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

February 6, 2012

Abstract

Improving Importance Sampling estimators for rare event probabilities requires sharp approximations of conditional densities. This is achieved for events $E_n := (u(X_1) + ... + u(X_n)) \in A_n$ 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.

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 in communication devices.The present paper proposes estimators for both large and moderate deviation probabilities; this latest 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 u is a real valued measurable function defined on $\mathbb{R}$. Define $\mathbf{U} := u(\mathbf{X})$ with density $p_{\mathbf{U}}$ and

$$\mathbf{U}_{1,n} := \sum_{i=1}^{n} \mathbf{U}_i.$$

We intend to estimate

$$P_n := P\left(\mathbf{U}_{1,n} \in nA\right)$$

for large but fixed n where

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

and a_n is a convergent sequence. The limit of this sequence either equals $E\mathbf{U}$ or is assumed to be larger than $E\mathbf{U}$. In the first case it will be assumed that a_n converges slowly in such a way that $P\left(\mathbf{U}_{1,n} \in nA\right)$ is not obtainable through the central limit theorem; we may call this case a moderate deviation case. The second situation is classically referred to as a large deviation case.

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

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

where

$$\mathcal{E}_n := \{(x_1, ..., x_n) \in \mathbb{R}^n : \left(u\left(x_1\right) + .. + u\left(x_n\right)\right) \in nA\} \tag{2}$$

with $u_i := u\left(x_i\right)$. The Importance Sampling estimator of P_n with sampling density g on $\mathbb{R}^n$ is

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

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 $E[u(\mathbf{X})]$ 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 $\mathbf{S}_{1,n} := \mathbf{X}_1 + ... + \mathbf{X}_n$ and $a > 0$.

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

2-The optimal sampling density g is similar to p_{nv} with conditioning event $(\mathbf{S}_{1,n} > na)$. The density g is obtained integrating p_{nv} 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 $p_{n\mathbf{V}}$ where $\mathbf{V}$ 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_{nv} 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 v 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 index 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 here-above, 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 $(|\mathbf{U}_{1,n}| > na_n)$ under $\mathbf{U}_i$'s proves however that the *observed bias* of the estimate through IS i.i.d. sampling can be important for reasonable 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 discussed also numerically.

Alternative methods have been extensively developed 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 $\mathbf{U}_{1,n}$ given $\mathbf{U}_{1,n} \in nA_k$ 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 u is the identity function, hence simulating runs $\mathbf{U}_1^k := (u(\mathbf{X}_1), ..., u(\mathbf{X}_k))$ under $(\mathbf{U}_{1,n} > na)$. However it often occurs that the conditioning event is defined through a joint set of conditions, say

$$u(\mathbf{X}_1) + ... + u(\mathbf{X}_n) > na \tag{5}$$

and

$$h(\mathbf{X}_1^n) \in B_n \tag{6}$$

for some function h and some measurable set B_n. Clearly in most cases the approximation of the density of $\mathbf{X}_1^k$ under both constraints is intractable and the approximation of the density of $\mathbf{X}_1^k$ conditioned upon $(\mathbf{X}_1^n \in \mathcal{E}_n)$ provides a good IS sampling scheme for the estimation of

$$P(u(\mathbf{X}_1) + ... + u(\mathbf{X}_n) > na \cap h(\mathbf{X}_1^n) \in B_n).$$

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 u and the value of a may be fitted such that (5) makes minimal the difference

$$\begin{aligned} &P(u(\mathbf{X}_1) + ... + u(\mathbf{X}_n) > na) \\ &- P(\mathbf{X}_1^n \in D_n). \end{aligned}$$

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 developed by [de Acosta 1992] and more recently by [Ermakov 2007]. At the contrary the ad hoc procedure developed 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 2011] 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 $\mathbb{R}^d$, $u(x)$ and a are $\mathbb{R}^s$ -valued. This will not be considered here.

1.2 Notations and Assumptions

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

1.2.1 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 normal density function on $\mathbb{R}$ with mean μ and variance τ at x is denoted $\mathfrak{n}(\mu, \tau, x)$.

Let p_{nv} denote the density of $\mathbf{X}_1^k$ under the local condition $(\mathbf{U}_{1,n} = nv)$

$$p_{nv}\left(\mathbf{X}_1^k = Y_1^k\right) := p(\mathbf{X}_1^k = Y_1^k \big| \mathbf{U}_{1,n} = nv) \tag{7}$$

where Y_1^k belongs to $\mathbb{R}^k$.

We will also consider the density p_{nA} of $\mathbf{X}_1^k$ conditioned upon $(\mathbf{U}_{1,n} > na)$

$$p_{nA}\left(\mathbf{X}_1^k = Y_1^k\right) := p(\mathbf{X}_1^k = Y_1^k \big| \mathbf{U}_{1,n} > na). \tag{8}$$

The approximating density of p_{nv} is denoted g_{nv}; the corresponding approximation of p_{nA} is denoted g_{nA}. Explicit formulas for those densities are presented in the next section.

1.2.2 Tilted densities and related quantities

The real valued measurable function u is assumed to be unbounded; standard transformations show that this assumption is not restrictive. It is assumed that $\mathbf{U} = u(\mathbf{X})$ has a density $p_{\mathbf{U}}$ w.r.t. the Lebesgue measure on $\mathbb{R}$. We also assume that the characteristic function of the random variable $\mathbf{U}$ is assumed to belong to L^r for some $r \geq 1$.

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

$$\pi^{\alpha}_{\mathbf{U}}(u) := \frac{\exp tu}{\phi_{\mathbf{U}}(t)} p_{\mathbf{U}}(u) \tag{9}$$

with $m(t) = \alpha$ and α belongs to the support of $P_{\mathbf{U}}$, the distribution of $\mathbf{U}$. The density $\pi^{\alpha}_{\mathbf{U}}$ is the *tilted* density with parameter α. Also it is assumed that this latest 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.

We also introduce the family of densities

$$\pi^{\alpha}_{u}(x) := \frac{\exp tu(x)}{\phi_{\mathbf{U}}(t)} p_{\mathbf{X}}(x). \tag{10}$$

with Π^{α}_{u} the associated distribution.

1.2.3 Specific sequences

The sequence a_n is introduced in the paper. For notational convenience its current terms will be denoted a without referring to the subscript n.

2 Conditioned samples

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

$$(\mathbf{U}_{1,n} = nv)$$

when v belongs to (a, ∞). We refer to [Broniatowski and Caron 2011] for this result.

We introduce a positive sequence ϵ_n which satisfies

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

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

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

$$g_0(y_1 | y_0) := \pi^{v}_{u}(y_1) \tag{11}$$

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

Set t_i the unique solution of the equation

$$m_i := m(t_i) = \frac{n}{n-i}\left(v - \frac{u_{1,i}}{n}\right) \tag{12}$$

where $u_{1,i} := u(y_1) + ... + u(y_i)$.

Define

$$g(y_{i+1} | y_1^i) = C_i p_{\mathbf{X}}(y_{i+1}) \mathfrak{n}(\alpha\beta + v, \alpha, u(y_{i+1})) \tag{13}$$

where C_i is a normalizing constant. Here

$$\alpha = s^2(t_i)(n-i-1) \tag{14}$$

$$\beta = t_i + \frac{\mu_3(t_i)}{2s^4(t_i)(n-i-1)}. \tag{15}$$

Set

$$g_{nv}(y_1^k) := g_0(y_1 | y_0) \prod_{i=1}^{k-1} g(y_{i+1} | y_1^i). \tag{16}$$

Theorem 1 *Assume (E1) and (E2). Then (i)*

$$p_{nv}\left(\mathbf{X}_1^k = Y_1^k\right) = g_{nv}(Y_1^k)(1 + o_{P_{nv}}(\epsilon_n (\log n)^2)) \tag{17}$$

and (ii)

$$p_{nv}\left(\mathbf{X}_1^k = Y_1^k\right) = g_{nv}(Y_1^k)(1 + o_{G_{nv}}(\epsilon_n (\log n)^2)). \tag{18}$$

The approximation stated in the above statement (i) holds on *typical paths* generated under the conditional density p_{ns}; in the same way, statement (ii) holds under the sampling scheme g_{ns}. Therefore they do not hold on the entire space $\mathbb{R}^k$ which would require more restrictive hypotheses on the characteristic function of $u(\mathbf{X}_1)$; 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 $u(x) = x$, since $g_{ns}\left(y_1^k\right)$ coincides with the conditional density. As stated in [Broniatowski and Caron 2011], the extension of our results from typical paths to the whole space $\mathbb{R}^k$ holds: convergence of the relative error on large sets imply that the total variation distance between the conditioned measure and its approximation goes to 0 on the entire space. So our results provide an extension of [Diaconis and Freedman 1988] and [Dembo and Zeitouni (1996)] who considered the case when k is of small order with respect to n; the conditions which are assumed in the present paper are weaker than those assumed in the just cited works; however, in contrast with their results, we do not provide explicit rates for the convergence to 0 of the total variation distance on $\mathbb{R}^k$.

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

We state the approximating density for p_{nA} defined in (8). It holds

$$p_{nA}(x_1^k) = \int_a^\infty p_{nv}\left(\mathbf{X}_1^k = x_1^k\right) p(\mathbf{U}_{1,n}/n = v |\, \mathbf{U}_{1,n} > na) dv \tag{19}$$

so that, in contrast with the classical IS approach for this problem we do 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_{nA}. 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_{nA} is handled on some small interval $(a, a+c)$, thus on the principal part of this integral.

Let c_n denote a positive sequence such that (C)

$$\lim_{n\to\infty} nc_n m^{-1}(a) = \infty$$
$$\sup_{n\geq 1} \frac{nc_n}{(n-k)} < \infty$$

and denote c the current term of the sequence c_n.

Denote (A) the following set of conditions

$$\lim_{n\to\infty} (n-k)\left(m^{-1}(a)\right)^2 = \infty$$
$$\lim_{n\to\infty} \frac{m^{-1}(a)}{\epsilon_n} = \infty$$

which trivially holds when $\lim_{n\to\infty} a_n > E\mathbf{U}$.

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

$$g_{nA}(y_1^k) := \frac{nm^{-1}(a)\int_a^{a+c} g_{nv}(y_1^k)\left(\exp -nm^{-1}(a)(v-a)\right) dv}{1 - \exp -nm^{-1}(a)\, c}. \tag{20}$$

The density

$$\frac{nm^{-1}(a)\left(\exp -nm^{-1}(a)(v-a)\right)\mathbb{1}_{(a,a+c)}(v)}{1 - \exp -nm^{-1}(a)\, c} \tag{21}$$

which appears in (20) approximates $p(\mathbf{U}_{1,n}/n = v |\, a < \mathbf{U}_{1,n}/n < a + c)$.

The *variance function* V of the distribution of $\mathbf{U}$ is defined on the span of $\mathbf{U}$ through

$$v \to V(v) := s^2(m^{-1}(v))$$

Denote (V) the condition

$$\sup_{n\geq 1}\left(\sqrt{n}m^{-1}(a)\int_a^{\infty}V'(v)\left(\exp -nm^{-1}(a)\left(v-a\right)\right)dv\right)<\infty. \qquad ((\mathrm{V}))$$

Theorem 2 *Assume (A), (C), ((V)), (E1) and (E2).. Then (i)*

$$p_{nA}\left(\mathbf{X}_1^k=Y_1^k\right)=g_{nA}(Y_1^k)(1+o_{P_{nA}}(\delta_n)) \tag{22}$$

and (ii)

$$p_{nA}\left(\mathbf{X}_1^k=Y_1^k\right)=g_{nA}(Y_1^k)(1+o_{G_{nA}}(\delta_n)) \tag{23}$$

where

$$\delta_n:=\max\left(\epsilon_n\left(\log n\right)^2,\left(\exp\left(-ncm^{-1}(a)\right)\right)^{\delta}\right). \tag{24}$$

for any positive $\delta<1$.

The proof of Theorem 2 is deferred to the Appendix.

Remark 3 *Most distributions used in statistics satisfy (V); numerous papers have focused on the properties of variance functions and classification of distributions. see e.g.[Letac and Mora (1990)] and references therein.*

Remark 4 *When a is fixed, the set of conditions (A) hold. In the case where $a=a_n$ converges to $E\mathbf{U}$, the set of conditions (A) does not cover the CLT zone. Indeed, the first condition of (A) implies that $m^{-1}(a)$ satisfies, for some $\delta>0$,*

$$|m^{-1}(a)n^{1/2+\delta}|<\infty.$$

Besides this limitation, choosing k and ϵ_n according to (A), (C), (E1) and (E2) is always possible. More a_n convergences slowly to $E\mathbf{U}$, more k can be choosen large with respect to n.

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_{nA} by g_{nA}. 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

$$ERE(k):=E_{G_{nA}}\left(1_{D_k}\left(Y_1^k\right)\frac{p_{nA}\left(Y_1^k\right)-g_{nA}\left(Y_1^k\right)}{p_{nA}\left(Y_1^k\right)}\right) \tag{25}$$

and

$$VRE(k):=Var_{G_{nA}}\left(1_{D_k}\left(Y_1^k\right)\frac{p_{nA}\left(Y_1^k\right)-g_{nA}\left(Y_1^k\right)}{p_{nA}\left(Y_1^k\right)}\right) \tag{26}$$

respectively the expectation and the variance of the relative error of the approximating scheme when evaluated on

$$D_k:=\left\{y_1^k\in\mathbb{R}^k\text{ such that }\left|g_{u_{1,n}}(y_1^k)/p_{u_{1,n}}\left(y_1^k\right)-1\right|<\delta_n\right\}$$

with $\epsilon_n\left(\log n\right)^2/\delta_n\to 0$ and $\delta_n\to 0$; therefore $G_{u_{1,n}}\left(D_k\right)\to 1$. The r.v's Y_1^k are sampled under g_{nA}. Note that the density p_{nA} is usually unknown. The argument is somehow heuristic and informal; 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 2011], 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 $(\mathbf{U}_{1,n} = na)$, namely when p_{nA} is replaced by p_{na} and g_{nA} by g_{na}.

Consider the ratio $g_{nA}(Y_1^k)/p_{nA}\left(Y_1^k\right)$ and use Cauchy's mean value theorem to obtain

$$g_{nA}(Y_1^k)/p_{nA}\left(Y_1^k\right)$$

$$= \frac{\int_a^{a+c} g_{nv}(\mathbf{X}_1^k = Y_1^k)\left(\exp -nm^{-1}(a)(v-a)\right)dv}{\int_a^{a+c} p_{nv}\left(\mathbf{X}_1^k = Y_1^k\right)\left(\exp -nm^{-1}(a)(v-a)\right)ds}$$

$$(1 + o_{G_{nA}}(1))$$

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

for some α between a and $a + c$. Since a and c are fixed, eventually small, 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 2011], 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

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

and

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

with m_k defined in (12). Define t by $m(t) = a$ and t^k by $m(t^k) = m_k$. Define

$$A\left(Y_1^k\right) := \frac{n-k}{n}\left(\frac{g_{nA}\left(Y_1^k\right)}{p_{\mathbf{X}}\left(Y_1^k\right)}\right)^3\left(\frac{N}{D}\right)^2\frac{s^2(t^k)}{s^2(t)}. \tag{27}$$

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

$$B\left(Y_1^k\right) := \sqrt{\frac{n-k}{n}}\left(\frac{g_{nA}\left(Y_1^k\right)}{p_{\mathbf{X}}\left(Y_1^k\right)}\right)^2\left(\frac{N}{D}\right)\frac{s^2(t^k)}{s^2(t)} \tag{28}$$

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

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

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

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

$$\left[\overline{ERE}(k) - 2\sqrt{\overline{VRE}(k)}, \overline{ERE}(k) + 2\sqrt{\overline{VRE}(k)}\right]. \tag{29}$$

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 $u(x) = x$. The case when $p_{\mathbf{X}}$ is a centered exponential distribution with variance 1 allows for an explicit evaluation of $CI(k)$ making no use of Lemma 11. The conditional density p_{nv} is calculated analytically, the density g_{nv} is obtained through (16), 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 large deviation range, with a such that $P_n := P(\mathbf{S}_{1,n} > na) \simeq 10^{-8}$. 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 $P_n := P(\mathbf{S}_{1,n} > na)$.

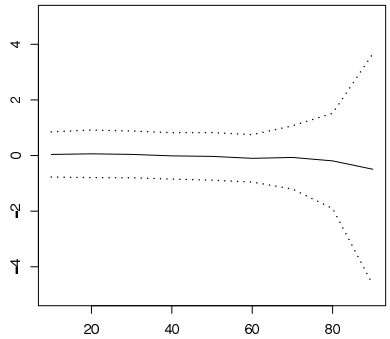

Figure 1: $\overline{ERE}(k)$(solid line) along with upper and lower bound of $\overline{CI}(k)$(dotted line) as a function of k with $n = 100$ and a such that $P_n \simeq 10^{-8}$.

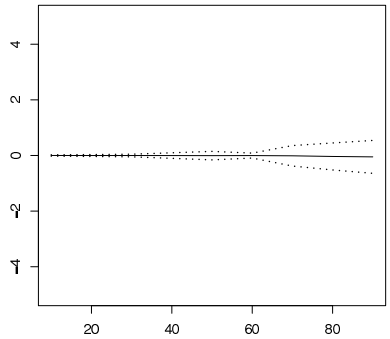

Figure 2: $ERE(k)$(solid line) along with upper and lower bound of $CI(k)$(dotted line) as a function of k with $n = 100$ and a such that $P_n \simeq 10^{-8}$.

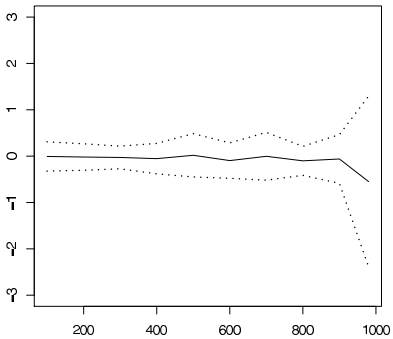

Figure 3: $\overline{ERE}(k)$(solid line) along with upper and lower bound of $\overline{CI}(k)$(dotted line) as a function of k with $n = 1000$ and a such that $P_n \simeq 10^{-8}$.

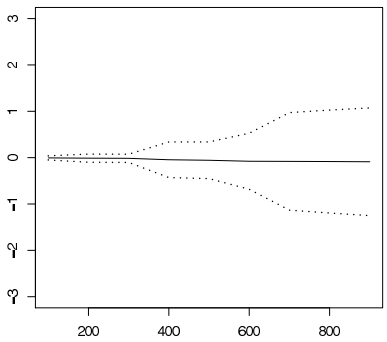

Figure 4: $ERE(k)$(solid line) along with upper and lower bound of $CI(k)$(dotted line) as a function of k with $n = 1000$ and a such that $P_n \simeq 10^{-8}$.

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_{nA}, which is a continuous mixture of densities g_{nv} as in (16), with exponential mixing measure with parameter $nm^{-1}(a)$ on (a, ∞)

$$\mathbb{1}_{(a,\infty)}(v) nm^{-1}(a) \exp\left[-nm^{-1}(a)(v-a)\right] \tag{30}$$

Since all IS schemes produce unbiased estimators, and since the truncation parameter c in (20) is immaterial, we consider untruncated versions of g_{nA} defined in (20) 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, and does not change the numerical results; this point has been checked carefully by the authors. Wee keep the notation g_{nA} for the untruncated density.

The density g_{nA} 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

$$g_{nA}\left(y_{k+1}^n \middle| y_1^k\right) := \prod_{i=k+1}^{n} \pi_u^{m_k}(y_i) \tag{31}$$

with $m_k := m(t^k) = \frac{n}{n-k}\left(v - \frac{u_{1,k}}{n}\right)$ as in (12) and

$$u_{1,k} = \sum_{i=1}^{k} u(y_i)$$

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 $P_n := P(\mathbf{U}_{1,n} > na)$.

Let $Y_1^n(l) := Y_1(l), ..., Y_n(l)$ be generated under g_{nA}. Let

$$\widehat{P_n}(l) := \frac{\prod_{i=0}^{n} p_{\mathbf{X}}(Y_i(l))}{g_{nA}(Y_1^n(l))} \mathbb{1}_{\mathcal{E}_n}(Y_1^n(l)) \tag{32}$$

and define

$$\widehat{P_n} := \frac{1}{L} \sum_{l=1}^{L} \widehat{P_n}(l). \tag{33}$$

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_{nA} a good proxy for p_{nA}.

Input : y_1^k, $p_{\mathbf{X}}$, n, v
Output : $g_{nv}\left(y_1^k\right)$

Initialization:
$t_0 \leftarrow m^{-1}(v)$;
$g_0(x_1|\, x_1^0) \leftarrow \pi_u^v(x_1)$;
$u_{1,1} \leftarrow u(y_1)$;

Procedure :
for $i \leftarrow 1$ **to** $k-1$ **do**
$m_i \leftarrow (12)$;
$t_i \leftarrow m^{-1}(m_i)$ $*$;
$\alpha \leftarrow (14)$;
$\beta \leftarrow (15)$;
Calculate C_i;
$g(y_{i+1}|\, y_1^i) \leftarrow (13)$;
$u_{1,i+1} \leftarrow u_{1,i} + u(y_{i+1})$;
end
Compute $g_{nv}\left(y_1^k\right) \leftarrow (16)$;

Return : $g_{nv}(y_1^k)$

Algorithm 1: Evaluation of $g_{nv}(y_1^k)$

Input : y_1^n, $p_{\mathbf{X}}$, n, k, a, M
Output : $g_{nA}\left(y_1^n\right)$

Procedure :
for $m \leftarrow 1$ **to** M **do**
Simulate v_m with density (30);
Calculate $g_{nv_m}\left(y_1^k\right)$ with Algorithm 1;
Calculate $g_{nv_m}\left(y_{k+1}^n|y_1^k\right) \leftarrow (31)$;
Calculate $g_{nv_m}\left(y_1^n\right) \leftarrow g_{nv_m}\left(y_1^k\right) g_{nv_m}\left(y_{k+1}^n|y_1^k\right)$
end
Compute $g_{nA}\left(y_1^n\right) \leftarrow \frac{1}{M}\sum_{m=1}^{M} g_{nv_m}\left(y_1^n\right)$;

Return : $g_{nA}\left(y_1^n\right)$

Algorithm 2: Evaluation of $g_{nA}\left(y_1^n\right)$

The calculation of $g_{nv}\left(y_1^k\right)$ above requires the value of

$$C_i = \left(\int p_{\mathbf{X}}(x) \mathfrak{n}\left(\alpha\beta + v, \beta, u(x)\right) dx \right)^{-1}.$$

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. It may happen that the reciprocal function of m is at hand, but even when $p_{\mathbf{X}}$ is the Weibull density and $u(x) = x$, such is not the case. We can replace step $*$ by*

$$t_{i+1} := t_i - \frac{\left(m\left(t_i\right) + u_i\right)}{(n-i)\, s^2\left(t_i\right)}. \tag{34}$$

Indeed since

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

with $u_i := u\left(y_i\right)$, 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 $s^2(t^a)$.

Input : $p_{\mathbf{X}}$, δ, n, a, L
Output : k_δ
Initialization: $k = 1$
Procedure :
while $\delta \notin \overline{CI}(k)$ **do**
for $l \leftarrow 1$ **to** L **do**
Simulate $Y_1^k(l)$ i.i.d. with density $p_{\mathbf{X}}$;
$A\left(Y_1^k(l)\right)$:=(27) using Algorithm 2 ;
$B\left(Y_1^k(l)\right)$:=(28) using Algorithm 2 ;
end
Calculate $\overline{CI}(k) \leftarrow$(29);
$k := k + 1$;
end
Return : $k_\delta := k$

Algorithm 3: Calculation of k_δ

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.

Input : p, μ, σ^2
Output : Y
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 $[0,1]$
Procedure :
while $Z < p\left(\mathfrak{N}^{-1}(X)\right)$ **do**
Simulate X with density f;
Simulate U uniform on $[0,1]$ independent of X;
Compute $Z := KUf(X)$;
end
Return : $Y := \mathfrak{N}^{-1}(X)$

Algorithm 4: Simulation of Y with density proportional to $p(x)\mathfrak{n}\left(\mu, \sigma^2, x\right)$

Remark 6 *The paper [Barbe and Broniatowski 1999] can be used in order to simulate* Y_1.

Input : $p_{\mathbf{X}}$, δ, n, v
Output : Y_1^k
Initialization:
Set $k \leftarrow k_\delta$ with Algorithm 3;
$t_0 \leftarrow m^{-1}(v)$;
Procedure :
Simulate Y_1 with density π_u^v;
$u_{1,1} \leftarrow u(Y_1)$;
for $i \leftarrow 1$ **to** $k-1$ **do**
$m_i \leftarrow$(12);
$t_i \leftarrow m^{-1}(m_i)$;
$\alpha \leftarrow$(14);
$\beta \leftarrow$(15);
Simulate Y_{i+1} with density $g(\, y_{i+1} | \, y_1^i)$ using Algorithm 4;
$u_{1,i+1} \leftarrow u_{1,i} + u(Y_{i+1})$;
end
Return : Y_1^k

Algorithm 5: Simulation of a sample Y_1^k with density g_{nv}

Input : $p_{\mathbf{X}}$, δ, n, a, M, L
Output : $\widehat{P_n}$
Initialization:
Set $k \to k_\delta$ with Algorithm 3;
Procedure :
for $l \leftarrow 1$ **to** L **do**
Simulate v_l with density (30);
Simulate $Y_1^k(l)$ with density g_{nv_l} with Algorithm 5;
Simulate $Y_{k+1}^n(l)$ i.i.d. with density $\pi_u^{\alpha_l}$;
Calculate $g_{nA}\left(Y_1^n(l)\right)$ with Algorithm 2;
Calculate $\widehat{P_n}(l) \leftarrow$ (32);
end
Compute $\widehat{P_n} \leftarrow$ (33);
Return : $\widehat{P_n}$

Algorithm 6: Calculation of $\hat{P}_n$

Remark 7 *$\pi_{\mathbf{U}}^{\alpha_l}$ is defined as in (31)*

$$\alpha_l := \frac{n}{n-k}\left(v_l - \frac{u_{1,k}}{n}\right)$$

as in (12) and

$$u_{1,k} = \sum_{i=1}^{k} u(Y_i(l)).$$

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_{nA} 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_{nA}. Not surprisingly the loss of performance with respect to the optimal sampling density 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 $g = \pi_u^a$, 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 $(\mathbf{U}_{1,n} > na)$ is Π_u^a; this is the basic form of the Gibbs conditioning principle.

Although developed for the large deviation case, the classical IS applies for the moderate deviation case since for $a \to E[u(\mathbf{X})]$ and $(a - E[u(\mathbf{X})])\sqrt{n} \to \infty$ it holds

$$P(\mathbf{X}_1 \in B | \mathbf{U}_{1,n} > na) = (1 + o(1))\,\Pi_u^a(B) \tag{35}$$

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 $Y_1^n(l)$, $1 \le l \le L$, with tilted density π_u^a. The standard IS estimate is defined through

$$\overline{P_n} := \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}_{\mathcal{E}_n}(Y_1^n(l)) \frac{\prod_{i=1}^n p_{\mathbf{X}}(Y_i(l))}{\prod_{i=1}^n \pi_u^a(Y_i(l))}$$

where the $X_i(l)$ are i.i.d. with density π_u^a and $\mathbb{1}_{\mathcal{E}_n}(Y_1^n(l))$ is as in (2). Set

$$\overline{P_n}(l) := \mathbb{1}_{\mathcal{E}_n}(Y_1^n(l)) \frac{\prod_{i=1}^n p_{\mathbf{X}}(Y_i(l))}{\prod_{i=1}^n \pi_u^a(Y_i(l))}.$$

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

$$Var\overline{P_n} = \frac{1}{L}\left(E_{\Pi_u^a}\left(\overline{P_n}(l)\right)^2 - P_n^2\right).$$

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

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

The following result holds.

Proposition 8 *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 $\Pi_u^a(B_n) \to 1$.

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

The variance of the IS estimate of $P(\mathbf{U}_{1,n} > na)$ is proportional to

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

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

(a)

$$\lim_{n\to\infty} E_{\Pi^a_{\mathbf{U}}} \mathbb{1}_{(b,\infty)} \left(\frac{\mathbf{U}_{1,n}}{n} \right) = 1$$

and

(b)

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

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

$$MSE(B_n) := E_{P_{\mathbf{U}}} \mathbb{1}_{(na,\infty)\cap(nb,\infty)} \frac{p_{\mathbf{U}}(\mathbf{U}_1^n)}{\pi^a_{\mathbf{U}}(\mathbf{U}_1^n)} - P_n^2$$

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_{nA} 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 9 *Assume $k/n \to 1$. It then holds,*

1. *There exist sets C_n in $\mathbb{R}^n$ such that*

   - $\lim_{n\to\infty} G_{nA}(C_n) = 1$
   - *for any y_1^n in C_n, $|\frac{p_{nA}}{g_{nA}}(y_1^k) - 1| < \delta_n$ with δ_n as in (24).*
   -
   - *when $a \to E\mathbf{U}$ (moderate deviation),*

   $$t^k s(t^k) = a(1 + o(1)) \tag{36}$$

   - *when $\lim_{n\to\infty} a_n$ is larger than $E\mathbf{U}$ (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

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

We prove that

Proposition 10 *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

$$E_{G_{nA}}\left(\mathbb{1}_{C_n}\widehat{P_n}(l)\right)^2$$

$$= P_n E_{P_{nA}}\mathbb{1}_{C_n}(Y_1^n)\frac{p_{\mathbf{X}}(Y_1^k)p_{\mathbf{X}}(Y_{k+1}^n)}{g_{nA}(Y_1^k)g_{nA}(Y_{k+1}^n\big|\, Y_1^k)}$$

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

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

$$= P_n^2(1+\delta_n)\sqrt{2\pi}\sqrt{n-k-1}$$

$$E_{P_{nA}}\mathbb{1}_{C_n}(Y_1^n)t^k s(t^k)(1+o(1))$$

$$= P_n^2 a\sqrt{2\pi}\sqrt{n-k-1}(1+o(1)).$$

The third line is Bayes formula. The fourth line is Lemma 11 (see the Appendix). The fifth line uses (36) and uniformity in Lemma 11, 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 restricted to such a set in our case), (ii) clearly holds and (iii) is a consequence of the assumption on the characteristic function of $u(\mathbf{X}_1)$. ■

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 $P_n = P\left(\frac{\mathbf{S}_{1,n}}{n} > a\right) = 0.009972$ with $n = 100$, 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 $L = 2000$.

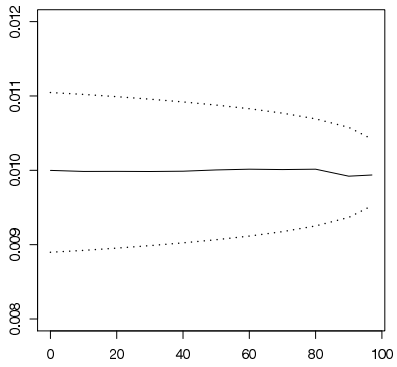

Figure 5: Curve of $\widehat{P_n}$ (solid line) in the normal case along with the two sigma confidence interval (dotted lines) as function of k with $n = 100$ for $L = 2000$ instances.

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 $P_n = P\left(\frac{\mathbf{S}_{1,n}}{n} > a\right) = 0.013887$ with $n = 100$, and $a = 0.232$. The solid lines is the estimate of P_n, the dot lines are the two sigma confidence interval. Abscissa is k.

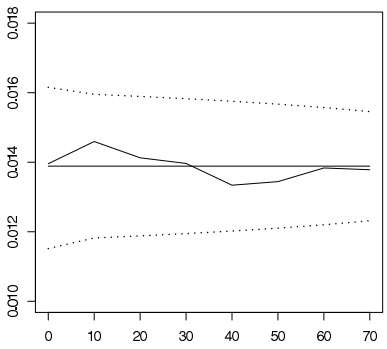

Figure 6: Curve of $\widehat{P_n}$ (solid line) in the exponential case along with the two sigma confidence interval (dotted lines) as function of k with $n = 100$ for $L = 2000$ instances.

Figure 7 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 $k \to \sqrt{n-k}/\sqrt{n}$.

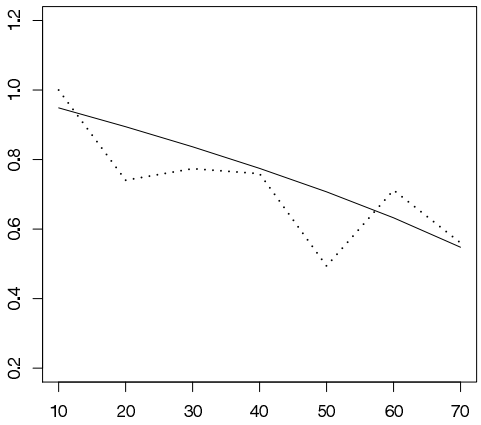

Figure 7: Ratio of the empirical value of the MSE of the adaptive estimate w.r.t. the empirical MSE of the i.i.d. twisted one (dotted line) along with the true value of this ratio (solid line) as a function of k.

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.

This example is not as artificial as it may seem; indeed it leads to a two dominating points situation which is quite often met in real life. Exploring at random the set of interest under the distribution of $(x_1 + ... + x_{100})/100$ under $\mathcal{E}_{100}$ avoids any search for dominating points. A further paper in $\mathbb{R}^d$ explores the advantage of this method, which already proves to compare favorably with usual methods on $\mathbb{R}$.

Drawing L i.i.d. points $v_1, ..., v_L$ according to the distribution of $S_{1,100}/100$ conditionally upon $|S_{1,100}|/100 > 0.28$ 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 v_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_{nA}; 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).

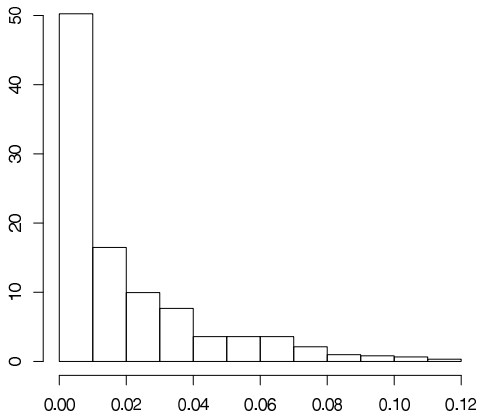

Figure 8: Histogram of Importance Factor with $k = 1$ and $n = 100$ for $L = 1000$ instances.

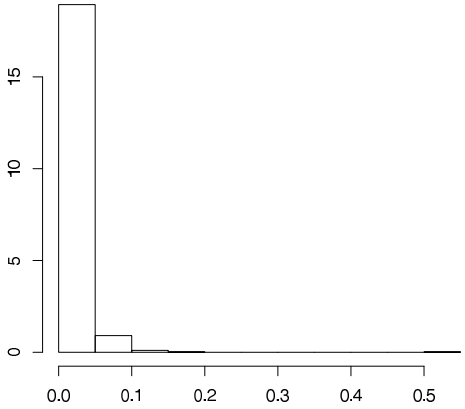

Figure 9: Histogram of Importance Factor with $k = 99$ and $n = 100$ for $L = 1000$ instances.

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.

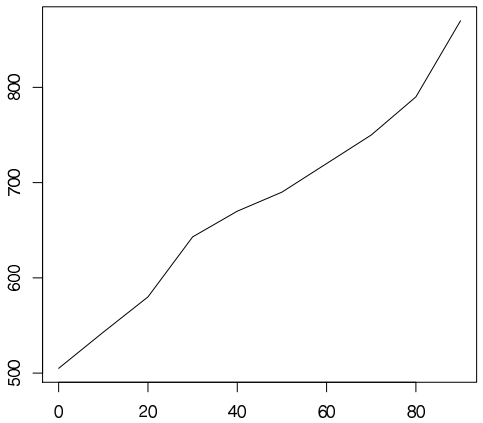

Figure 10: Curve of the hit Rate as a function of k with $n = 100$ for $L = 1000$ instances.

7 Appendix

The following lemma provide asymptotic formula for the tail probability of $\mathbf{U}_{1,n}$ under the hypothesis and notations of section 3. Define

$$I_{\mathbf{U}}(x) := xm^{-1}(x) - \log \phi_{\mathbf{U}}\left(m^{-1}(x)\right)$$

Lemma 11 *(see [Jensen 1995], Corollary 6.4.1) Under the same hypotheses and notations as section 3,*

$$P\left(\frac{\mathbf{U}_{1,n}}{n} > a\right) = \frac{\exp -nI_{\mathbf{U}}(a)}{\sqrt{2\pi}\sqrt{n}\psi(a)}\left(1 + O(\frac{1}{\sqrt{n}})\right)$$

where $\psi(a) := m^{-1}(a)s(m^{-1}(a))$.

7.1 Proof of Theorem 2

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

Lemma 12 *Suppose that (V) holds. Then (i)*$E_{P_{nA}}\mathbf{U}_1 = a + o(1)$, *(ii)* $E_{P_{nA}}\mathbf{U}_1^2 = a^2 + s^2\left(m^{-1}(a)\right) + o(1)$ *and (iii)* $E_{P_{nA}}\mathbf{U}_1\mathbf{U}_2 = a^2 + o(1)$.

Proof. We make use of Lemma 23 of [Broniatowski and Caron 2011], meaning $E_{P_{nv}}[\mathbf{U}_1] = v$. It holds

$$E_{P_{nA}}\mathbf{U}_1 = \int_a^\infty (E_{P_{nv}}\mathbf{U}_1)\, p\left(\mathbf{U}_{1,n}/n = v |\, \mathbf{U}_{1,n} > na\right) dv.$$

Integration by parts yields,

$$E_{P_{nA}}\mathbf{U}_1 = a + \int_a^\infty P\left(\mathbf{U}_{1,n}/n > v |\, \mathbf{U}_{1,n} > na\right) dv.$$

Using Lemma 11 and Chernoff inequality,

$$\int_a^\infty P\left(\mathbf{U}_{1,n}/n > v |\, \mathbf{U}_{1,n} > na\right) dv \le \sqrt{2\pi}\psi(a)\sqrt{n}\int_a^\infty \exp[n\left(I_{\mathbf{U}}(a) - I_{\mathbf{U}}(v)\right)]$$

where $\psi(a)$ is defined in Lemma 11.

Finally, using $I_{\mathbf{U}}(v) > I_{\mathbf{U}}^{'}(a)v + I_{\mathbf{U}}(a) - aI_{\mathbf{U}}^{'}(a)$, and integrating

$$\int_a^\infty P\left(\mathbf{U}_{1,n}/n > v |\, \mathbf{U}_{1,n} > na\right) dv \le \frac{\sqrt{2\pi}s(m^{-1}(a))}{\sqrt{n}}.$$

Hence, $E_{P_{nA}}\mathbf{U}_1 = a + o(1)$.

Insert $E_{P_{nv}}\mathbf{U}_1^2 = v^2 + s^2\left(m^{-1}(a)\right) + O\left(\frac{1}{n}\right)$ in

$$E_{p_{nA}}\mathbf{U}_1^2 = \int_a^\infty E_{P_{nv}}\mathbf{U}_1^2 p\left(\mathbf{U}_{1,n}/n = v |\, \mathbf{U}_{1,n} > na\right) dv.$$

Firstly, by integration by parts, Lemma 11 and Chernoff inequality,

$$\int_a^\infty v^2 p\left(\mathbf{U}_{1,n}/n = v |\, \mathbf{U}_{1,n} > na\right) dv = a^2 + o(1)$$

Indeed, since (C) implies $nm^{-1}(a) \to \infty$ when n tends to ∞, it holds

$$\int_a^\infty vp\left(\mathbf{U}_{1,n}/n > v |\, \mathbf{U}_{1,n} > na\right) dv \le \frac{s(m^{-1}(a))}{\sqrt{n}}\left(a + \frac{1}{nm^{-1}(a)}\right).$$

Secondly,

$$\int_a^\infty V(v)p\left(\mathbf{U}_{1,n}/n = v |\, \mathbf{U}_{1,n} > na\right) dv =$$

$$s^2(m^{-1}(a)) + 2\int_a^\infty V^{'}(v)P\left(\mathbf{U}_{1,n}/n > v |\, \mathbf{U}_{1,n} > na\right) dv.$$

Using Lemma 11, Chernoff inequality and $I_{\mathbf{U}}(v) > I_{\mathbf{U}}^{'}(a)v + I_{\mathbf{U}}(a) - aI_{\mathbf{U}}^{'}(a)$, it holds under condition (V),

$$\begin{aligned}
&\int_a^\infty V^{'}(v)P\left(\mathbf{U}_{1,n}/n > v |\, \mathbf{U}_{1,n} > na\right) dv \\
\le\ & s(m^{-1}(a))\left(\sqrt{n}m^{-1}(a)\int_a^\infty V^{'}(v)\exp\left(-nm^{-1}(a)(v-a)\right) dv\right)
\end{aligned}$$

and

$$\int_a^\infty V(v)p\left(\mathbf{U}_{1,n}/n = v |\, \mathbf{U}_{1,n} > na\right) dv = s^2(m^{-1}(a)) + o(1).$$

The third term is handled similarly due to the fact that the $O(1/n)$ consists in a sum of powers of v.

For $E_{P_{nA}}\mathbf{U}_1\mathbf{U}_2 = a^2 + o(1)$, the proof is similar. ■

Lemma 12 yields the maximal inequality stated in Lemma 22 of [Broniatowski and Caron 2011] under the condition $(\mathbf{U}_{1,n} > na)$. We also need the order of magnitude of the maximum of $(|\mathbf{U}_1|, ..., |\mathbf{U}_k|)$ under P_{nA} which is stated in the following result.

Lemma 13 *It holds for all k between 1 and n*

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

Proof. Using the same argument as in Lemma 23 of [Broniatowski and Caron 2011], we consider the case when the r.v's $\mathbf{U}_i$ take non negative values. We prove that

$$\lim_{n\to\infty} P_{nA}(\max(\mathbf{U}_1, ..., \mathbf{U}_k) > t_n) = 0$$

when

$$\lim_{n\to\infty} \frac{t_n}{\log n} = \infty.$$

It holds

$$\begin{aligned} P_{nA}(\max(\mathbf{U}_1, ..., \mathbf{U}_k) > t_n) &= \int_a^{a+c} P_{nv}(\max(\mathbf{U}_1, ..., \mathbf{U}_k) > t_n | \mathbf{U}_{1,n}/n = v) \\ &\quad p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} > na)\, dv \\ &\quad + \int_{a+c}^{\infty} P_{nv}(\max(\mathbf{U}_1, ..., \mathbf{U}_k) > t_n | \mathbf{U}_{1,n}/n = v) \\ &\quad p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} > na)\, dv \\ &= : I + II. \end{aligned}$$

Now, using the same arguments as before,

$$II \le \frac{P(\mathbf{U}_{1,n}/n > a + c)}{P(\mathbf{U}_{1,n}/n > a)} \le \frac{m^{-1}(a)s(m^{-1}(a))}{m^{-1}(a+c)s(m^{-1}(a+c))} \exp\left(-ncm^{-1}(a)\right)$$

Since c is fixed and $m^{-1}(a)$ is bounded , $II \to 0$ under (C).

Furthermore by Lemma 23 of [Broniatowski and Caron 2011],

$$\lim_{n\to\infty} P(\max(\mathbf{U}_1, ..., \mathbf{U}_n) > t_n | \mathbf{U}_{1,n}/n = v) =: \lim_{n\to\infty} r_n = 0$$

when $v \in (a, a + c)$. Hence

$$I \le r_n(1 + o(1)) \to 0.$$

This proves the Lemma. ■

We now prove Theorem 2(i).

*Step 1.*We first prove that the integral (19) can be reduced to its principal part, namely that

$$p_{nA}(Y_1^k) = (1 + o_{P_{nA}}(1))$$

$$\int_a^{a+c} p(\mathbf{X}_1^k = Y_1^k \big| \mathbf{U}_{1,n}/n = v) p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} > na) dv \tag{37}$$

holds for any fixed $c > 0$.

Apply Bayes formula to obtain

$$p_{nA}(Y_1^k) = \frac{np_{\mathbf{X}}(Y_1^k)}{(n-k)} \frac{\int_a^{\infty} p\left(\frac{\mathbf{U}_{k+1,n}}{n-k} = \frac{n}{n-k}\left(t - \frac{k\overline{U_{1,k}}}{n}\right)\right) dt}{P(\mathbf{U}_{1,n} > na)}$$

where $\overline{U_{1,k}} := \frac{U_{1,k}}{k}$.

Denote

$$I := \frac{P\left(\frac{\mathbf{U}_{k+1,n}}{n-k} > m_k + \frac{nc}{n-k}\right)}{P\left(\frac{\mathbf{U}_{k+1,n}}{n-k} > m_k\right)}.$$

with

$$m_k = \frac{n}{n-k}\left(a - \frac{k\overline{U_{1,k}}}{n}\right).$$

Then (37) holds whenever $I \to 0$ (under P_{nA}).

Under P_{nA} it holds

$$\overline{U_{1,n}} = a + O_{P_{nA}}\left(\frac{1}{nm^{-1}(a)}\right).$$

A similar result as Lemma 22 holds under condition $(\mathbf{U}_{1,n} > na)$, using Lemma 21; namely it holds

$$\max_{0\le i\le k-1} \left|\overline{U_{i+1,n}}\right| = a + o_{P_{nA}}(\epsilon_n).$$

Using both results, it holds

$$m_k = a + O_{P_{nA}}(v_n) \tag{38}$$

with $v_n = \max\left(\epsilon_n, \frac{1}{(n-k)m^{-1}(a)}\right)$ which tends to 0 under (C).

We now prove that $I \to 0$. Using once more Lemma 11 yields

$$I \le \frac{m^{-1}(m_k)s(m^{-1}(m_k))}{m^{-1}(m_k + \frac{nc}{n-k})s(m^{-1}(m_k) + \frac{nc}{n-k})}$$

$$\exp\left(-(n-k)\left(I_{\mathbf{U}}\left(m_k + \frac{nc}{n-k}\right) - I_{\mathbf{U}}(m_k)\right)\right).$$

Now by convexity of the function $I_{\mathbf{U}}$, and (38),

$$\begin{aligned} & \exp -(n-k)\left(I_{\mathbf{U}}\left(m_k + \frac{nc}{n-k}\right) - I_{\mathbf{U}}(m_k)\right) \\ \le\ & \exp -ncm^{-1}(m_k) = \exp -nc\left[m^{-1}(a) + \frac{1}{V(a+\theta O_{P_{nA}}(v_n))}O_{P_{nA}}(v_n)\right] \end{aligned}$$

for some θ in $(0,1)$. which tends to 0 under P_{nA} when (A) and (C) hold. By monotonicity of $t \to m(t)$ and condition (C) the ratio in I is bounded.

We have proved that

$$I = O_{P_{nA}}\left(\exp -ncm^{-1}(a)\right).$$

Step 2. Theorem (2)(i) holds uniformly in v in $(a, a+c)$ where Y_1^k is generated under P_{nA}. This result follows from a similar argument as used in Theorem 1 where (22) is proved under the local sampling P_{nv}. A close look at the proof shows that (22) holds whenever Lemmas 22 and 23, stated in [Broniatowski and Caron 2011] for the variables $\mathbf{U}_i$'s instead of $\mathbf{X}_i$'s hold under P_{nA}. Those lemmas are substituted by Lemmas 12 and 13 here above.

Inserting (22) in (37) yields

$$\begin{aligned} p_{nA}(Y_1^k) = & \left(\int_a^{a+c} g_{nv}(Y_1^k) p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} > na) dv\right) \\ & \left(1 + o_{p_{nA}}\left(\max\left(\epsilon_n (\log n)^2, \left(\exp\left(-ncm^{-1}(a)\right)\right)^{\delta}\right)\right)\right). \end{aligned}$$

dor any positive $\delta < 1$.

The conditional density of $\mathbf{U}_{1,n}/n$ given $(\mathbf{U}_{1,n} > na)$ is given in (30) which holds uniformly in v on $(a, a+c)$.

Summing up we have proved

$$p_{nA}(Y_1^k) =$$

$$\begin{aligned} & \left(nm^{-1}(a)\int_a^{a+c} g_{nv}(Y_1^k)\left(\exp -nm^{-1}(a)(v-a)\right)dv\right) \\ & \left(1 + o_{p_{nA}}\left(\max\left(\epsilon_n (\log n)^2, \left(\exp\left(-ncm^{-1}(a)\right)\right)^{\delta}\right)\right)\right) \end{aligned}$$

as $n \to \infty$ for any positive δ.

In order to get the approximation of p_{nA} by the density g_{nA} it is enough to observe that

$$nm^{-1}(a)\int_a^{a+c} g_{nv}(Y_1^k)\left(\exp -nm^{-1}(a)(v-a)\right)dv$$

$$= 1 + o_{P_{nA}}\left(\exp -ncm^{-1}(a)\right)$$

as $n \to \infty$ which completes the proof of (22). The proof of (23) follows from (22) and Lemma 14 cited hereunder.

The following Lemma proves that approximating p_{nA} by g_{nA} under p_{nA} is similar to approximating p_{nA} by g_{nA} under g_{nA}.

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

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

as n tends to ∞. Then

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

Proof. Denote

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

It holds for all positive δ

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

where

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

Since

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

it follows that

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

which proves the claim. ■

7.2 Proof of Lemma 9

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

$$\left|\frac{p_{nA}(y_1^k)}{g_{nA}\left(y_1^k\right)} - 1\right| < \delta_n$$

with δ_n as in (24) 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 $u_{1,k} := \sum_{i=1}^k u(y_i)$ and α_n satisfies

$$\lim_{n\to\infty} \alpha_n = 0 \tag{39}$$

together with

$$\lim_{n\to\infty} \alpha_n a\sqrt{n-k} = \infty. \tag{40}$$

We prove that

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

Let

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

with

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

By the above definition

$$\lim_{n\to\infty} P_{nA}\left(A_{n,\varepsilon_n}\right) = 1. \tag{41}$$

Note also that

$$\begin{aligned} G_{nA}\left(A_{n,\varepsilon_n}\right) &:= \int \mathbb{1}_{A_{n,\varepsilon_n}}(x_1^n) g_{nA}\left(x_1^n\right) dx_1^n \\ &= \int \mathbb{1}_{A^k_{\varepsilon_n}}(x_1^k) g_{nA}\left(x_1^k\right) dx_1^n \\ &\geq \frac{1}{1+\delta_n} \int \mathbb{1}_{A^k_{\varepsilon_n}}(x_1^k) p_{nA}(x_1^k) dx_1^k \\ &= \frac{1}{1+\delta_n}\left(1+o(1)\right) \end{aligned}$$

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_{nA}.

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_{nv} uniformly upon v in $(a, a+c)$ and integrate out with respect to the distribution of $\mathbf{U}_{1,n}/n$ conditionally on $\mathbf{U}_{1,n}/n \in (a, a+c)$.

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

$$t^k s(t^k) = a\left(1+o(1)\right) \tag{42}$$

holds.

By Lemma 22 in [Broniatowski and Caron 2011] and integrating w.r.t. p_{nv} on $(a, a+c)$ it holds, under (39) and (40)

$$\lim_{n\to\infty} P_{nA}\left(B_n\right) = 1. \tag{43}$$

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

$$\left| \frac{t^k}{a} - 1 \right| < \delta_n'. \tag{44}$$

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 (44).

Further (44) 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 (42) holds.

Define

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

Since

$$\int \mathbb{1}_{C_n}(x_1^n) g_{nA}\left(x_1^k\right) dx_1^n \geq \frac{1}{1+\delta_n} \int \mathbb{1}_{C_n} p_{nA}(x_1^n) dx_1^n$$

and by (41) and (43)

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

we obtain

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

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

References

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

[Barbe and Broniatowski 1999] BARBE, P., BRONIATOWSKI, M. 1992. Simulation in exponential families. *ACM Transactions on Modeling and Computer Simulation (TOMACS) 9(3)*, 203-223.

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

[Barndorff-Nielsen 1978] BARNDORFF-NIELSEN, O. 1978. *Information and Exponential Families in Statistical Theory.* Wiley, New-York (1978)

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

[Broniatowski and Caron 2011] BRONIATOWSKI, M., CARON, V. 2010. Long runs under a conditional limit distribution. arxiv:1010.3616,2010.

[Broniatowski and Ritov 2009] BRONIATOWSKI, M., RITOV, Y. 2009 Importance sampling for rare events and conditioned random walks. arxiv:0910.1819,2009.

[Bucklew 2004] BUCKLEW, J.A. 2004. *Introduction to rare event simulation.* Springer Series in Statistics, Springer-Verlag, New York.

[Dembo and Zeitouni (1996)] DEMBO, A. and ZEITOUNI, O. 1996. Refinements of the Gibbs conditioning principle. *Probab. Theory Related Fields* **104** 1–14.

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

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

[Glynn and Whitt 1992] GLYNN, P.W., WHITT, W. 1992. The asymptotic efficiency of simulation estimators. *Oper. Res., 40(3),* 505-520.

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

[Jensen 1995] JENSEN, J.L. 1995. *Saddlepoint Approximations. Oxford Statistical Science Series, vol 16.* The Clarendon Press Oxford University Press, New York, Oxford Science Publications.

[Letac and Mora (1990)] LETAC, G., MORA, M.(1990) Natural real exponential families with cubic variance functions. Ann. Statist. 18, no. 1, 1–37.

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

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