Since η''' is continuous, it follows from the three term Taylor series expansion,

$$\eta(v, \theta) = \eta(0, \theta) + \eta'(0, \theta)v + \frac{1}{2}\eta''(0, \theta)v^2 + \frac{1}{6}\eta'''(\tilde{v}, \theta)v^3,$$

(where $\tilde{v} \leq |v|$) (3) and (4) above, that for any given ε we can choose δ_1 small enough so that

$$|\eta(v, \theta^*) - \frac{1}{3!}\eta'''(0, \theta^*)v^3| \leq \varepsilon(\Lambda''(\theta^*))^{\frac{3}{2}}|v|^3 \quad \text{for } |v| < \delta_1.$$

Equivalently,

$$|\eta(v, \theta^*) - \frac{1}{3!}\Lambda'''(\theta^*)(\iota v)^3| \leq \varepsilon(\Lambda''(\theta^*))^{\frac{3}{2}}|v|^3 \quad \text{for } |v| < \delta_1. \tag{5}$$

We choose δ_1 so that we also have for $|v| < \delta_1$:

$$\left| \frac{1}{3!}\Lambda'''(\theta^*)(\iota v)^3 \right| < \frac{1}{8}\Lambda''(\theta^*)|v|^2, \tag{6}$$

and

$$|\eta(v, \theta^*)| < \frac{1}{8}\Lambda''(\theta^*)|v|^2. \tag{7}$$

Now, since

$$\begin{aligned} \psi\left(n^{-\frac{1}{2}}\Lambda''(\theta^*)^{-\frac{1}{2}}v, \theta^*, n\right) &= \exp\left\{n\eta\left(n^{-\frac{1}{2}}\Lambda''(\theta^*)^{-\frac{1}{2}}v, \theta^*\right)\right\} \\ &= \exp\left\{\frac{n}{6}\Lambda'''(\theta^* + \iota n^{-\frac{1}{2}}\Lambda''(\theta^*)^{-\frac{1}{2}}\tilde{v})\left(n^{-\frac{1}{2}}\Lambda''(\theta^*)^{-\frac{1}{2}}\iota v\right)^3\right\}, \end{aligned}$$

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

3.1.2 Proposed Importance Sampling Density

We now define the form of the IS density g . We first show the structure of g . Later we specify how the parameters are chosen to achieve asymptotically vanishing relative error. For $a \in (0, \infty)$, $b \in (0, \infty)$, and $\alpha \in (1, \infty)$,

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

Note that if we put

$$p := \int_{|v| < a} g(v) dv = b \int_{|v| < a} \phi(v) dv = b \times (2\Phi(a) - 1),$$

where $\Phi(\cdot)$ is the standard Normal CDF, then

$$C = \frac{(1-p)}{\int_{|v| \geq a} \frac{dv}{|v|^\alpha}} > 0$$

if $p < 1$.

Assumption 1 below is important for coming up with the proposed IS density. The following observation helps in its statement: Note, from the definitions of ψ and η , that for any $\theta \in \Theta^o$,

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

is a characteristic function. In fact, defining

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

where E_θ denotes the expectation operator under the distribution F_θ , we have

$$\begin{aligned} \exp\left\{-\frac{v^2}{2}\right\} \psi\left(n^{-\frac{1}{2}}\Lambda''(\theta)^{-\frac{1}{2}}v, \theta, n\right) &= \left[\exp\left\{-\frac{v^2}{2n} + \eta\left(n^{-\frac{1}{2}}\Lambda''(\theta)^{-\frac{1}{2}}v, \theta\right)\right\} \right]^n \\ &= \left(E_\theta \left[e^{in^{-\frac{1}{2}}\Lambda''(\theta)^{-\frac{1}{2}}v(X_1 - \bar{x})} \right] \right)^n = \left[\varphi_\theta\left(n^{-\frac{1}{2}}\Lambda''(\theta)^{-\frac{1}{2}}v\right) \right]^n. \end{aligned}$$

Therefore, $\exp\{-v^2\} \psi^2\left(n^{-\frac{1}{2}}\Lambda''(\theta)^{-\frac{1}{2}}v, \theta, n\right)$ is also a characteristic function.

Assumption 1 There exist $\alpha_0 > 1$ and $\gamma \geq 1$ such that

$$\int_{u \in \mathfrak{R}} |u|^{\alpha_0} |\varphi_\theta(u)|^\gamma du = \int_{u \in \mathfrak{R}} |u|^{\alpha_0} |M(iu)|^\gamma du < \infty.$$

Now we are in position to specify the parameters of the proposed IS density. Recall that δ_1 was chosen so that (5), (6) and (7) hold. We choose $q_{\delta_1} < 1$ so that $|\varphi_{\theta^*}(v)| < q_{\delta_1}$ for $|v| \geq \delta_1$. Set

$$\alpha = \alpha_0$$

and

$$a_n = \delta_1 \sqrt{n\Lambda''(\theta^*)}.$$

Let $p_n = b_n \times (2\Phi(a_n) - 1)$. Due to the well known fact that

$$1 - \Phi(x) \sim \frac{1}{\sqrt{2\pi}x} \exp(-x^2/2),$$

we have

$$(2\Phi(a_n) - 1) \sim 1 - \frac{2}{\sqrt{2\pi n\Lambda''(\theta^*)}\delta_1} \exp(-n\delta_1^2\Lambda''(\theta^*)/2).$$

Note that for g to be a valid density function, we need $p_n < 1$. Select b_n to be a sequence of positive real numbers that converges to 1 in such a way that $b_n < 1/(2\Phi(a_n) - 1)$ and

$$\lim_{n \rightarrow \infty} \frac{q_{\delta_1}^{2n-\gamma} n^{\frac{1+\alpha}{2}}}{(1 - b_n(2\Phi(a_n) - 1))} = 0. \tag{9}$$

For example, it is easy to see that $b_n = 1 - 1/n^\xi$ for any $\xi > 0$ satisfies (9). For each n , let g_n denote the pdf of the form (8) with parameters α , a_n and b_n chosen as above. Figure 1, illustrates the shape of g_n . Let E_n and Var_n denote the expectation and variance, respectively, w.r.t. the density g_n .

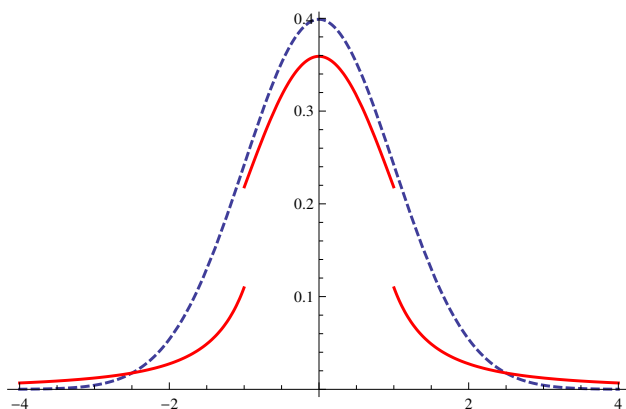


Figure 1: Dotted curve is the normal density function, while solid line is the density of the proposed IS density.

Theorem 1 Under Assumption 1,

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

Consequently, from Proposition 2 it follows that

$$Var_n \left[\frac{\psi \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} V, \theta^*, n \right) \phi(V)}{g(V)} \right] \rightarrow 0$$

so that the proposed estimator for $f_n(\bar{x})$ has asymptotically vanishing relative error.

The following lemma from Feller (1971) is useful to prove Theorem 1.

Lemma 1 For any $\lambda, \beta \in \mathbb{C}$,

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

Proof of Proposition 2: Writing $\zeta_3(\theta^*) = \Lambda'''(\theta^*)\Lambda''(\theta^*)^{-3/2}$ we have

$$\begin{aligned} \left| \int_{v \in \mathfrak{R}} \psi(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v, \theta^*, n) \phi(v) dv - 1 \right| &= \left| \int_{v \in \mathfrak{R}} \{ \psi(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v, \theta^*, n) - 1 \} \phi(v) dv \right| \\ &\leq \int_{v \in \mathfrak{R}} \left| \psi(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v, \theta^*, n) - 1 - \frac{\zeta_3(\theta^*)}{6\sqrt{n}} (tv)^3 \right| \phi(v) dv \\ &= \frac{1}{\sqrt{2\pi}} (I_1 + I_2), \end{aligned}$$

where

$$\begin{aligned} I_1 &= \int_{|n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v| < \delta_1} \left| \exp \left\{ n \eta(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v, \theta^*) \right\} - 1 - n \frac{\Lambda'''(\theta^*)}{3!} \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v \right)^3 \right| \exp \left\{ -\frac{v^2}{2} \right\} dv, \\ I_2 &= \int_{|n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v| \geq \delta_1} \left| \exp \left\{ n \eta(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v, \theta^*) \right\} - 1 - n \frac{\Lambda'''(\theta^*)}{3!} \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v \right)^3 \right| \exp \left\{ -\frac{v^2}{2} \right\} dv. \end{aligned}$$

We apply Lemma (1) with

$$\lambda = n \times \eta \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v, \theta^* \right) \text{ and } \beta = n \frac{\Lambda'''(\theta^*)}{3!} \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v \right)^3.$$

Since $\frac{|\beta|^2}{2} = \frac{Cv^6}{n}$, where C is a constant independent of n , and for $|n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v| < \delta_1$, we have from (7), (6) and (5), respectively,

$$|\lambda| = n \left| \eta \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v, \theta^* \right) \right| \leq n \frac{1}{8} \Lambda''(\theta^*) (n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v)^2 = \frac{|v|^2}{8},$$

$$|\beta| = n \left| \frac{1}{3!} \Lambda'''(\theta^*) \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v \right)^3 \right| \leq n \frac{1}{8} \Lambda''(\theta^*) (n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v)^2 = \frac{|v|^2}{8}$$

and

$$|\lambda - \beta| = n \left| \eta \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v, \theta^* \right) - \frac{1}{3!} \Lambda'''(\theta^*) \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v \right)^3 \right|$$

$$\leq n \varepsilon (\Lambda''(\theta^*))^{\frac{3}{2}} |n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v|^3 = \frac{\varepsilon |v|^3}{\sqrt{n}}.$$

From Lemma (1) it follows that the integrand in I_1 is dominated by

$$\exp \left\{ \frac{v^2}{8} \right\} \times \left(\frac{\varepsilon |v|^3}{\sqrt{n}} + \frac{Cv^6}{n} \right) \times \exp \left\{ -\frac{v^2}{2} \right\} = \exp \left\{ -\frac{3v^2}{8} \right\} \left(\frac{\varepsilon |v|^3}{\sqrt{n}} + \frac{Cv^6}{n} \right).$$

Since ε is arbitrary we have $I_1 = o(n^{-\frac{1}{2}})$.

Next we have

$$I_2 \leq \int_{|n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v| \geq \delta_1} \left| \exp \left\{ -\frac{v^2}{2} \right\} \psi(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v, \theta^*, n) \right| dv$$

$$+ \int_{|n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v| \geq \delta_1} \left(1 + \left| \frac{\zeta_3(\theta^*) v^3}{6} \right| \right) \exp \left\{ -\frac{v^2}{2} \right\} dv$$

$$= \int_{|\Lambda''(\theta^*)^{-\frac{1}{2}} v| \geq \delta_1 \sqrt{n}} |\varphi_{\theta^*} \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v \right)|^n dv$$

$$+ \int_{|\Lambda''(\theta^*)^{-\frac{1}{2}} v| \geq \delta_1 \sqrt{n}} \left(1 + \left| \frac{\zeta_3(\theta^*) v^3}{6} \right| \right) \exp \left\{ -\frac{v^2}{2} \right\} dv.$$

Recall that $q_{\delta_1} < 1$ is such that $|\varphi_{\theta^*}(v)| < q_{\delta_1}$ for $|v| \geq \delta_1$. Therefore, we have

$$I_2 \leq q_{\delta_1}^{n-\gamma} \int_{v \in \mathfrak{R}} |\varphi_{\theta^*} \left(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v \right)|^\gamma dv + \int_{|\Lambda''(\theta^*)^{-\frac{1}{2}} v| \geq \delta_1 \sqrt{n}} \left(1 + \left| \frac{\zeta_3(\theta^*) v^3}{6} \right| \right) \exp \left\{ -\frac{v^2}{2} \right\} dv$$

$$= q_{\delta_1}^{n-\gamma} n^{\frac{1}{2}} \sqrt{|\Lambda''(\theta^*)|} \int_{v \in \mathfrak{R}} |\varphi_{\theta^*}(u)|^\gamma du + \int_{|\Lambda''(\theta^*)^{-\frac{1}{2}} v| \geq \delta_1 \sqrt{n}} \left(1 + \left| \frac{\zeta_3(\theta^*) v^3}{6} \right| \right) \exp \left\{ -\frac{v^2}{2} \right\} dv.$$

It follows that $I_2 = o(n^{-\alpha})$ for any α .

The key step in the proof of Theorem 1 relies on expressing

$$\int_{v \in \mathfrak{R}} \frac{\psi^2(n^{-\frac{1}{2}} \Lambda''(\theta^*)^{-\frac{1}{2}} v, \theta^*, n) \phi^2(v)}{g_n(v)} dv$$

as the sum of I_3 and I_4 , where

$$I_3 = \int_{|\Lambda''(\theta^*)^{\frac{1}{2}}v| < \delta_1\sqrt{n}} \frac{\psi^2(n^{-\frac{1}{2}}\Lambda''(\theta^*)^{-\frac{1}{2}}v, \theta^*, n)\phi^2(v)}{g_n(v)} dv$$

and

$$I_4 = \int_{|\Lambda''(\theta^*)^{\frac{1}{2}}v| \geq \delta_1\sqrt{n}} \frac{\psi^2(n^{-\frac{1}{2}}\Lambda''(\theta^*)^{-\frac{1}{2}}v, \theta^*, n)\phi^2(v)}{g_n(v)} dv.$$

Substituting the value of $g_n(v)$, we get

$$I_3 = \frac{1}{b_n} \int_{|\Lambda''(\theta^*)^{\frac{1}{2}}v| < \delta_1\sqrt{n}} \psi^2(n^{-\frac{1}{2}}\Lambda''(\theta^*)v, \theta^*, n)\phi(v) dv$$

and

$$I_4 = \frac{1}{C_n} \int_{|\Lambda''(\theta^*)^{\frac{1}{2}}v| < \delta_1\sqrt{n}} |v|^\alpha \psi^2(n^{-\frac{1}{2}}\Lambda''(\theta^*)v, \theta^*, n)\phi^2(v) dv.$$

The proof now proceeds as the proof of Proposition 2 and shows that $I_3 = 1 + o(n^{-\frac{1}{2}})$ and that $I_4 \rightarrow 0$ as $n \rightarrow \infty$. The proof can be found in Dey and Juneja (2011).

4 EFFICIENT ESTIMATION OF THE TAIL PROBABILITY $P(\bar{X}_n \geq \bar{x})$

In this section we consider the problem of efficient estimation of $P(\bar{X}_n \geq \bar{x})$ for $\bar{x} > E[X_1]$. As discussed in the introduction, this relies on using a saddlepoint representation of $P(\bar{X}_n \geq \bar{x})$ obtained using Parseval's relation. Let θ^* denote the solution to the equation $\Lambda'(\theta) = \bar{x}$. Since, $\Lambda'(0) = EX_1$ and Λ' is strictly increasing, it follows that $\theta^* > 0$.

Let

$$Y_n = \sqrt{n}(\bar{X}_n - \bar{x}).$$

Let $h_{n,\theta,\bar{x}}(\cdot)$ be the density function of Y_n when each X_i has distribution function F_θ , where,

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

The following relation follows:

$$\begin{aligned} P[\bar{X} \geq \bar{x}] &= e^{n\Lambda(\theta^*) - n\theta^*\bar{x}} \int_0^\infty e^{-\sqrt{n}\theta^*y} h_{n,\theta^*,\bar{x}}(y) dy \\ &= \frac{e^{n\{\Lambda(\theta^*) - \theta^*\bar{x}\}}}{\sqrt{n}\theta^*} \int_0^\infty \sqrt{n}\theta^* e^{-\sqrt{n}\theta^*y} h_{n,\theta^*,\bar{x}}(y) dy \\ &= \frac{e^{n\{\Lambda(\theta^*) - \theta^*\bar{x}\}}}{2\pi\sqrt{n}\theta^*} \int_{-\infty}^\infty \frac{e^{-t\sqrt{n}\bar{x}}}{1 + \frac{t}{\sqrt{n}\theta^*}} \left[\frac{M\left(\theta^* + \frac{t}{\sqrt{n}}\right)}{M(\theta^*)} \right]^n dt, \end{aligned}$$

where the last equality follows from Parseval's relation applied to the exponential density $\sqrt{n}\theta^* e^{-\sqrt{n}\theta^*y}$ and the density of Y . Now by the change of variable $t = \Lambda''(\theta^*)^{-1/2}v$ and rearrangement of terms, the last integral may be re-expressed as,

$$\begin{aligned} &\frac{e^{n\{\Lambda(\theta^*) - \theta^*\bar{x}\}}}{2\pi\theta^*\sqrt{n\Lambda''(\theta^*)}} \int_{-\infty}^\infty \frac{1}{1 + \frac{lv}{\theta^*\sqrt{n\Lambda''(\theta^*)}}} \exp \left[n \left\{ \Lambda \left(\theta^* + \frac{lv}{\sqrt{n\Lambda''(\theta^*)}} \right) - \Lambda(\theta^*) - \frac{lv\bar{x}}{\sqrt{n\Lambda''(\theta^*)}} \right\} \right] dv \\ &= \frac{e^{n\{\Lambda(\theta^*) - \theta^*\bar{x}\}}}{\theta^*\sqrt{2\pi n\Lambda''(\theta^*)}} \int_{-\infty}^\infty \frac{\psi(v, \theta^*, n)\phi(v)}{1 + \frac{lv}{\theta^*\sqrt{n\Lambda''(\theta^*)}}} dv, \end{aligned} \tag{10}$$

Table 1: Comparison of the proposed methodology (PM) with optimal state independent exponential twisting (OET).

n=60			$c_n = 3.54 \times 10^{-4}$
N	OET	PM	Variance reduction
1000	3.17×10^{-4}	3.16×10^{-4}	42.78
10000	3.25×10^{-4}	3.23×10^{-4}	47.28
100000	3.24×10^{-4}	3.26×10^{-4}	47.75
n=120			$c_n = 8.62 \times 10^{-7}$
N	OET	PM	Variance reduction
1000	8.20×10^{-7}	8.29×10^{-7}	77.85
10000	8.33×10^{-7}	8.36×10^{-7}	75.89
100000	8.24×10^{-7}	8.25×10^{-7}	91.99
n=180			$c_n = 2.42 \times 10^{-9}$
N	OET	PM	Variance reduction
1000	2.33×10^{-9}	2.31×10^{-9}	80.60
10000	2.36×10^{-9}	2.44×10^{-9}	105.56
100000	2.30×10^{-9}	2.36×10^{-9}	98.79

where $\psi(v, \theta^*, n)$ is as defined in the last section. Expansion of (10) is well-known (see, Jensen (1995)) from where it follows that (10) converges to 1 as $n \rightarrow \infty$.

We again estimate the integral in (10) by Monte-Carlo simulation. Specifically, we re-express it as

$$\int_{-\infty}^{+\infty} \frac{\psi(v, \theta^*, n)\phi(v)}{\left(1 + \frac{tv}{\theta^* \sqrt{n\Lambda''(\theta^*)}}\right)} g(v) dv,$$

where $g(v)$ is a density whose support is \mathfrak{R} . Again, if V_1, V_2, \dots, V_N are iid samples from the density g , then the estimator given by

$$\hat{P}[\bar{X} \geq \bar{x}] := \frac{e^{n\{\Lambda(\theta^*) - \theta^* \bar{x}\}}}{\theta^* \sqrt{2\pi n\Lambda''(\theta^*)}} \frac{1}{N} \sum_{i=1}^N \frac{\psi(V_i, \theta^*, n)\phi(V_i)}{\left(1 + \frac{tV_i}{\theta^* \sqrt{n\Lambda''(\theta^*)}}\right)} g(V_i)$$

is an unbiased estimator for $P(\bar{X} \geq \bar{x})$. Then, analogous to Theorem 1, we can show that:

Theorem 2 Under Assumption 1, for g_n as in Theorem (1),

$$E_n \left[\frac{\psi^2(V, \theta^*, n)\phi^2(V)}{\left(1 + \frac{tV}{\theta^* \sqrt{n\Lambda''(\theta^*)}}\right)^2 g_n^2(V)} \right] = \int_{-\infty}^{+\infty} \frac{\psi^2(v, \theta^*, n)\phi^2(v)}{\left(1 + \frac{tv}{\theta^* \sqrt{n\Lambda''(\theta^*)}}\right)^2 g_n(v)} dv = 1 + o(n^{-\frac{1}{2}}) \text{ as } n \rightarrow \infty.$$

The proof of Theorem 2 can be found in Dey and Juneja (2011). It follows that our estimator for $P(\bar{X} \geq \bar{x})$ has asymptotically vanishing relative error.

5 NUMERICAL EXPERIMENTS

In this section we consider a simple numerical experiment to compare efficiency of the proposed method with the one involving state independent exponential twisting proposed by Sadowsky and Bucklew (1990).

We consider a sequence of rv's $(X_i : i \geq 1)$ that are iid exponentially distributed with mean 1. We estimate the probability $P(\bar{X}_n \geq 1.5)$ for different values of n . Table 5 reports the estimates based on N generated samples. c_n denotes the exact asymptotic corresponding to the probability. In these experiments we set $a_n = 2$, $\alpha = 2$ and $p_n = 0.9$ so that $b_n = p_n / (2 * \Phi(2) - 1) = 0.9 / 0.9545 = 0.9429$. We also report the variance reduction achieved by the proposed method over the one proposed by Sadowsky and Bucklew (1990). This is substantial and it increases with increasing n .

REFERENCES

- Asmussen, S., and P. Glynn. 2008. *Stochastic Simulation: Algorithms and Analysis*. New York: Springer.
- Blanchet, J., D. Leder, and P. Glynn. 2009. "Strongly efficient algorithms for light-tailed random walks: An old folk song sung to a faster new tune...". In *MCQMC 2008 edited by Pierre LECuyer and Art Owen.*, 227–248. Springer.
- Butler, R. 2007. *Saddlepoint Approximation with Applications*. Cambridge: Cambridge University Press.
- Daniels, H. E. 1954, ". "Saddlepoint Approximations in Statistics". *Annals of Mathematical Statistics* 25 (4): 631–650.
- Dey, S., and S. Juneja. 2011. "Efficient Simulation of Density and Large Deviations Probability of Sums of Random Vectors Using Saddle Point Representations". *Preprint*.
- Dieker, A. B., and M. Mandjes. 2005. "On Asymptotically Efficient Simulation of Large Deviations Probability". *Adv. in Appl. Probab.* 37 (2): 539–552.
- Feller, W. 1971. *An Introduction to Probability Theory and its Applications, Vol-2*. New York: John Wiley and Sons Inc.
- Glasserman, P., and S. Juneja. 2008. "Uniformly Efficient Importance Sampling for the Tail distribution of Sums of Random Variables". *Mathematics of Operation Research* 33 (1): 36–50.
- Glasserman, P., and J. Li. 2005. "Importance Sampling for Portfolio Credit Risk". *Management Science* 51:1643–1656.
- Glasserman, P., and Y. Wang. 1997. "Counterexamples in Importance Sampling for Large Deviation Probabilities". *The Annals of Applied Probability* 7 (3): 731–746.
- Jensen, J. 1995. *Saddlepoint Approximations*. Oxford: Oxford University Press.
- Juneja, S., and P. Shahabuddin. 2006. "Rare Event Simulation Techniques". *Handbooks in Operation Research and Management Science* 13:291–350.
- Parekh, S., and J. Walrand. 1989. "A Quick Simulation Method for Excessive Backlogs in Networks of Queue". *IEEE Transactions on Automatic Control* 34 (1): 54–66.
- Sadowsky, J. 1996. "On Monte Carlo Estimation of Large Deviation Probabilities". *The Annals of Applied Probability* 6:399–422.
- Sadowsky, J., and J. Bucklew. 1990. "On Large Deviation Theory and Asymptotically Efficient Monte Carlo Simulation Estimation". *IEEE Transactions on Information Theory* 36 (1): 579–588.

AUTHOR BIOGRAPHIES

SANTANU DEY is a PhD Student in the School of Technology and Computer Sciences at the Tata Institute of Fundamental Research, Mumbai, India. He received a Bachelors of Statistics from Indian Statistical Institute (ISI)-Kolkata, India, a Master of Statistics from Indian Statistical Institute (ISI)-Kolkata, India. His research focuses on application of stochastic processes and simulation to computational finance. His email address is dsantanu@tcs.tifr.res.in.

SANDEEP JUNEJA is an Associate Professor School of Technology and Computer Science at the Tata Institute of Fundamental Research. His Ph.D. was from Department of Operations Research at Stanford University (1993). His research interests include applied probability, computational finance, simulation and

Dey and Juneja

rare-event analysis. He serves on the editorial board on Mathematics of Operations Research. His email address is juneja@tifr.res.in and his web page is www.tcs.tifr.res.in/~sandeepj.