Towards zero variance estimators for rare event probabilities

Michel Broniatowski, Virgile Caron

Towards zero variance estimators for rare event probabilities

Michel Broniatowski and Virgile Caron

01/08/11

Abstract

Improving Importance Sampling estimators for rare event probabilities requires sharp approximations of conditional densities. This is achieved for events $E_n := (f(X_1) + ... + f(X_n)) \in A_n$ where the summands are i.i.d. and E_n is a large or moderate deviation event. The approximation of the conditional density of the vector $(X_1, ..., X_{k_n})$ with respect to E_n on long runs, when $k_n/n \to 1$, is handled. The maximal value of k_n compatible with a given accuracy is discussed; simulated results are presented, which enlight the gain of the present approach over classical IS schemes. Detailed algorithms are proposed.

Subclass :MSC 60-08 and MSC 65C05

Keywords: Importance sampling, rare event and large deviation, moderate deviation

[1]

1 Introduction and notation

1.1 Motivation and context

Importance Sampling procedures aim at reducing the calculation time which is necessary in order to evaluate integrals, often in large dimension. We consider the case when the integral to be numerically computed is the probability of an event defined by a large number of random components; this case has received quite a lot of attention, above all when the event is of *small* probability, typically of order 10^{-8} or so, as occurs frequently in industrial applications or

[1]M. Broniatowski
LSTA, Université Paris 6, France
michel.broniatowski@upmc.fr
V. Caron
LSTA, Université Paris 6, France
virgile.caron@upmc.fr

in communication devices.The present paper proposes estimators for both large and moderate deviation probabilities; this lattest case is of interest for statistics. The situation which is considered is the following.

The r.v's $\mathbf{X}, \mathbf{X}_i's$ are i.i.d. with known common density $p_{\mathbf{X}}$ on $\mathbb{R}$, and f is a real valued measurable function defined on $\mathbb{R}$. Let $\mu := Ef(\mathbf{X})$ and $\sigma^2 := Varf(\mathbf{X})$. Define

$$\mathbf{U} := \frac{f(\mathbf{X}) - \mu}{\sigma}$$

and

$$\overline{\mathbf{U}_1^n} := \frac{1}{n}\sum_{i=1}^{n} \mathbf{U}_i.$$

We intend to estimate

$$P_n := P\left(\overline{\mathbf{U}_1^n} \in A\right)$$

for large but fixed n where

$$A := (a_n, \infty) \tag{1}$$

and a_n is positive, either fixed (large deviation case) or satisfies $a_n \to 0$ slowly from above (moderate deviation case).

The basic estimate of $P(\overline{\mathbf{U}_1^n} \in A)$ is defined as follows: generate L i.i.d. samples $X_1^n(l)$ with underlying density $p_{\mathbf{X}}$ and define

$$P^{(n)}(A) := \frac{1}{L}\sum_{l=1}^{L} \mathbf{1}_{\mathcal{E}_n}\left(X_1^n(l)\right)$$

where

$$\mathcal{E}_n := \{(x_1, ..., x_n) \in \mathbb{R}^n : (u_1 + .. + u_n)/n \in A\} \tag{2}$$

with $u_i := (f(x_i) - \mu)/\sigma$. The Importance Sampling estimator of $P(\overline{\mathbf{U}_1^n} \in A)$ with sampling density g on $\mathbb{R}^n$ is

$$P_g^{(n)}(A) := \frac{1}{L}\sum_{l=1}^{L} \hat{P}_n(l)\mathbf{1}_{\mathcal{E}_n}\left(Y_1^n(l)\right) \tag{3}$$

where $\hat{P}_n(l)$ is called "importance factor" and writes:

$$\hat{P}_n(l) := \frac{\prod_{i=1}^{n} p_{\mathbf{X}}\left(Y_i(l)\right)}{g\left(Y_1^n(l)\right)} \tag{4}$$

and where the L samples $Y_1^n(l) := (Y_1(l), ..., Y_n(l))$ are i.i.d. with common density g.

The problem of finding a good sampling density g has been widely explored when $a_n = a$ is fixed and positive; this is the large deviation case; see e.g. Bucklew (2004). The case when a tends slowly to zero from above (the moderate deviation case) is considered in Ermakov (2007);

Under hypotheses to be recalled later, the *classical* IS scheme consists in the simulation of n i.i.d. replications $Y_1^{(l)}, ..., Y_n^{(l)}$ with density π^{a_n} on $\mathbb{R}$ and therefore $g(y_1, ..., y_n) = \pi^{a_n}(y_1)...\pi^{a_n}(y_n)$. The density π^{a_n} is the so-called *tilted* (or *twisted*) density at point a_n which , in case when $a_n = a$ is fixed, is called the *dominating point* of the set (a, ∞); see Bucklew (2004). In spite of the fact that this terminology is usually used in the large deviation case, we adopt it also in the moderate deviation one, for reasons to be stated later on.

This approach produces efficient IS schemes, in the sense that the computational burden necessary to obtain a relative precision of the estimate with respect to P_n does not grow exponentially as a function of n. It can be proved that in the large deviation range the variance of the classical IS is proportional to $P_n^2\sqrt{n}$.

The numerator in the expression (4) is the product of the $p_{\mathbf{X}_1}(Y_i)$'s while the denominator need not be a density of i.i.d. copies evaluated on the $Y_i's$. Indeed the optimal choice for g is the density of $\mathbf{X}_1^n := (\mathbf{X}_1, ..., \mathbf{X}_n)$ conditioned upon $\mathbf{X}_1^n \in \mathcal{E}_n$, leading to a zero variance estimator. We will propose an IS sampling density which approximates this conditional density very sharply on its first components $y_1, ..., y_k$ where $k = k_n$ is very large, namely $k/n \to 1$. This motivates the title of this paper.

Let us introduce a toy case in order to define the main step of the procedure, namely the simulation of a sample under a proxy of the conditional density. Assume $\mathbf{X}_1^n$ is a vector of n i.i.d. standard normal real valued random variables and $P_n := P(\mathbf{S}_1^n > na)$ with $\mathbf{S}_1^n := (\mathbf{X}_1 + ... + \mathbf{X}_n)$ with $a > 0$.

1- For any $s > a$ the joint density p_s of $\mathbf{X}_1, ...\mathbf{X}_{n-1}$ conditionally upon $(\mathbf{S}_1^n = ns)$ is known analytically and simulation under p_s is easy for any s. A general form of this statement is Theorem 1, Section 2.

2-The optimal sampling density g is similar to p_s with conditioning event $(\mathbf{S}_1^n > na)$ .The density g is obtained integrating p_s with respect to the the conditional distribution of $\mathbf{S}_1^n/n$ under $(\mathbf{S}_1^n > na)$ which is well approximated by an exponential distribution on (a, ∞) with expectation $a + 1/na$. The corresponding general statement is Theorem 2 Section 2. Therefore samples under a proxy of g are obtained through Monte Carlo simulation as follows: draw Y_1^n with density $p_{\mathbf{s}}$ where $\mathbf{s}$ follows the just cited exponential density. Insert these terms in (4) repeatedly to get $P_g^{(n)}$.

In the general case the joint distribution p_s cannot be approximated sharply on the very long run $1, ..., n-1$, but merely on $1, ..., k_n$ with k_n close to n. The approximation provided in Theorem 1 and, as a consequence in Theorem 2, is valid on the first k_n coordinates; a precise tuning of k_n is provided in Section 3. Since s is simulated on the whole set $(a, +\infty)$, no search is done in order to identify dominating points and no part of the target set $(a, +\infty)$ is neglected in the simulation of runs; the example in section 6, where the classical IS scheme is compared to the present one, is illuminating in this respect.

The merits of an IS estimator are captured through a number of criterions:

1. The asymptotic variance of the estimate
2. The stability of the Importance Factor
3. The hit rate of the IS scheme, which is the number of times the set $\mathcal{E}_n$ is reached by the simulated samples
4. Some run time indicator.

Some mixed indices have been proposed (see Glynn and Whitt (1992)), combining 1 and 4 with noticeable extension. The present paper provides an improvement over classical IS schemes as measured by 1, 2, 3 hereabove, as shown numerically on some examples. These progresses are also argued on a theoretical basis, following the quasi-optimality of the proposed IS scheme resulting from the approximation of the conditional density. When the r.v. $\mathbf{U}_i$'s are real-valued, the present method might be costly. The toy case which we present in the simulation study, pertaining to events $\left(|\overline{\mathbf{U}_1^n}|/n > a_n\right)$ under dissymetric $\mathbf{U}_i$'s proves however that the *observed bias* of the estimate through IS i.i.d. sampling can be important for reasonnable L, which does not happen with the present approach. Also the hit rate of the present proposal is close to 100%.

The criterion which we consider is different from the variance, and results as an evaluation of the MSE of our estimate on specific subsets of the runs generated by the sampling scheme, which we call *typical subsets*, namely having probability going to 1 under the sampling scheme as n increases. On such sets, the MSE is proved to be of very small order with respect to the variance of the classical estimate, which cannot be diminished on any such typical subsets. It will be shown that the relative gain in terms of simulation runs necessary to perform an α% relative error on P_n drops by a factor $\sqrt{n-k}/\sqrt{n}$ with respect to the classical IS scheme. Since k is allowed to be close to n, the resulting gain in variance is noticeable. Numerical evidence of this reduction in MSE is produced. Also we present a way of choosing the value of k_n with respect to n in such a way that the accuracy of the sampling scheme with respect to the optimal one is somehow controlled. This rule is argumented also numerically.

Alternative methods have been extensively developped for rare event simulation (see Botev and Kroese (2010) and references therein). The splitting technique results in an adhoc covering $A_1 \subset A_2 \subset ... \subset A$. It is assumed that the conditional distribution P_k of $\overline{\mathbf{U}_1^n}$ given $\overline{\mathbf{U}_1^n} \in A_k$ is known. An ad hoc choice of the A_k's leading to a common value for the P_k's provides efficient estimator for P_n, with small run-times. However in the present static case the calculation of the conditional distribution is difficult, even in the real case, and requires a sharp asymptotic analysis of large or moderate deviation probabilities.

It may seem that we could have reduced this paper to the case when f is the identity function, hence simulating runs $\mathbf{U}_1^k := (f(\mathbf{X}_1), ..., f(\mathbf{X}_k))$ under $\overline{\mathbf{U}_1^n} > a$. However it often occurs that the conditioning event is defined through a joint set of conditions, say

$$f(\mathbf{X}_1) + ... + f(\mathbf{X}_n) > n(\sigma a + \mu) \tag{5}$$

and

$$g\left(\mathbf{X}_1^n\right) \in B_n \tag{6}$$

for some function g and some measurable set B_n. Clearly in most cases the approximation of the density of $\mathbf{X}_1^k$ under both constraints is untractable and the approximation of the density of $\mathbf{X}_1^k$ conditionally on $\mathcal{E}_n$ provides a good IS sampling scheme for the estimation of

$$P\left(f\left(\mathbf{X}_1\right) + ... + f\left(\mathbf{X}_n\right) > n\left(\sigma a + \mu\right) \cap g\left(\mathbf{X}_1^n\right) \in B_n\right).$$

A simple example is when the constraint writes

$$\mathbf{X}_1^n \in D_n$$

and D_n is included in a set defined through (5). The function f and the value of a may be fitted such that (5) makes minimal the difference

$$\begin{aligned} &P\left(f\left(\mathbf{X}_1\right) + ... + f\left(\mathbf{X}_n\right) > n\left(\sigma a + \mu\right)\right) \\ &- P\left(\mathbf{X}_1^n \in D_n\right). \end{aligned}$$

Our proposal therefore hinges on the local approximation of the conditional distribution of longs runs $\mathbf{X}_1^k$ from $\mathbf{X}_1^n$. This cannot be achieved through the classical theory of large deviations, nor through the moderate deviations one, first developped by de Acosta (1992) and more recently by Ermakov (2007). At the contrary the ad hoc procedure developped in the range of large deviations by Diaconis and Freedman (1988) for the local approximation of the conditional distribution of $\mathbf{X}_1^k$ given the value of $\mathbf{S}_1^n := \mathbf{X}_1 + ... + \mathbf{X}_n$ is the starting point of the method leading to the present approach. We rely on Broniatowski and Caron (2010) where the basic approximation used in the present paper can be found. A first draft in the direction of the present work is in Broniatowski and Ritov (2009).

The present approach can be extended to the case of a multivariate constraint for a multidimensional problem, i.e. when for all x in $\mathbb{R}^d$, $f\left(x\right)$ and a are $\mathbb{R}^s$ -valued. This will not be considered here.

1.2 Regularity assumptions

The following notation and assumptions are kept throughout the paper without further reference.

The real valued measurable function f is assumed to be unbounded; standard transformations show that this assumption is not restrictive.

It is assumed that $\mathbf{U} = \left(f(\mathbf{X}) - \mu\right)/\sigma$ has a density $p_{\mathbf{U}}$ w.r.t. the Lebesgue measure on $\mathbb{R}$. We consider various assumptions on the regularity of $p_{\mathbf{U}}$ which each of them implies the validity of the improved Importance Sampling scheme. For these densities the approximation of the density of $\overline{\mathbf{U}_1^n}$ in the moderate and large deviation range, as well as the tail approximation of its distribution function is uniform. See Jensen (1995), Chapter 6.

1. Log-concave and almost Log-concave densities: $p_{\mathbf{U}}$ can be written as

$$p_{\mathbf{U}}(x) = c(x) \exp -h(x), \qquad x < \infty$$

with h a convex function, and where for some $x_0 > 0$ and constants $0 < c_1 < c_2 < \infty$, we have

$$c_1 < c(x) < c_2 \text{ for } x_0 < x < \infty.$$

Examples of densities which satisfy the above conditions include the Normal, the Gamma, the hyperbolic density, etc. An other example is when $\mathbf{U} := (\mathbf{X} - \psi)^2$ and $\mathbf{X}$ has log-normal distribution, $\mathbf{X} = \exp(\mathbf{Z})$ with $\mathbf{Z} \sim \mathcal{N}\left(\mu, \tau^2\right)$ and $\psi = E\mathbf{X} = \mu + \frac{1}{2}\tau$. Then

$$p_{\mathbf{U}}(x) = \frac{\exp(-\frac{3}{8}\tau)}{\sqrt{6\pi\tau^3}}$$

$$\left\{1 - \exp\left(-\sqrt{3x}\right)\right\} \exp\left\{\frac{\sqrt{3x}}{2} - \frac{x}{2\tau}\right\}$$

which is log-concave.

2. Gamma-like densities: the density of the r.v. $\mathbf{U}$ satisfies

$$p_{\mathbf{U}}(x) = c(x) \exp -h(x)$$

for all x with $0 < c_1 < c(x) < c_2 \leq \infty$ when x is larger than some $x_0 > 0$ and $h(x)$ is a convex function which satisfies $h(x) = \tau + h_1(x)$ with, for $x_1 < x_2$,

$$a_1 \log \frac{x_2}{x_1} - b_1 < h_1(x_2) - h_1(x_1) < a_2 \log \frac{x_2}{x_1} - b_2$$

where a_1, a_2, b_1 and b_2 are positive constants with $a_2 < 1$.

A wide class of densities for which our results apply is when there exist constants $x_0 > 0$, $\alpha > 0$, $\tau > 0$ and A such that

$$p_{\mathbf{U}}(x) = Ax^{\alpha-1} l(x) \exp(-\tau x) \qquad x > x_0$$

where $l(x)$ is slowly varying at infinity.

3. Densities defined through conditions on their characteristic function. Here is a set of conditions which can easily be verified, once given the characteristic function; see Jensen (1995) Chapter 6 for an exhaustive set of conditions which imply uniformity in the local and tail approximations of the distribution of the sample mean in the large (and, indeed, moderate) deviation case.

Denote $\varphi(\theta + it) := E \exp(\theta + it)\mathbf{U}$, assuming $|\varphi(\theta + it)| < \infty$ when θ belongs to some non void neighborhood Θ_0 of 0. Set $\kappa(\theta + it) :=$

$\log \varphi(\theta + it)$, $\mu(\theta + it) := (d/d\theta)\,\kappa(\theta + it)$ and $\sigma(\theta + it) := (d/d\theta)\ \mu(\theta + it)$.

Assume that there exist constants $c_1, c_2 > 0$, and ξ such that for θ in Θ_0

(i) $\left| \frac{\partial^k \kappa}{\partial t^k}(\theta + it\sigma(\theta)) \right|_{t=0} \leq c_1$ for $2 \leq k \leq 6$

(ii)for any $c > 0$ there exists $\rho < 1$ such that
$|\varphi(\theta + it\sigma(\theta))/\varphi(\theta)| \leq \rho$ for $|t| > c$

(iii) $\int |\varphi(\theta + it\sigma(\theta))/\varphi(\theta)|^{\xi} dt \leq c_2$.

We denote by (R) any of the above condition pertaining to the regularity of the density $p_{\mathbf{U}}(x)$.

We also assume that the characteristic function of the random variable $\mathbf{U}$ is assumed to belong to L^r for some $r \geq 1$.

1.3 Notation

1.3.1 Conditional densities and their approximations

Throughout the paper the value of a density $p_{\mathbf{Z}}$ of some continuous random vector $\mathbf{Z}$ at point z may be written $p_{\mathbf{Z}}(z)$ or $p(\mathbf{Z} = z)$, which may prove more convenient according to the context. The notation $\mathfrak{n}\left(\mu, \sigma^2, x\right)$ is defined through

$$\mathfrak{n}\left(\mu, \sigma^2, x\right) := \left(1/\sigma\sqrt{2\pi}\right) \exp\left[-\left((x-\mu)/\sigma\right)^2\right]$$

for all real x.

Let p_s denote the density of $\mathbf{X}_1^k$ under the local condition $\overline{\mathbf{U}_1^n} = s$

$$p_s\left(\mathbf{X}_1^k = y_1^k\right) := p(\mathbf{X}_1^k = y_1^k \big| \overline{\mathbf{U}_1^n} = s) \tag{7}$$

where y_1^k belongs to $\mathbb{R}^k$.

We will also consider the density p_A of $\mathbf{X}_1^k$ conditioned upon $\left(\overline{\mathbf{U}_1^n} > a\right)$

$$p_A\left(\mathbf{X}_1^k = y_1^k\right) := p(\mathbf{X}_1^k = y_1^k \big| \overline{\mathbf{U}_1^n} > a). \tag{8}$$

The approximating density of p_s is denoted g_s; the corresponding approximation of p_A is denoted g_A. Explicit formulas for those densities are presented in the next section.

For any of the above densities the symbols O_{p_s} (respectvely o_{p_s}), etc , designate remainder terms which are bounded (resp. go to 0) under the sampling with density p_s on $\mathbb{R}^k$. The same definition applies to O_{p_A} and o_{p_A} and the corresponding terms under the other sampling densities.

1.3.2 Tilted densities and related quantities

The r.v. $\mathbf{U}$ is supposed to fulfill the Cramer condition: its moment generating function satisfies

$$\phi_{\mathbf{U}}(t) := E \exp t\mathbf{U} < \infty$$

for t in a non void neighborhood of 0. Define the functions $m(t), s^2(t)$ and $\mu_3(t)$ as the first, second and third derivatives of $\log \phi_{\mathbf{U}}(t)$, and m^{-1} denote the reciprocal function of m.

Denote

$$\pi_{\mathbf{U}}^{\alpha}(x) := \frac{\exp t\left(f(x)-\mu\right)/\sigma}{\phi_{\mathbf{U}}(t)} p_{\mathbf{X}}(x)$$

with $m(t) = \alpha$ and α belongs to the support of $P_{\mathbf{U}}$, the distribution of $\mathbf{U}$. The density $\pi_{\mathbf{U}}^{\alpha}$ is the *tilted* density with parameter α. Also it is assumed that this lattest definition of t makes sense for all α in the support of $\mathbf{U}$. Conditions on $\phi_{\mathbf{U}}(t)$ which ensure this fact are referred to as *steepness properties*, and are exposed in Barndorff-Nielsen (1978), p153.

1.3.3 Specific sequences

Two sequences a_n and c_n are introduced in the paper. For notational convenience their current terms will be denoted a and c without referring to the subscript n. This convention holds only for these specific sequences, satisfying therefore conditions (A) and (C), to be introduced in the sequel.

2 Conditioned samples

The starting point is the approximation of p_s defined in (7) on $\mathbb{R}^k$ for large values of k under the point condition

$$\overline{\mathbf{U}_1^n} = s$$

when s belongs to (a, ∞). We refer to Broniatowski and Caron (2010) for this result.

We introduce a positive sequence ϵ_n which satisfies

$$\lim_{n\to\infty} \epsilon_n \sqrt{n-k} = \infty \tag{E1}$$

$$\lim_{n\to\infty} \epsilon_n \left(\log n\right)^2 = 0. \tag{E2}$$

Assume that

$$\lim_{n\to\infty} \frac{s^2}{\epsilon_n \left(\log n\right)} = \infty. \tag{As}$$

Define a density $g_s(y_1^k)$ on $\mathbb{R}^k$ as follows. Set

$$g_0(y_1 | y_0) := \pi_{\mathbf{U}}^{s}(y_1) \tag{9}$$

with y_0 arbitrary and, for $1 \le i \le k-1$, define $g(y_{i+1} | y_1^i)$ recursively.

Set t_i the unique solution of the equation

$$m_i := m(t_i) = \frac{n}{n-i}\left(s - \frac{\Sigma_1^i}{n}\right) \tag{10}$$

where $\Sigma_1^i := \sum_{j=1}^{i} (f(y_j) - \mu)/\sigma$.

Define

$$g(\, y_{i+1} |\, y_1^i) = C_i p_{\mathbf{X}}(y_{i+1})\ \mathfrak{n}\,(\alpha\beta + s, \alpha, (f(y_{i+1}) - \mu)/\sigma) \tag{11}$$

where C_i is a normalizing constant. Here

$$\alpha = s^2(t_i)\,(n - i - 1) \tag{12}$$

$$\beta = t_i + \frac{\mu_3\,(t_i)}{2s^4(t_i)\,(n - i - 1)}. \tag{13}$$

Set

$$g_s\left(y_1^k\right) := g_0(\, y_1 |\, y_0) \prod_{i=1}^{k-1} g(\, y_{i+1} |\, y_1^i). \tag{14}$$

Theorem 1 *Assume (As) and (E1,2) . Then (i)*

$$p_s\left(\mathbf{X}_1^k = Y_1^k\right) = g_s(Y_1^k)(1 + o_{p_s}(\epsilon_n\,(\log n)^2)) \tag{15}$$

and (ii)

$$p_s\left(\mathbf{X}_1^k = Y_1^k\right) = g_s(Y_1^k)(1 + o_{g_s}(\epsilon_n\,(\log n)^2)). \tag{16}$$

The approximation stated in the above statement (i) holds on *typical paths* generated under the conditional density p_s; in the same way, statement (ii) holds under the sampling scheme g_s. Therefore they do not hold on the entire space $\mathbb{R}^k$ which would require more restrictive hypotheses on the characteristic function of $f\,(\mathbf{X}_1)$; see Diaconis and Freedman (1988) for such conditions in the case when k is allowed to grow slowly with respect to n and a is fixed. However the above theorem provides optimal approximations on the entire space $\mathbb{R}^k$ for all k between 1 and $n-1$ in the gaussian case and $f(x) = x$, since $g_s\left(y_1^k\right)$ coincides with the conditional density.

As stated above the optimal choice for the sampling density is p_A for which we state an approximation result, extending Theorem 1.

Assume that a satisfies

$$\lim_{n\to\infty} \frac{a^2}{\epsilon_n\,(\log n)} = \infty. \tag{A}$$

The case when a does not depend on n satisfies (A) for any sequence ϵ_n under (E1,2). Conditions (A) and (E1,2) jointly imply that a cannot satisfy $a = O\left(\sqrt{\log\log n/n}\right)$; the Iterated Logarithm zone , as the Central Limit one, is not covered by our result.

Under these assumptions k can be fixed or can grow together with n with the restriction that $n-k$ should tend to infinity; when a is fixed this rate is governed through (E1) (or reciprocally given k ,ϵ_n is governed by k) independently on a.

In the moderate deviation case for a given sequence a close to 0, ϵ_n has rapid decrease, which in turn forces $n-k$ to grow rapidly.

We state the approximating density for p_A defined in (8). It holds

$$p_A(x_1^k) = \int_a^\infty p_s\left(\mathbf{X}_1^k = x_1^k\right) p(\overline{\mathbf{U}_1^n} = s \big| \overline{\mathbf{U}_1^n} > a) ds \tag{17}$$

so that, in contrast with the classical IS approach for this problem we will not consider the dominating point approach but merely realize a sharp approximation of the integrand at any point of the domain (a,∞) and consider the dominating contribution of all those distributions in the evaluation of the conditional density p_A. A similar point of view has been considered in Barbe and Broniatowski (2004) for sharp approximations of Laplace type integrals in $\mathbb{R}^d$.

The approximation of p_A is handled on some small interval $(a, a+c)$, thus on the principal part of this integral.

Let $c := c_n$ denote a positive sequence satisfying the following set of conditions, denoted (C).

$$\begin{gathered}(n-k)^{2-\delta} c \to \infty \\ nac \to \infty \\ c \to 0\end{gathered}$$

for some positive δ.

Let t_a be defined through the equation

$$m(t_a) = a.$$

Define on $\mathbb{R}^k$ the density

$$\begin{aligned} & g_A(y_1^k) \\ & := \frac{nm^{-1}(a) \int_a^{a+c} g_s(y_1^k) \left(\exp -nm^{-1}(a)(s-a)\right) ds}{\exp -nm^{-1}(a)\, c} \end{aligned} \tag{18}$$

where

$$\frac{nm^{-1}(a) \left(\exp -nm^{-1}(a)(s-a)\right) \mathbb{1}_{(a,a+c)}(s)}{\exp -nm^{-1}(a)\, c}$$

approximates $p(\overline{\mathbf{U}_1^n} = s \big| \overline{\mathbf{U}_1^n} \in (a, a+c))$.

It then holds

Theorem 2 *Assume (A) , (C) and (E1,2) . Then (i)*

$$p_A\left(\mathbf{X}_1^k = Y_1^k\right) = g_A(Y_1^k)(1 + o_{p_A}(\delta_n)) \tag{19}$$

and (ii)

$$p_A\left(\mathbf{X}_1^k = Y_1^k\right) = g_A(Y_1^k)(1 + o_{g_A}(\delta_n)) \tag{20}$$

where

$$\delta_n := \max\left(\epsilon_n (\log n)^2, \frac{1}{(n-k)^{2-\delta} c}, \exp -nca\right). \tag{21}$$

The proof is deferred to the Appendix.

3 How far is the approximation valid?

This section provides a rule leading to an effective choice of the crucial parameter $k = k_n$ in order to achieve a given accuracy bound for the relative error committed substituting p_A by g_A. The largest k the best the estimate of the rare event probability. We consider the large deviation case, assuming a fixed. The accuracy of the approximation is measured through

$$ERE(k) := E_{g_A}\left(1_{D_k}\left(Y_1^k\right)\frac{p_A\left(Y_1^k\right) - g_A\left(Y_1^k\right)}{p_A\left(Y_1^k\right)}\right)$$

and

$$VRE(k) := Var_{g_A}\left(1_{D_k}\left(Y_1^k\right)\frac{p_A\left(Y_1^k\right) - g_A\left(Y_1^k\right)}{p_A\left(Y_1^k\right)}\right) \tag{22}$$

respectively the expectation and the variance of the relative error of the approximating scheme when evaluated on D_k , the subset of $\mathbb{R}^k$ where $\left|g_A(Y_1^k)/p_A\left(Y_1^k\right) - 1\right| < \delta_n$; therefore $\int_{D_k} g_A(x)\,dx \to 1$.The r.v′s Y_1^k are sampled under g_A. Note that the density p_A is usually unknown. The argument is somehow heuristic and unformal; nevertheless the rule is simple to implement and provides good results. We assume that the set D_k can be substituted by $\mathbb{R}^k$ in the above formulas, therefore assuming that the relative error has bounded variance, which would require quite a lot of work to be proved under appropriate conditions, but which seems to hold, at least in all cases considered by the authors. We keep the above notation omitting therefore any reference to D_k .

Consider a two-sigma confidence bound for the relative accuracy for a given k, defining

$$CI(k) :=$$

$$\left[ERE(k) - 2\sqrt{VRE(k)}, ERE(k) + 2\sqrt{VRE(k)}\right].$$

Let δ denote an acceptance level for the relative accuracy. Accept k until δ belongs to $CI(k)$. For such k the relative accuracy is certified up to the level 5% roughly.

In Broniatowski and Caron (2010) a similar question is addressed and a proxy of the curve $\delta \to k_\delta$ is provided in order to define the maximal k leading to a given relative accuracy under the point condition $\overline{\mathbf{U}_1^n} = a_n$, namely when p_A is replaced by p_{a_n} and g_A by g_{a_n}.

Consider the ratio $g_A(Y_1^k)/p_A\left(Y_1^k\right)$ and use Cauchy's mean value theorem to obtain

$$g_A(Y_1^k)/p_A\left(Y_1^k\right)$$

$$= \frac{\int_a^{a+c} g_s(\mathbf{X}_1^k = Y_1^k)\left(\exp -nm^{-1}(a)(s-a)\right)ds}{\int_a^{a+c} p_s\left(\mathbf{X}_1^k = Y_1^k\right)\left(\exp -nm^{-1}(a)(s-a)\right)ds}$$

$$(1 + o_{g_A}(1))$$

$$= \frac{g_\alpha(\mathbf{X}_1^k = Y_1^k)}{p_\alpha\left(\mathbf{X}_1^k = Y_1^k\right)}\left(1 + o_{g_A}(1)\right)$$

for some α between a and $a + c$. Since a is fixed and $c \to 0$ it is reasonable to substitute α by a in order to evaluate the accuracy of the approximation. We thus inherit of the relative efficiency curve in Broniatowski and Caron (2010), to which we refer for definition and derivation.

We briefly state the necessary steps required for the calculation of the graph of a proxy of $k \to CI(k)$.

Introduce

$$D := \left[\frac{\pi_{\mathbf{U}}^a(a)}{p_{\mathbf{X}}(a)}\right]^n$$

and

$$N := \left[\frac{\pi_{\mathbf{U}}^{m_k}(m_k)}{p_{\mathbf{X}}(m_k)}\right]^{(n-k)}$$

with m_k defined in (10). Define t by $m(t) = a$. Define

$$A\left(Y_1^k\right) := \frac{n-k}{n}\left(\frac{g_A\left(Y_1^k\right)}{p_{\mathbf{X}}\left(Y_1^k\right)}\right)^3\left(\frac{D}{N}\right)^2\frac{s^2(t_k)}{s^2(t)}. \tag{23}$$

Simulate L i.i.d. samples $Y_1^k(l)$, each one made of k i.i.d. replications under $p_{\mathbf{X}}$; set

$$\widehat{A} := \frac{1}{L}\sum_{l=1}^{L} A\left(Y_1^k(l)\right).$$

We use the same approximation for B. Define

$$B\left(Y_1^k\right) := \sqrt{\frac{n-k}{n}}\left(\frac{g_A\left(Y_1^k\right)}{p_{\mathbf{X}}\left(Y_1^k\right)}\right)^2\left(\frac{D}{N}\right)\frac{s^2(t_k)}{s^2(t)} \tag{24}$$

and

$$\widehat{B} := \frac{1}{L}\sum_{l=1}^{L} B\left(Y_1^k(l)\right)$$

with the same $Y_1^k(l)'s$ as above.

Set

$$\overline{VRE}(k) := \widehat{A} - \widehat{B}^2. \tag{25}$$

which is a fair approximation of $VRE(k)$.

In the same way a proxy for ERE is defined through

$$\overline{ERE}(k) := 1 - \widehat{B}.$$

A proxy of $CI(k)$ can now be defined through

$$\overline{CI}(k) :=$$

$$\left[\overline{ERE}(k) - 2\sqrt{\overline{VRE}(k)}, \overline{ERE}(k) + 2\sqrt{\overline{VRE}(k)}\right]. \tag{26}$$

We now check the validity of the just above approximation, comparing $\overline{CI}(k)$ with $CI(k)$ on a toy case. Detailed algorithms leading to effective procedures are exposed in the next section.

Consider $f(x) = x$. The case when p is a centered exponential distribution with variance 1 allows for an explicity evaluation of $CI(k)$ making no use of Lemma 16. The conditional density p_s is calculated analytically, the density g_s is obtained through (14), hence providing a benchmark for our proposal. The terms $\widehat{A}$ and $\widehat{B}$ are obtained by Monte Carlo simulation following the algorithm presented hereunder. Tables 1,2 and 3,4 show the increase in δ w.r.t. k in the moderate deviation range, with a such that $P_n := P\left(\mathbf{S}_1^n > na\right) \simeq 10^{-2}$. In Table 5,6 and 7,8, a is such that $P\left(\mathbf{S}_1^n > na\right) \simeq 10^{-8}$ corresponding to a large deviation case. We have considered two cases, when $n = 100$ and when $n = 1000$. These tables show that the approximation scheme is quite accurate, since the relative error is fairly small even in very high dimension spaces. Also they show that $\overline{ERE}$ et $\overline{CI}$ provide good tools for the assessing the value of k. Denote $P_n := P\left(\mathbf{S}_1^n > na\right)$. The abscissa is k.

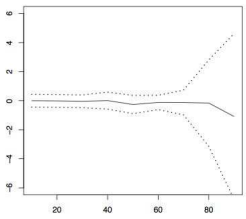

Figure 1: $\overline{CI}(k)$ and $P_n \simeq 10^{-2}$.

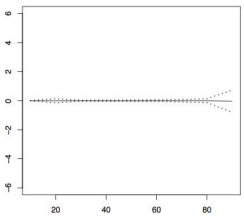

Figure 2: $CI(k)$ and $P_n \simeq 10^{-2}$.

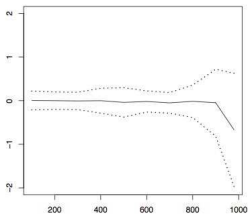

Figure 3: $\overline{CI}(k)$ and $P_n \simeq 10^{-2}$.

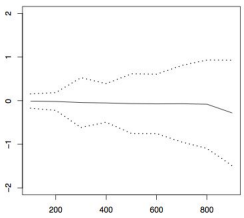

Figure 4: $CI(k)$ and $P_n \simeq 10^{-2}$.

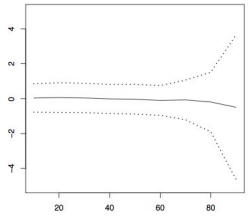

Figure 5: $\overline{CI}(k)$ and $P_n \simeq 10^{-8}$.

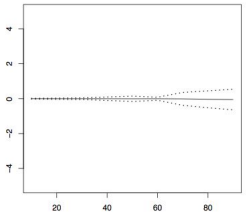

Figure 6: $CI(k)$ and $P_n \simeq 10^{-8}$.

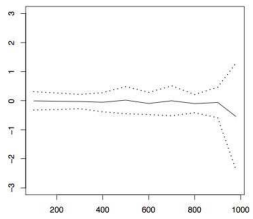

Figure 7: $\overline{CI}(k)$ and $P_n \simeq 10^{-8}$.

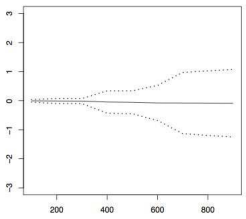

Figure 8: $CI(k)$ and $P_n \simeq 10^{-8}$.

4 The new Estimator and the algorithms

4.1 Adaptive IS Estimator for rare event probability

The IS scheme produces samples $Y := (Y_1, ..., Y_k)$ distributed under g_A, which is a continuous mixture of densities g_s as in (14), with exponential mixing measure with parameter $nm^{-1}(a)$ on (a, ∞)

$$\mathbb{1}_{(a,\infty)}(x) nm^{-1}(a) \exp\left[-nm^{-1}(a)(x-a)\right] \tag{27}$$

Since all IS schemes produce unbiaised estimators, and since the trucation parameter c in (18) is unmaterial, we consider untruncated versions of g_A defined in (18) integrating on (a, ∞) instead of $(a, a+c)$. This avoids a number of computational and programming questions, a difficult choice of an extra parameter c_n, and does not change the numerical results; this point has been checked carefully by the authors. Wee keep the notation g_A for the untruncated density.

The density g_A is extended from $\mathbb{R}^k$ onto $\mathbb{R}^n$ completing the $n-k$ remaining coordinates with i.i.d. copies of r.v's $Y_{k+1}, ..., Y_n$ with common tilted density

$$g_A\left(y_{k+1}^n \middle| y_1^k\right) := \prod_{i=k+1}^{n} \pi_{\mathbf{U}}^{\alpha}(y_i) \tag{28}$$

with $\alpha := m(t_k) = \frac{n}{n-k}\left(s - \frac{\Sigma_1^k}{n}\right)$ as in (10) and

$$\Sigma_1^k = \sum_{i=1}^{k} \frac{f(y_i) - \mu}{\sigma}$$

The last $n-k$ r.v's $\mathbf{Y}_i$'s are therefore drawn according to the classical i.i.d. scheme in phase with Sadowsky and Bucklew (1990) or Ermakov (2007) schemes in the large or moderate deviation setting.

We now define our IS estimator of $P_n := P\left(\overline{\mathbf{U}_1^n} > a\right)$.

Let $Y_1^n(l) := Y_1(l), ..., Y_n(l)$ be generated under g_A and let $S_1^n(l) := Y_1(l) + ... + Y_n(l)$. Let

$$\widehat{P_n}(l) := \frac{\prod_{i=0}^{n} p_{\mathbf{X}}(Y_i(l))}{g_A(Y_1^n(l))} \mathbf{1}_{\mathcal{E}_n}\left(S_1^n(l)\right) \tag{29}$$

and define

$$\widehat{P_n} := \frac{1}{L}\sum_{l=1}^{L} \widehat{P_n}(l). \tag{30}$$

in accordance with (3).

4.2 Algorithms

First, we present a series of three algorithms (Algorithms 1, 2 and 3) which produces the curve $k \to \overline{RE}(k)$. The resulting $k = k_\delta$ is the longest size of the runs which makes g_A a good proxy for p_A.

Algorithm 3 ***Evaluation of*** $g_s(y_1^k)$

1. ***INPUT*** *:* y_1^k, $p_{\mathbf{X}}$, n s.
2. ***OUTPUT*** *:* $g_s\left(y_1^k\right)$.
3. ***INITIALIZATION*** *:*

   $t_0 := m^{-1}(s)$
   $h_0(x_1 | x_1^0) := \pi^s(x_1)$
   $\Sigma_1^1 := (f(y_1) - \mu)/\sigma$

4. ***PROCEDURE*** *:*

   For i *from* 1 *to* $k-1$
   $m_i :=$*(10)*
   $t_i := m^{-1}(m_i)$ $*$
   $\alpha :=$*(12)*
   $\beta :=$*(13)*

Calculate C_i

$g(y_{i+1}|\, y_1^i) :=$*(11)*

endFor

Compute $g_s\left(y_1^k\right) :=$*(14)*

5. ***RETURN*** $g_s\left(y_1^k\right)$

Algorithm 4 ***Evaluation of*** $g_A\left(y_1^n\right)$

1. ***INPUT****:* y_1^n, $p_{\mathbf{X}}$, n, k, a, M
2. ***OUTPUT*** *:* $g_A\left(y_1^n\right)$
3. ***PROCEDURE***

For m *from* 1 *to* M

Simulate s_m *with density (27)*

Calculate $g_{s_m}\left(y_1^k\right)$ *with Algorithm 1*

Calculate $g_{s_m}\left(y_{k+1}^n|y_1^k\right) :=$ (28)

Calculate $g_{s_m}\left(y_1^n\right) =: g_{s_m}\left(y_1^k\right) g_{s_m}\left(y_{k+1}^n|y_1^k\right)$

endFor

Compute $g_A\left(y_1^n\right) := \frac{1}{M}\sum_{m=1}^{M} g_{s_m}\left(y_1^n\right)$

4. ***RETURN*** $g_A\left(y_1^n\right)$

The calculation of $g_s\left(y_1^k\right)$ above requires the value of

$$C_i = \left(\int p_{\mathbf{X}}(x)\mathfrak{n}\left(\alpha\beta + s, \beta, \left(f(x) - \mu\right)/\sigma\right) dx\right)^{-1}.$$

This can be done through Monte Carlo simulation. The value of M need not be very large.

Remark 5 *Solving* $t_i = m^{-1}(m_i)$ *might be difficult, even through a Newton Raphson technique and time consuming in simple cases. It may happen that the reciprocal function of* m *is at hand, as is assumed in Dupuis and Wang [DupuisWang2004], but even in such current situation as the Weibull distribution and* $f(x) = x$*, such is not the case. We can replace step* $*$ *by*

$$t_{i+1} := t_i - \frac{\left(m\left(t_i\right) + u_i\right)}{\left(n - i\right) s^2\left(t_i\right)}. \tag{31}$$

Indeed since

$$m(t_{i+1}) - m(t_i) = -\frac{1}{n-i}\left(m(t_i) + u_i\right)$$

with $U_i(l) := \left(f\left(Y_i(l)\right) - \mu\right)/\sigma$, *use a first order approximation to derive that* t_{i+1} *can be subtituted by* τ_{i+1} *defined through*

$$\tau_{i+1} := t_i - \frac{1}{(n-i)\, s^2(t_i)} \left(m(t_i) + u_i\right).$$

In the moderate deviation scale the function $s^2(.)$ *does not vary from* 1 *and the above approximation is fair. For the large deviation case, the same argument applies, since* $s^2(t_i)$ *keeps close to* $s^2(t_a)$.

Algorithm 6 ***Calculation of*** k_δ

1. ***INPUT*** *:* $p_{\mathbf{X}}$, δ, n, s, L
2. ***OUTPUT*** *:* k_δ
3. ***INITIALIZE*** *:* $k = 1$
4. ***PROCEDURE***

   Do
   For l *from* 1 *to* L
   Simulate $Y_1^k(l)$ *i.i.d. with density* $p_{\mathbf{X}}$
   $A\left(Y_1^k(l)\right)$:=*(23) using Algorithm 2*
   $B\left(Y_1^k(l)\right)$:=*(24) using Algorithm 2*
   endFor
   Calculate $\overline{CI}(k)$:=*(26)*
   $k := k + 1$
   While $\delta \notin \overline{CI}(k)$
   endDo

5. ***RETURN*** *:* $k_\delta := k$

The next algorithms 4, 5 and 6 provide the estimate of P_n.

The following algorithm provides a simple acceptance/rejection simulation tool for Y_{i+1} with density $g(\, y_{i+1} |\, y_1^i)$.

Denote $\mathfrak{N}$ the c.d.f. of a normal variate with parameter $\left(\mu, \sigma^2\right)$,and $\mathfrak{N}^{-1}$ its inverse.

Algorithm 7 ***Simulation of*** Y ***with density proportional to*** $p(x)\mathfrak{n}\left(\mu, \sigma^2, x\right)$

1. ***INPUT*** *:* p, μ, σ^2
2. ***OUTPUT*** *:* Y

3. ***INITIALIZATION*** :
   Select a density f *on* $[0,1]$ *and*
   a positive constant K *such that*
   $p\left(\mathfrak{N}^{-1}(x)\right) \leq Kf(x)$ *for all* x *in* $[0,1]$

4. ***PROCEDURE***

   Do

   Simulate X *with density* f
   Simulate U *uniform on* $[0,1]$ *independent of* X
   $Z := KUf(X)$
   While $Z < p\left(\mathfrak{N}^{-1}(X)\right)$

   endDo

5. ***RETURN*** $Y := \mathfrak{N}^{-1}(X)$

Algorithm 8 ***Simulation of a sample*** Y_1^k ***with density*** g_s

1. ***INPUT*** : $p_{\mathbf{X}}$, δ, n, s, a
2. ***OUTPUT*** : Y_1^k
3. ***INITIALIZATION*** :

   Set k*:=* k_δ *with Algorithm 3*
   $t_0 := m^{-1}(s)$

4. ***PROCEDURE***

   Simulate Y_1 *with density* $\pi_{\mathbf{U}}^s$
   $\Sigma_1^1 := (f(Y_1) - \mu)/\sigma$
   For i *from* 1 *to* $k-1$

   m_i :=*(10)*
   $t_i := m^{-1}(m_i)$
   α :=*(12)*
   β :=*(13)*
   Simulate Y_{i+1} *with density* $g(y_{i+1}|\, y_1^i)$
   using Algorithm 4
   $\Sigma_1^{i+1} := \Sigma_1^i + (f(Y_{i+1}) - \mu)/\sigma$

   endFor

5. ***RETURN*** Y_1^k

Remark 9 *The paper Barbe and Broniatowski (1999) can be used in order to simulate* Y_1.

Algorithm 10 ***Calculation of*** $\hat{P}_n$

1. ***INPUT*** *:* $p_{\mathbf{X}}$*,* δ n*,* a*,* M*,* L
2. ***OUTPUT*** *:*$\hat{P}_n$
3. ***INITIALIZATION***
   Set $k = k_\delta$ *with Algorithm 3*
4. ***PROCEDURE***

   For l *from* 1 *to* L

   Simulate s_l *with density (27)*
   Simulate $Y_1^k(l)$ *with density* g_{s_l} *with Algorithm 5*
   Simulate $Y_{k+1}^n(l)$ *i.i.d. with density* $\pi_{\mathbf{U}}^{\alpha_l}$
   Calculate $g_A\left(Y_1^n(l)\right)$ *with Algorithm 2*
   Calculate $\widehat{P_n}(l) :=$ (29)

   endFor

   Compute $\widehat{P_n} :=$ (30)
5. ***RETURN*** *:* $\widehat{P_n}$

Remark 11 $\pi_{\mathbf{U}}^{\alpha_l}$ *is defined as in (28)*

$$\alpha_l := m(t_k) = \frac{n}{n-k}\left(s_l - \frac{\Sigma_1^k}{n}\right)$$

as in (10) and

$$\Sigma_1^k = \sum_{i=1}^{k} \frac{f(Y_i(l)) - \mu}{\sigma}.$$

5 Compared efficiencies of IS estimators

The situation which we face with our proposal lacks the possibility to provide an order of magnitude of the variance our our IS estimate , since the properties necessary to define it have been obtained only on *typical paths* under the sampling density g_A and not on the whole space $\mathbb{R}^n$. This leads to a quasi-MSE measure for the performance of the proposed estimator, which quantifies the variability evaluated on classes of subsets of $\mathbb{R}^n$ whose probability goes to 1 under the sampling g_A. Not surprisingly the loss of performance with respect to the optimal sampling density $p_{\mathbf{X}_1^n/\mathcal{E}_n}$ is due to the $n-k$ last i.i.d. simulations, leading a quasi- MSE of the estimate proportional to $\sqrt{n-k}$.

5.1 The efficiency of the classical IS scheme

We first recall the definition of the classical IS sampling scheme and its asymptotic performance. The r.v.'s Y_i 's in (4) are i.i.d. and have density $g = \pi^a_{\mathbf{U}}$, hence with $m(t) = a$. See Sadowsky and Bucklew (1990) in the LDP case and Ermakov (2007) in the MDP case. The reason for this sampling scheme is the fact that in the large deviation case, a is the "dominating point" of the set (a, ∞) i.e. a is such that the proxy of the conditional distribution of $\mathbf{X}_1$ given $\overline{\mathbf{U}_1^n} > a$ is $\Pi^a_{\mathbf{U}}$; this is the basic form of the Gibbs conditioning principle.

Although developped for the large deviation case, the classical IS applies for the moderate deviation case since for $a \to 0$ and $a\sqrt{n} \to \infty$ it holds

$$P\left(\mathbf{X}_1 \in B | \, \overline{\mathbf{U}_1^n} > a\right) = (1 + o(1))\, \Pi^a_{\mathbf{U}}(B) \tag{32}$$

for any Borel set B as $n \to \infty$. This follows as a consequence of Sanov Theorem for moderate deviations (see Ermakov (2007) and de Acosta (1992)) and justifies the classical IS scheme in this range.

The classical IS is defined simulating L times a random sample of n i.i.d. r.v's $Y_1^n(j)$, $1 \le j \le L$, with tilted density $\pi^a_{\mathbf{U}}$. The standard IS estimate is defined through

$$\overline{P_n} := \frac{1}{L} \sum_{l=1}^{L} \mathbf{1}_{\mathcal{E}_n}(Y_1^n(l)) \frac{\prod_{i=1}^n p_{\mathbf{U}}(Y_i(l))}{\prod_{i=1}^n \pi^a_{\mathbf{U}}(Y_i(l))}$$

where the $X_i(l)$ are i.i.d. with density $\pi^a_{\mathbf{U}}$ and $\mathbf{1}_{\mathcal{E}_n}(l)$ is as in (2). Set

$$\overline{P_n}(l) := \mathbf{1}_{\mathcal{E}_n}(Y_1^n(l)) \frac{\prod_{i=1}^n p_{\mathbf{U}}(Y_i(l))}{\prod_{i=1}^n \pi^a_{\mathbf{U}}(Y_i(l))}.$$

The variance of $\overline{P_n}$ is given by

$$Var\overline{P_n} = \frac{1}{L}\left(E_{\pi^a_{\mathbf{U}}}\left(\overline{P_n}(l)\right)^2 - P_n^2\right).$$

The *relative accuracy* of the estimate $\overline{P_n}$ is defined through

$$RE(\overline{P_n}) := \frac{Var\overline{P_n}}{P_n^2} = \frac{1}{L}\left(\frac{E_{\pi_{\mathbf{U}}^a}\left(\overline{P_n}(l)\right)^2}{P_n^2} - 1\right).$$

The following result holds.

Proposition 12 *The relative accuracy of the estimate $\overline{P_n}$ is given by*

$$RE(\overline{P_n}) = \frac{\sqrt{2\pi}\sqrt{n}}{L}a(1+o(1))$$

as n tends to infinity.

We will prove that no reduction of the variance of the estimator can be achieved on subsets B_n of $\mathbb{R}^n$ such that $\Pi^a(B_n) \to 1$.

The easy case when $\mathbf{U}_1, ..., \mathbf{U}_n$ are i.i.d. with standard normal distribution and $f(x) = x$ is sufficient for our need.

The variance of the IS estimate of $P\left(\overline{\mathbf{U}_1^n} > a\right)$is proportional to

$$\begin{aligned} V &:= E_{p_{\mathbf{U}}}\mathbf{1}_{(a,\infty)}\left(\overline{\mathbf{U}_1^n}\right)\frac{p_{\mathbf{U}}\left(\mathbf{U}_1^n\right)}{\pi_{\mathbf{U}}^a\left(\mathbf{U}_1^n\right)} - P_n^2 \\ &= E_{p_{\mathbf{U}}}\mathbf{1}_{(a,\infty)}\left(\overline{\mathbf{U}_1^n}\right)\left(\exp\frac{na^2}{2}\right)\left(\exp -na\overline{\mathbf{U}_1^n}\right) - P_n^2 \end{aligned}$$

A set B_n resulting as reducing the MSE should penalize large values of $-\left(\mathbf{U}_1 + ... + \mathbf{U}_n\right)$ while bearing nearly all the realizations of $\mathbf{U}_1 + ... + \mathbf{U}_n$ under the i.i.d. sampling scheme $\pi_{\mathbf{U}}^a$ as n tends to infinity. It should therefore be of the form (b, ∞) for some $b = b_n$ so that

(a)

$$\lim_{n\to\infty} E_{\pi_{\mathbf{U}}^a}\mathbf{1}_{(b,\infty)}\left(\overline{\mathbf{U}_1^n}\right) = 1$$

and

(b)

$$\lim_{n\to\infty}\sup\frac{E_{p_{\mathbf{U}}}\mathbf{1}_{(a,\infty)\cap(b,\infty)}\left(\overline{\mathbf{U}_1^n}\right)\frac{p_{\mathbf{U}}(\mathbf{U}_1^n)}{\pi_{\mathbf{U}}^a(\mathbf{U}_1^n)}}{V} < 1$$

which means that the IS sampling density $\pi_{\mathbf{U}}^a$ can lead a MSE defined by

$$MSE(B_n) := E_{p_{\mathbf{U}}}\mathbf{1}_{(na,\infty)\cap(nb,\infty)}\frac{p_{\mathbf{U}}\left(\mathbf{U}_1^n\right)}{\pi_{\mathbf{U}}^a\left(\mathbf{U}_1^n\right)} - P_n^2$$

with a clear gain over the variance indicator. However when $b \leq a$ (b) does not hold and when $b > a$ (a) does not hold.

So no reduction of this variance can be obtained by taking into account the properties of the *typical paths* generated under the sampling density: a reduction of the variance is possible only by conditioning on "small" subsets of the sample paths space. On no classes of subsets of $\mathbb{R}^n$ with probability going to 1 under the sampling is it possible to reduce the variability of the estimate, whose rate is definitely proportional to $\sqrt{n}$, imposing a burden of order $L\sqrt{n}\alpha$ in order to achieve a relative efficiency of $\alpha\%$ with respect to P_n.

5.2 Efficiency of the adaptive twisted scheme

We first put forwards a Lemma which assesses that large sets under the sampling distribution g_A bear all what is needed to achieve a dramatic improvement of the relative efficiency of the IS procedure. Its proof is deferred to the Appendix.

Lemma 13 *Assume $k/n \to 1$. It then holds,*

1. *There exist sets C_n in $\mathbb{R}^n$ such that*
   - $\lim_{n\to\infty} G_A(C_n) = 1$
   - *for any y_1^n in C_n, $|\frac{p_A}{g_A}(y_1^k) - 1| < \delta_n$ with δ_n as in (21).*
2. - *when $a \to 0$ (moderate deviation),*
   $$t_k s(t_k) = a\,(1 + o(1)) \tag{33}$$
   - *when a is fixed (large deviation) , $t_k s(t_k)$ remains bounded away from 0 and infinity.*

We now evaluate the Mean Square Error of the adaptive twisted IS algorithm on this family of sets. Let

$$RE\left(\widehat{P_n}\right) = \frac{1}{L}\left(\frac{E_{g_A}\left(\mathbf{1}_{C_n}\widehat{P_n}(l)\right)^2}{P_n^2} - 1\right).$$

We prove that

Proposition 14 *The relative accuracy of the estimate $\hat{P}_n$ is given by*

$$RE(\widehat{P_n}) = \frac{\sqrt{2\pi}\sqrt{n-k-1}}{L} a(1 + o(1))$$

as n tends to infinity.

Proof. Using the definition of C_n we get

$$E_{g_A}\left(\mathbf{1}_{C_n}\widehat{P_n}(l)\right)^2$$

$$= P_n E_{p_A}\mathbf{1}_{C_n}(Y_1^n)\frac{p_{\mathbf{X}}(Y_1^k)p_{\mathbf{X}}(Y_{k+1}^n)}{g_A(Y_1^k)g_A(Y_{k+1}^n|\,Y_1^k)}$$

$$\leq P_n(1+\delta_n)E_{p_A}\mathbf{1}_{C_n}(Y_1^n)\frac{p_{\mathbf{X}}(Y_1^k)}{p(Y_1^k|\,\mathcal{E}_n)}\frac{p_{\mathbf{X}}(Y_{k+1}^n)}{g_A(Y_{k+1}^n|\,Y_1^k)}$$

$$= P_n^2(1+\delta_n)E_{p_A}\mathbf{1}_{C_n}(Y_1^n)\frac{1}{p(\mathcal{E}_n|\,Y_1^k)}\frac{p_{\mathbf{X}}(Y_{k+1}^n)}{g_A(Y_{k+1}^n|\,Y_1^k)}$$

$$= P_n^2(1+\delta_n)\sqrt{2\pi}\sqrt{n-k-1}$$
$$E_{p_A}\mathbf{1}_{C_n}(Y_1^n)t_k s(t_k)(1+o(1))$$
$$= P_n^2 a\sqrt{2\pi}\sqrt{n-k-1}(1+o(1)).$$

The third line is Bayes formula. The fourth line is Lemma 16 (see the Appendix). The fifth line uses (33) and uniformity in Lemma 16, where the conditions in Corollary 6.1.4 of Jensen (1995) are easily checked since, in his notation, $J(\theta) = \mathbb{R}$, condition (i) holds for θ in a neighborhood of 0 (Θ_0 indeed is resticted to such a set in our case), (ii) clearly holds and (iii) is a consequence of the assumption on the characteristic function of $f(\mathbf{X}_1)$. ■

6 Simulation results

6.1 The gaussian case

The random variables $X_i's$ are i.i.d. with normal distribution with mean 0 and variance 1. The case treated here is $P\left(\frac{\mathbf{S}_n}{n} > a\right) = P_n$ with $n = 100$, $P_n = 0.009972$ and $a = 0.232$. We build the curve of the estimate of P_n (solid lines) and the two sigma confidence interval (dot lines) with respect to k. The value of L is $L = 2000$.

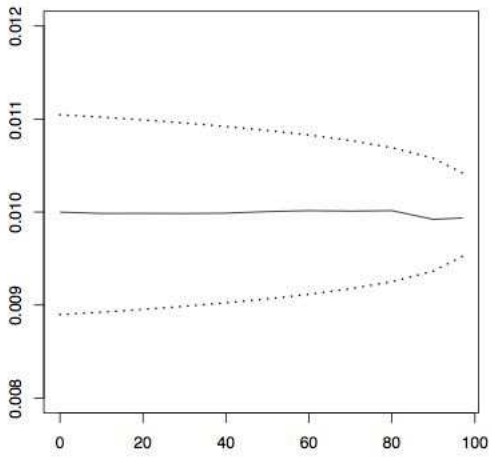

Figure 9: Estimator and confidence interval

6.2 The exponential case

The random variables $X_i's$ are i.i.d. with exponential distribution with parameter 1 on $(-1, \infty)$. The case treated here is $P\left(\frac{\mathbf{S}_n}{n} > a\right) = P_n$ with $n = 100$, $P_n = 0.013887$ and $a = 0.232$. The solid lines is the estimate of P_n, the dot lines are the two sigma confidence interval. Abcissa is k; $L = 2000$.

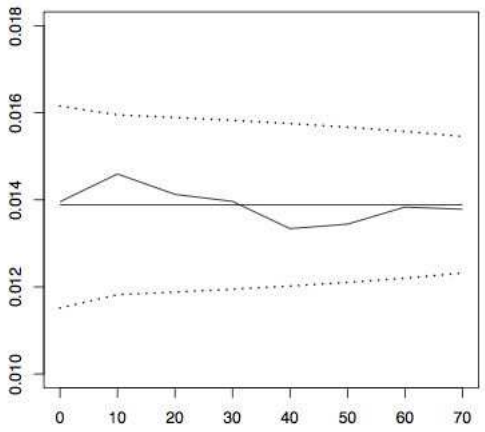

Figure 10: Estimator and confidence interval

Figure 11 shows the ratio of the empirical value of the MSE of the adaptive estimate w.r.t. the empirical MSE of the i.i.d. twisted one, in the exponential case with $P_n = 10^{-2}$ and $n = 100$. The value of k is growing from $k = 0$ (i.i.d. twisted sample) to $k = 70$ (according to the rule of section 3). This ratio stabilizes to $\sqrt{n-k}/\sqrt{n}$ for $L = 2000$. The abscissa is k and the solid line is $k \to \sqrt{n-k}/\sqrt{n}$.

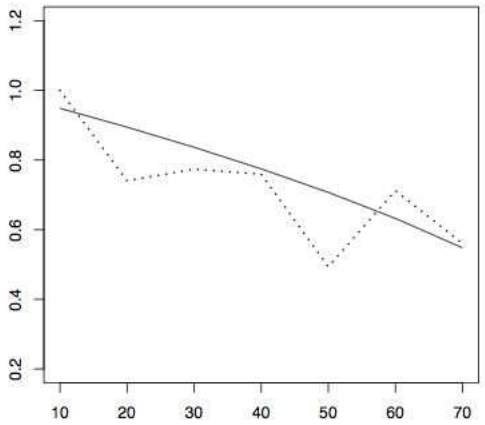

Figure 11: Empirical Ratio versus Theoretical Ratio

6.3 A comparison study with the classical twisted IS scheme

This section compares the performance of the present approach with respect to the standard tilted one as described in Section 1.

Consider a random sample $X_1, ..., X_{100}$ where X_1 has a normal distribution $N(0.05, 1)$ and let

$$\mathcal{E}_{100} := \left\{ x_1^{100} : \frac{|x_1 + ... + x_{100}|}{100} > 0.28 \right\}$$

for which

$$P_{100} = P\left((X_1, ..., X_{100}) \in \mathcal{E}_{100}\right) = 0.01120.$$

Our interest is to show that in this simple dissymetric case a direct extension of our proposal provides a good estimate, while the standard IS scheme ignores a part of the event $\mathcal{E}_{100}$. The standard i.i.d. IS scheme introduces the dominating point $a = 0.28$ and the family of i.i.d. tilted r.v's with common $N(a, 1)$ distribution. The resulting estimator of P_{100} is $0,01074$ (with $L = 1000$), indicating that the event $S_1^{100}/100 < -0.28$ is ignored in the evaluation of P_{100}, inducing

a bias in the estimation. Since the simulated r.v's are independent under the tilted distribution the Importance factor oscillates wildly. Also the hit rate is of order 50%. It can also be seen that $S_1^{100}/100 < -0.28$ is never visited through the procedure.

Drawing L i.i.d. points $s_1, ..., s_L$ according to the distribution of $S_1^{100}/100$ conditionally upon $\left|S_1^{100}\right|/100 > 0.28$ we evaluate P_{100} with $k = 99$; note that in the gaussian case Theorem 1 provides an exact description of the conditional density of X_1^k for all k between 1 and n , and therefore the same nearly holds in Theorem 2. Simulating the s_i's in this toy case is easy; just simulate samples $X_1, ..., X_{100}$ under $N(0.05, 1)$ until $\mathcal{E}_{100}$ is reached. The resulting value of the estimate is 0.01125 which is fairly close to P_{100}.

As expected the Importance factor is very close to P_{100} for all sample paths X_1^n simulated under G_A; this is in accordance with Theorem 1. Also the hit rate is very close to 100%.

The histograms pertaining to the Importance factor are as follows (Figures 12 and 13).

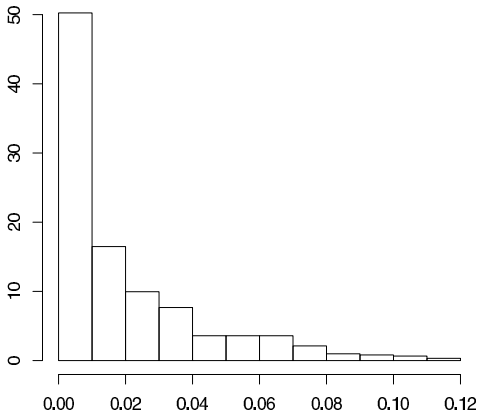

Figure 12: Histogram of Importance Factor for k=1

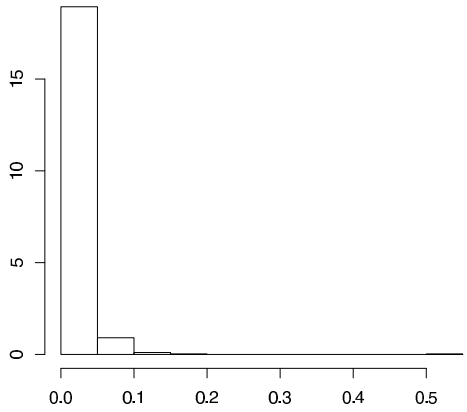

Figure 13: Histogram of Importance Factor for k=99

It is also interesting to draw the hit rate as a function of k. When $k = 1$ then this rate is close to 50%, since the present algorithms coincides with the classical i.i.d. IS scheme. As k increases, the hit rate approaches 100%; the value of L is 1000.

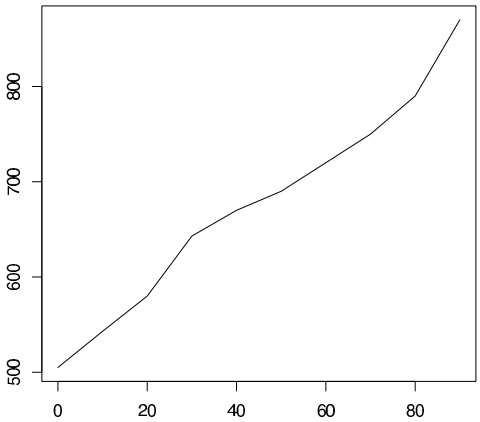

Figure 14: Hit Rate versus k

7 Appendix

The following two lemmas are used intensively throughout the paper; they provide asymptotic formulas for the density and the tail probability of the sample mean of i.i.d. summands $\mathbf{X}_i$ satisfying the Cramer condition; the functions $m(t)$ and $s^2(t)$ hereunder are the first and second derivatives of the moment generating function $\log Ee^{t\mathbf{X}}$. The function I is defined through

$$I(x) := xm^{-1}(x) - \log \phi\left(m^{-1}(x)\right)$$

where m is defined as in section 1 with $f(x) = x$. Their proofs are given in the large deviation scale in Rihter (1957) and Jensen (1995) and can easily be extended in the moderate deviation one (see also Hoglund (1979)); we omit details.

Lemma 15 *Under the general hypotheses and notation of this paper, when $a = a$ satisfies $\sqrt{n}a \to \infty$ it holds*

$$p\left(\frac{\mathbf{S}_1^n}{n} = a\right) = \frac{\sqrt{n}\exp -nI(a)}{s(t_a)\sqrt{2\pi}}\left(1 + o(1)\right).$$

Under condition (R) uniformity upon a holds. Here t_a satisfies $m(t_a) = a$.

The *global* counterpart of Lemma 15 in the form used here is due to Jensen (see Jensen (1995), corollary 6.4.1) and states

Lemma 16 *Under the same hypotheses as above*

$$P\left(\frac{\mathbf{S}_1^n}{n} > a\right) = \frac{\exp -nI(a)}{\sqrt{2\pi}\sqrt{n}\psi(a)}\left(1 + O(\frac{1}{\sqrt{n}})\right)$$

where $\psi(a) := t_a s(t_a)$.

7.1 Two Lemmas pertaining to the partial sum under its final value

Lemma 17 *It holds (i)$E_{p_A}\mathbf{U}_1 = a + o(1)$, (ii) $E_{p_A}\mathbf{U}_1^2 = 1 + s^2(t) + o(1)$ and (iii) $E_{p_A}\mathbf{U}_1\mathbf{U}_2 = a^2 + o(1)$ where $m(t) = a$.*

Proof. Under the point conditioning $\overline{\mathbf{U}_1^n} = s$ Lemma 12 in Broniatowski and Caron (2010) applies and provides $E_{p_s}\mathbf{U}_1 = s$, $E_{p_s}\mathbf{U}_1^2 = 1 + s_{\mathbf{U}}^2(t) + O\left(\frac{1}{n}\right)$ and $E_{p_s}\mathbf{U}_1\mathbf{U}_2 = s^2 + O\left(\frac{1}{n}\right)$ where p_s denotes the distribution of $\mathbf{U}_1^n$ under $\overline{\mathbf{U}_1^n} = s$. It holds

$$E_{p_A}\mathbf{U}_1 = \int_a^\infty E_{p_s}\mathbf{U}_1 p\left(\overline{\mathbf{U}_1^n} = s \middle| \overline{\mathbf{U}_1^n} > a\right) ds.$$

Using Condition (R) and Lemmas 15 and 16 it follows through a second order Taylor expansion that uniformly on s on (a,∞)

$$p\left(\overline{\mathbf{U}_1^n}=s\middle|\,\overline{\mathbf{U}_1^n}>a\right)=nm^{-1}(a) \tag{34}$$
$$\exp\left(-nm^{-1}(a)(s-a)\right)\frac{s\left(m^{-1}(a)\right)}{s\left(m^{-1}(s)\right)}(1+o(1))$$

where $s^2(t):=\frac{d}{dt}m(t)$ is bounded away from 0 as t is large, since $s^2(t)=Var_{\mathbf{U}_t}$ where $\mathbf{U}_t$ has d.f $\pi_{\mathbf{U}}^t(x):=e^{tx}p_{\mathbf{U}}(x)/Ee^{t\mathbf{U}}$. Furthemore for $c\to 0$ with $nca\to\infty$ it holds

$$\frac{\int_a^{a+c}E_{p_s}\mathbf{U}_1 p\left(\overline{\mathbf{U}_1^n}=s\middle|\,\overline{\mathbf{U}_1^n}>a\right)ds}{\int_a^{\infty}E_{p_S}\mathbf{U}_1 p\left(\overline{\mathbf{U}_1^n}=s\middle|\,\overline{\mathbf{U}_1^n}>a\right)ds}\to 1.$$

Indeed

$$\int_a^{a+c}E_{p_s}\mathbf{U}_1 p\left(\overline{\mathbf{U}_1^n}=s\middle|\,\overline{\mathbf{U}_1^n}>a\right)ds \tag{35}$$
$$=(1+o(1))$$
$$\int_a^{a+c}E_{p_s}\mathbf{U}_1 nm^{-1}(a)\exp\left(-nm^{-1}(a)(s-a)\right)ds$$

where we used the continuity of $t\to s^2(t)$ in $(a,a+c_n)$ and its regularity; indeed when a is fixed (large deviation case), use the continuity of $s^2(t)$ at $t=m^{-1}(a)$ whilst in the moderate deviation case ($a\to 0$), a first order expansion of $s\left(m^{-1}(s)\right)$ on $(a,a+c)$ provides (35). We have proved (i). The same arguments yield (ii) and (iii). We omit details. ■

Lemma 17 yields the maximal inequality stated in Lemma 13 in Broniatowski and Caron (2010) under the condition $\overline{\mathbf{U}_1^n}>a$.

We also need the order of magnitude of the maximum of $(\mathbf{U}_1,...,\mathbf{U}_k)$ under p_A which is stated in the following result.

Lemma 18 *It holds for all k between 1 and n*

$$\max(\mathbf{U}_1,...,\mathbf{U}_k)=O_{p_A}(\log n).$$

Proof. This Lemma is a slight modification of Lemma 14 in Broniatowski and Caron (2010). Denote P_s the distribution of $(\mathbf{X}_1,...,\mathbf{X}_n)$ under the point conditioning event $\overline{\mathbf{U}_1^n}=s$. It holds

$$p_A(\max(\mathbf{U}_1,...,\mathbf{U}_k)>t)\le kp_A(\mathbf{U}_n>t) \tag{36}$$

$$=k\int_a^{\infty}p_s(\mathbf{U}_n>t)\,p\left(\overline{\mathbf{U}_1^n}=s\middle|\,\overline{\mathbf{U}_1^n}>a\right)ds.$$

When c satisfies

$$\lim_{n\to\infty}nca=\infty$$

it holds

$$\frac{\int_a^{a+c} p_s\left(\mathbf{U}_n > t\right) p\left(\overline{\mathbf{U}_1^n} = s \middle| \overline{\mathbf{U}_1^n} > a\right) ds}{\int_a^{\infty} p_s\left(\mathbf{U}_n > t\right) p\left(\overline{\mathbf{U}_1^n} = s \middle| \overline{\mathbf{U}_1^n} > a\right) ds} = 1 + o(1).$$

Indeed

$$\begin{aligned}
&\int_{a+c}^{\infty} p_s\left(\mathbf{U}_n > t\right) p\left(\overline{\mathbf{U}_1^n} = s \middle| \overline{\mathbf{U}_1^n} > a\right) ds \\
&\leq C \int_{a+c}^{\infty} p_s\left(\mathbf{U}_n > t\right) \\
&\qquad nm^{-1}\left(a\right) \exp\left(-nm^{-1}\left(a\right)\left(s-a\right)\right) ds \\
&\leq C \exp -ncm^{-1}\left(a\right).
\end{aligned}$$

Turning back to (36) we provide an upper bound for $P_s\left(\mathbf{U}_n > t\right)$. We have, following the proof in Lemma 14 in Broniatowski and Caron (2010), substituting the r.v. $\mathbf{X}_n$ by $\mathbf{U}_n$.

$$p_s\left(\mathbf{U}_n > t\right) \leq Cste \frac{\phi_{\mathbf{U}}\left(\tau^s + \lambda\right)}{\phi_{\mathbf{U}}\left(\tau^s\right)} \int_t^{\infty} \frac{p\left(\mathbf{U}_n = v\right)}{\pi_{\mathbf{U}}^s\left(v\right)} e^{-\lambda v} dv$$

where $m\left(\tau^s\right) = s$ and λ is such that $\phi_{\mathbf{U}}\left(\tau^s + \lambda\right)$ is finite (such a λ exists since s is in $(a, a+c)$, hence bounded away from ∞). Inserting the above bound in (36) and restricting this integral to its principal part on $(a, a+c)$ concludes the proof. ■

Lemma 18 provides a similar statement as Lemma 14 in Broniatowski and Caron (2010) under $\overline{\mathbf{U}_1^n} > a$, which in turn proves Theorem 19 making use of the same Edgeworth expansions as in Theorem 7 of Broniatowski and Caron (2010).

7.2 Proof of Theorem 2

Proof of Theorem 2(i).

Step 1.We first prove that the integral (17) can be reduced to its principal part, namely that

$$p_A(Y_1^k) = \left(1 + o_{p_A}\left(1\right)\right)$$

$$\int_a^{a+c} p(\mathbf{X}_1^k = Y_1^k \middle| \overline{\mathbf{U}_1^n} = s) p(\overline{\mathbf{U}_1^n} = s \middle| \overline{\mathbf{U}_1^n} > a) ds \tag{37}$$

holds.

Apply Bayes formula to obtain

$$p_A(Y_1^k) = \frac{n p_{\mathbf{X}}\left(Y_1^k\right)}{(n-k)}$$

$$\frac{\int_a^{\infty} p\left(\overline{\mathbf{U}_{k+1}^n} = \frac{n}{n-k}\left(t - \frac{k\overline{\Sigma_1^k}}{n}\right)\right) dt}{P\left(\overline{\mathbf{U}_1^n} > a\right)}$$

where
$\overline{\Sigma_1^k} := \frac{1}{k}\sum_{i=1}^k U_i := \frac{1}{k}\sum_{i=1}^k \frac{f(Y_i)-\mu}{\sigma}$.
Denote

$$I := \frac{P\left(\overline{\mathbf{U}_{k+1}^n} > \frac{n}{n-k}\left(a+c-\frac{k\overline{\Sigma_1^k}}{n}\right)\right)}{P\left(\overline{\mathbf{U}_{k+1}^n} > \frac{n}{n-k}\left(a-\frac{k\overline{\Sigma_1^k}}{n}\right)\right)}.$$

Then (37) holds whenever $I \to 0$ (under p_A).

Under p_A it holds

$$\overline{\Sigma_1^k} = a + O_{p_A}\left(\frac{1}{nm^{-1}(a)}\right). \tag{38}$$

Indeed using classical results for moderate and large deviations (see Lemmas 15 and 16) for all positive u it holds, for $k = k_n \quad$ with $k/n \to 1$

$$\lim_{n\to\infty} P\left(\overline{\Sigma_1^k} > \left(a + \frac{u}{nm^{-1}(a)}\right)\middle| \overline{\Sigma_1^n} > a\right) = \exp -u. \tag{39}$$

Thus (38) holds whenever

$$nam^{-1}(a) \to \infty$$

which holds true when a is fixed or when $a \to 0$ together with $a\sqrt{n} \to \infty$, which follows from (E1,2) and (A). We now prove that $I \to 0$. Inserting (38) in I and using once more Lemmas 15 and 16 yields

$$I = (1 + o_{p_A}(1)) \exp -(n-k)$$

$$\left(I_{\mathbf{U}}\left(a + c - \frac{k\overline{\Sigma_1^k}}{n}\right) - I_{\mathbf{U}}\left(a - \frac{k\overline{\Sigma_1^k}}{n}\right)\right).$$

A second order expansion in the above display using (38) yields $I \to 0$ (under p_A) with an appropriate rate when

$$(n-k)^{2-\delta} c \to \infty$$

for some positive δ. which allows for fixed or small c. It then holds

$$I = O_{p_A}\left(\frac{1}{(n-k)^2 c}\right) = o_{p_A}\left(\frac{1}{(n-k)^{2-\delta} c}\right).$$

Step 2. (15) holds uniformly in s in $(a, a+c)$ where Y_1^k is generated under p_A. This result follows from a similar proof as in Broniatowski and Caron (2010) where (15) is proved under the local sampling p_s. A close look at the proof shows that (15) holds whenever Lemmas 12 and 14 in Broniatowski and Caron (2010), stated for the variables $\mathbf{U}_i$'s instead of $\mathbf{X}_i$'s hold under p_A. Those lemmas are substituted by Lemmas 17 and 18 hereabove.

Inserting (15) in (37) yields

$$p_A(Y_1^k) = \left(\int_a^{a+c} g_s(Y_1^k) p(\overline{\mathbf{U}_1^n} = s \big| \overline{\mathbf{U}_1^n} > a) ds \right) \left(1 + o_{p_A} \left(\max \left(\epsilon_n (\log n)^2, \frac{1}{(n-k)^{2-\delta} c} \right) \right) \right).$$

The conditional density of $\overline{\mathbf{U}_1^n}$ given $\left(\overline{\mathbf{U}_1^n} > a\right)$ is given in (39) which holds uniformly in s on the class of compact sets $(a, a+c)$ for a and c bounded.

Summing up we have proved

$$p_A(Y_1^k) =$$

$$\left(nm^{-1}(a) \int_a^{a+c} g_s(Y_1^k) \left(\exp -nm^{-1}(a)(s-a) \right) ds \right) \left(1 + o_{p_A} \left(\max \left(\epsilon_n (\log n)^2, \frac{1}{(n-k)^{2-\delta} c} \right) \right) \right)$$

as $n \to \infty$.

In order to get the approximation of p_A by the density g_A it is enough to observe that

$$nm^{-1}(a) \int_a^{a+c} g_s(Y_1^k) \left(\exp -nm^{-1}(a)(s-a) \right) ds = 1 + o_{p_A} \left(\exp -ncm^{-1}(a) \right)$$

as $n \to \infty$. When a is bounded away from 0 then so is $m^{-1}(a)$; when $\lim_{n\to\infty} a = 0$, $m^{-1}(a) \sim a$ and the condition

$$\lim_{n\to\infty} nca = \infty$$

leads to the completion of the proof.

Proof of Theorem 2(ii)

The following Lemma proves that approximating p_A by g_A under p_A is similar to approximating p_A by g_A under g_A.

Let $\mathfrak{R}_n$ and $\mathfrak{S}_n$ denote two p.m's on $\mathbb{R}^n$ with respective densities $\mathfrak{r}_n$ and $\mathfrak{s}_n$.

Lemma 19 *Suppose that for some sequence ε_n which tends to 0 as n tends to infinity*

$$\mathfrak{r}_n(Y_1^n) = \mathfrak{s}_n(Y_1^n)(1 + o_{\mathfrak{R}_n}(\varepsilon_n)) \tag{40}$$

as n tends to ∞. Then

$$\mathfrak{s}_n(Y_1^n) = \mathfrak{r}_n(Y_1^n)(1 + o_{\mathfrak{S}_n}(\varepsilon_n)). \tag{41}$$

Proof. Denote

$$A_{n,\delta\varepsilon_n} :=$$
$$\left\{y_1^n : (1-\varepsilon_n)\mathfrak{s}_n\left(y_1^n\right) \leq \mathfrak{r}_n\left(y_1^n\right) \leq \mathfrak{s}_n\left(y_1^n\right)(1+\varepsilon_n)\right\}.$$

It holds for all positive δ

$$\lim_{n\to\infty} I(n,\delta) = 1$$

where

$$I(n,\delta) := \int \mathbf{1}_{A_{n,\delta\varepsilon_n}}\left(y_1^n\right)\frac{\mathfrak{r}_n\left(y_1^n\right)}{\mathfrak{s}_n(y_1^n)}\mathfrak{s}_n(y_1^n)dy_1^n.$$

Since

$$I(n,\delta) \leq (1+\delta\varepsilon_n)\mathfrak{S}_n\left(A_{n,\delta\varepsilon_n}\right)$$

it follows that

$$\lim_{n\to\infty} \mathfrak{S}_n\left(A_{n,\delta\varepsilon_n}\right) = 1,$$

which proves the claim. ■

7.3 Proof of Lemma 13

Assume $k/n \to 1$. Let C_n in $\mathbb{R}^n$ such that for all y_1^n in C_n,

$$\left|\frac{p_A(y_1^k)}{g_A\left(y_1^k\right)} - 1\right| < \delta_n$$

with δ_n as in (21) and

$$\left|\frac{m(t_k)}{a} - 1\right| < \alpha_n$$

where t_k is defined through

$$m(t_k) := \frac{n}{n-k}\left(a - \frac{u_1^k}{n}\right)$$

with $u_1^k := \sum_{i=1}^k \left(f(y_i)-\mu\right)/\sigma$ and α_n satisfies

$$\lim_{n\to\infty} \alpha_n = 0 \tag{42}$$

together with

$$\lim_{n\to\infty} \alpha_n a\sqrt{n-k} = \infty. \tag{43}$$

We prove that

$$\lim_{n\to\infty} G_A\left(C_n\right) = 1.$$

Let

$$A_{n,\varepsilon_n} := A_{\varepsilon_n}^k \times \mathbb{R}^{n-k}$$

with

$$A_{\varepsilon_n}^k := \left\{x_1^k : \left|\frac{p_A(x_1^k)}{g_A\left(x_1^k\right)} - 1\right| < \delta_n\right\}.$$

By the above definition

$$\lim_{n\to\infty} P_A\left(A_{n,\varepsilon_n}\right) = 1. \tag{44}$$

Note also that

$$\begin{aligned} G_A\left(A_{n,\varepsilon_n}\right) &:= \int \mathbf{1}_{A_{n,\varepsilon_n}}(x_1^n) g_A\left(x_1^n\right) dx_1^n \\ &= \int \mathbf{1}_{A^k_{\varepsilon_n}}(x_1^k) g_A\left(x_1^k\right) dx_1^n \\ &\geq \frac{1}{1+\delta_n} \int \mathbf{1}_{A^k_{\varepsilon_n}}(x_1^k) p_A(x_1^k) dx_1^k \\ &= \frac{1}{1+\delta_n}\left(1+o(1)\right) \end{aligned}$$

which goes to 1 as n tends to ∞. We have just proved that the sequence of sets A_{n,ε_n} contains roughly all the sample paths X_1^n under the importance sampling density g_A.

We use the fact that t_k defined through

$$m(t_k) = \frac{n}{n-k}\left(a - \frac{U_1^k}{n}\right)$$

is close to a under p_s uniformly upon s in $(a, a+c)$ and integrate out with respect to the distribution of $\overline{\mathbf{U}_1^n}$ conditionally on $\overline{\mathbf{U}_1^n} \in (a, a+c)$.

Let δ_n tend to 0 and $\lim_{n\to\infty} a\alpha_n\sqrt{n-k} = \infty$ and

$$B_n := \left\{x_1^n : \left|\frac{m(t_k)}{a} - 1\right| < \alpha_n\right\}.$$

We prove that on B_n

$$t_k s(t_k) = a\left(1+o(1)\right) \tag{45}$$

holds.

By Lemma 13 in Broniatowski and Caron (2010) and integrating w.r.t. p_s on $(a, a+c)$ it holds, under (42) and (43)

$$\lim_{n\to\infty} P_A\left(B_n\right) = 1. \tag{46}$$

There exists δ_n' such that for any x_1^n in B_n

$$\left|\frac{t_k}{a} - 1\right| < \delta_n'. \tag{47}$$

Indeed

$$\left|\frac{m(t_k)}{a} - 1\right| = \left|\frac{t_k\left(1+v_k\right)}{a} - 1\right| < \alpha_n$$

and $\lim_{n\to\infty} v_k = 0$. Therefore

$$1 - \frac{v_k t_k}{a} - \alpha_n < \frac{t_k}{a} < 1 - \frac{v_k t_k}{a} + \alpha_n.$$

Since $\frac{m(t_k)}{a}$ is bounded so is $\frac{t_k}{a}$ and therefore $\frac{v_k t_k}{a} \to 0$ as $n \to \infty$ which implies (47).

Further (47) implies that there exists δ_n" such that

$$\left| \frac{t_k s(t_k)}{a} - 1 \right| < \delta_n".$$

Indeed

$$\begin{aligned} \left| \frac{t_k s(t_k)}{a} - 1 \right| &= \left| \frac{t_k \left(1 + u_k\right)}{a} - 1 \right| \\ &\leq \delta_n' + \left(1 + \delta_n'\right) u_k = \delta_n" \end{aligned}$$

where $\lim_{n\to\infty} u_k = 0$. Therefore (45) holds.

Define

$$C_n := B_n \cap A_{n,\varepsilon_n}$$

Since

$$\int \mathbf{1}_{C_n}(x_1^n) g_A\left(x_1^k\right) dx_1^n \geq \frac{1}{1+\delta_n} \int \mathbf{1}_{C_n} p_A(x_1^n) dx_1^n$$

and by (44) and (46)

$$\lim_{n\to\infty} P_A\left(C_n\right) = 1$$

we obtain

$$\lim_{n\to\infty} G_A\left(C_n\right) = 1.$$

which concludes the proof of (i) and (ii).

References

de Acosta, A.: Moderate deviations and associated Laplace approximations for sums of independent random vectors. Trans Amer Math Soc 329(1), 357-375 (1992)

Barbe, P., Broniatowski, M.: Simulation in exponential families. ACM Transactions on Modeling and Computer Simulation (TOMACS) 9(3), 203-223 (1992)

Barbe, P., Broniatowski, M.: On sharp large deviations for sums of random vectors and multidimensional Laplace approximation. Teor Veroyatn Primen 49(4), 743-774 (2004)

Barndorff-Nielsen, O.: Information and Exponential Families in Statistical Theory. Wiley, New-York (1978)

Botev, Z.I., Kroese, D.P.: Efficient Monte Carlo simulation via the Generalized Splitting Method. Statistics and Computing, DOI: 10.1007/s11222-010-9201-4 (2010)

Broniatowski, M., Caron, V.: Long runs under point conditioning. The real case. arxiv:1010.3616,2010 (2010)

Broniatowski, M., Ritov, Y.: Importance sampling for rare events and conditioned random walks. arxiv:0910.1819,2009 (2009)

Bucklew, J.A. Introduction to rare event simulation. Springer Series in Statistics, Springer-Verlag, New York (2004)

Diaconis, P., Freedman, D.A.: Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. J Theoret Probab 1(4), 381-410 (1988)

Ermakov, M.: Importance sampling for simulations of moderate deviation probabilities of statistics. Statist Decisions 25(4), 265-284 (2007)

Glynn, P.W., Whitt, W.: The asymptotic efficiency of simulation estimators. Oper. Res., 40(3), 505-520 (1992)

Höglund, T.: A unified formulation of the central limit theorem for small and large deviations from the mean. Z. Wahrsch. Verw. Gebiete 49(1), 105-117 (1979)

Jensen, J.L.: Saddlepoint Approximations. Oxford Statistical Science Series, vol 16. The Clarendon Press Oxford University Press, New York, Oxford Science Publications (1995)

Rihter, V.: Local limit theorems for large deviations. Dokl Akad Nauk SSSR (NS) 115:53-56 (1957)

Sadowsky, J.S., Bucklew, J.A. On large deviations theory and asymptotically efficient Monte Carlo estimation. IEEE Trans Inform Theory 36(3), 579-588 (1990)