The Annals of Applied Probability
2014, Vol. 24, No. 6, 2246–2296
DOI: 10.1214/13-AAP975

LONG RUNS UNDER A CONDITIONAL LIMIT DISTRIBUTION

BY MICHEL BRONIATOWSKI AND VIRGILE CARON

Université Pierre Et Marie Curie, LSTA

This paper presents a sharp approximation of the density of long runs of a random walk conditioned on its end value or by an average of a function of its summands as their number tends to infinity. In the large deviation range of the conditioning event it extends the Gibbs conditional principle in the sense that it provides a description of the distribution of the random walk on long subsequences. An approximation of the density of the runs is also obtained when the conditioning event states that the end value of the random walk belongs to a thin or a thick set with a nonempty interior. The approximations hold either in probability under the conditional distribution of the random walk, or in total variation norm between measures. An application of the approximation scheme to the evaluation of rare event probabilities through importance sampling is provided. When the conditioning event is in the range of the central limit theorem, it provides a tool for statistical inference in the sense that it produces an effective way to implement the Rao–Blackwell theorem for the improvement of estimators; it also leads to conditional inference procedures in models with nuisance parameters. An algorithm for the simulation of such long runs is presented, together with an algorithm determining the maximal length for which the approximation is valid up to a prescribed accuracy.

1. Context and scope. This paper explores the asymptotic distribution of a random walk conditioned on its final value as the number of summands increases. Denote $\mathbf{X}_1^n := (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ a set of n independent copies of a real random variable $\mathbf{X}$ with density $p_{\mathbf{X}}$ on $\mathbb{R}$ and $\mathbf{S}_{1,n} := \mathbf{X}_1 + \cdots + \mathbf{X}_n$. We consider approximations of the density of the vector $\mathbf{X}_1^k = (\mathbf{X}_1, \ldots, \mathbf{X}_k)$ on $\mathbb{R}^k$ when $\mathbf{S}_{1,n} = na_n$, and a_n is a convergent sequence. The integer valued sequence $k := k_n$ is such that

$$0 \le \lim \sup_{n\to\infty} k/n \le 1 \tag{K1}$$

together with

$$\lim_{n\to\infty} n - k = \infty. \tag{K2}$$

Therefore we may consider the asymptotic behavior of the density of the trajectory of the random walk on long runs. For the sake of applications we also address the case when $\mathbf{S}_{1,n}$ is substituted by $\mathbf{U}_{1,n} := u(\mathbf{X}_1) + \cdots + u(\mathbf{X}_1)$ for some real

Received February 2012; revised September 2013.
MSC2010 subject classifications. Primary 60B10; secondary 65C50.
Key words and phrases. Gibbs principle, conditioned random walk, large deviation, moderate deviation, simulation, importance sampling, Rao–Blackwell theorem.

valued measurable function u, and when the conditioning event is $(\mathbf{U}_{1,n} = u_{1,n})$ where $u_{1,n}/n$ converges as n tends to infinity. A complementary result provides an estimation for the case when the conditioning event is a large set in the large deviation range, $(\mathbf{U}_{1,n} \in nA)$ where A is a Borel set with nonempty interior with $Eu(\mathbf{X}) < \texttt{essinf}A$; two cases are considered, according to the local dimension of A at its essential infimum point $\texttt{essinf}A$.

The interest in this question stems from various sources. When k is fixed (typically $k = 1$) this is a version of the *Gibbs conditional principle* which has been studied extensively for fixed $a_n \neq E\mathbf{X}$, therefore under a *large deviation* condition. Diaconis and Freedman [13] have considered this issue also in the case $k/n \to \theta$ for $0 \leq \theta < 1$, in connection with de Finetti's theorem for exchangeable finite sequences. Their interest was related to the approximation of the density of $\mathbf{X}_1^k$ by the *product density* of the summands $\mathbf{X}_i$'s, and therefore on the validity of the independence of the $\mathbf{X}_i$'s under conditioning. Their result is in the spirit of Van Camperhout and Cover [22], and parallels can be drawn with Csiszár's [10] asymptotic conditional independence result, when the conditioning event is $(\mathbf{S}_{1,n} > na_n)$ with a_n fixed and larger than $E\mathbf{X}$. In the same vein and under the same *large deviation* condition Dembo and Zeitouni [11] considered similar problems. This question is also of importance in statistical physics. Numerous papers pertaining to structural properties of polymers deal with this issue, and we refer to [12] and [23] for a description of those problems and related results. In the moderate deviation case, Ermakov [15] also considered a similar problem when $k = 1$.

The approximation of conditional densities is the basic ingredient for the numerical estimation of integrals through improved Monte Carlo techniques. Rare event probabilities may be evaluated through importance sampling techniques; efficient sampling schemes consist of the simulation of random variables under a proxy of a conditional density, often with respect to conditioning events of the form $(\mathbf{U}_{1,n} > na_n)$; optimizing these schemes has been a motivation for this work.

In parametric statistical inference, conditioning on the observed value of a statistic leads to a reduction of the mean square error of some estimate of the parameter; the famous Rao–Blackwell and Lehmann–Scheffé theorems can be implemented when a simulation technique produces samples according to the distribution of the data conditioned on the value of some observed statistics. In these applications the conditioning event is local, and when the statistic is of the form $\mathbf{U}_{1,n}$, then the observed value $u_{1,n}$ satisfies $\lim_{n\to\infty} u_{1,n}/n = Eu(\mathbf{X})$. Such is the case in exponential families when $\mathbf{U}_{1,n}$ is a sufficient statistic for the parameter. Other fields of applications pertain to parametric estimation where conditioning by the observed value of a sufficient statistic for a nuisance parameter produces optimal inference via maximum likelihood in the conditioned model. In general this conditional density is unknown; the approximation produced in this paper provides a tool for the solution of these problems.

For both importance sampling and for the improvement of estimators, the approximation of the conditional density of $\mathbf{X}_1^k$ on long runs should be of a special form: it has to be a density on $\mathbb{R}^k$, easy to simulate, and the approximation should be sharp. For these applications the relative error of the approximation should be small on the simulated paths only. Also for inference via maximum likelihood under nuisance parameters the approximation has to be accurate on the sample itself and not on the entire space.

Our first set of results provides a very sharp approximation scheme; numerical evidence on exponential runs with length $n = 1000$ provide a *relative error* of the approximation of order less than 100% for the density of the first 800 terms when evaluated on the sample paths themselves, thus on the significant part of the support of the conditional density; this very sharp approximation rate is surprising in such a large dimensional space, and it illustrates the fact that the conditioned measure occupies a very small part of the entire space. Therefore the approximation of the density of $\mathbf{X}_1^k$ is not performed on the sequence of entire spaces $\mathbb{R}^k$, but merely on a sequence of subsets of $\mathbb{R}^k$ which contain the trajectories of the conditioned random walk with probability going to 1 as n tends to infinity; the approximation is performed on *typical paths*.

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 [13] and Dembo and Zeitouni [11] 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 previously 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 $\mathbb{R}^k$.

It would have been of interest to consider sharper convergence criteria than the total variation distance; the χ^2-distance, which is the mean square relative error, cannot be bounded through our approach on the entire space $\mathbb{R}^k$, since it is only suitable for large sets of trajectories (whose probability goes to 1 as n increases); this is not sufficient to bound its expected value under the conditional sampling.

This paper is organized as follows. Section 2 presents the approximation scheme for the conditional density of $\mathbf{X}_1^k$ under the conditioning point sequence $(\mathbf{S}_{1,n} = na_n)$. In Section 3, it is extended to the case when the conditioning family of events is written as $(\mathbf{U}_{1,n} = u_{1,n})$. The value of k for which this approximation is appropriate is discussed; an algorithm for the implementation of this rule is proposed. Algorithms for the simulation of random variables under the approximating scheme are also presented. Section 4 extends the results of Section 3 when conditioning on large sets. Two applications are presented in Section 5; the first one pertains to Rao–Blackwellization of estimators, hence on the application of the results of Section 3 when the conditioning point is such that $\lim_{n\to\infty} u_{1,n}/n = Eu(\mathbf{X})$; in the second application the result of Section 4 is used to derive small variance estimators of rare event probabilities through importance sampling; in this case the conditioning event is in the range of the large deviation scale.

The main steps of the proofs are in the core of the paper; some of the technicalities are left to the Appendix.

2. Random walks conditioned on their sum.

2.1. *Notation and hypothesis.* In this section the conditioning point event is written as

$$\mathcal{E}_n := (\mathbf{S}_{1,n} = na_n).$$

We assume that $\mathbf{X}$ satisfies the Cramér condition; that is, $\mathbf{X}$ has a finite moment generating function $\Phi(t) := E[\exp(t\mathbf{X})]$ in a nonempty neighborhood of 0. Denote

$$m(t) := \frac{d}{dt}\log\Phi(t),$$

$$s^2(t) := \frac{d}{dt}m(t),$$

$$\mu_3(t) := \frac{d}{dt}s^2(t).$$

The values of $m(t)$, s^2 and $\mu_3(t)$ are the expectation, the variance and the kurtosis of the *tilted* density

$$\pi^\alpha(x) := \frac{\exp(tx)}{\Phi(t)}p(x), \tag{1}$$

where t is the unique solution of the equation $m(t) = \alpha$ when α belongs to the support of $\mathbf{X}$. Conditions on $\Phi(t)$ which ensure existence and uniqueness of t are referred to as *steepness properties;* we refer to [4], page 153 ff., for all properties of moment generating functions used in this paper. Denote Π^α the probability measure with density π^α.

We also assume that the characteristic function of $\mathbf{X}$ is in L^r for some $r \geq 1$ which is necessary for the Edgeworth expansions to be performed.

The probability measure of the random vector $\mathbf{X}_1^n$ on $\mathbb{R}^n$ conditioned upon $\mathcal{E}_n$ is denoted P_{na_n}. We also denote P_{na_n} the corresponding distribution of $\mathbf{X}_1^k$ conditioned upon $\mathcal{E}_n$; the vector $\mathbf{X}_1^k$ then has a density with respect to the Lebesgue measure on $\mathbb{R}^k$ for $1 \leq k < n$, which will be denoted p_{na_n}. For a general r.v. $\mathbf{Z}$ with density p, $p(\mathbf{Z} = z)$ denotes the value of p at point z. Hence, $p_{na_n}(x_1^k) = p(\mathbf{X}_1^k = x_1^k | \mathbf{S}_{1,n} = na_n)$. The normal density function on $\mathbb{R}$ with mean μ and variance τ at x is denoted $\mathfrak{n}(\mu, \tau, x)$. When $\mu = 0$ and $\tau = 1$, the standard notation $\mathfrak{n}(x)$ is used.

2.2. *A first approximation result.* We first put forward a simple result which provides an approximation of the density p_{na_n} of the measure P_{na_n} on $\mathbb{R}^k$ when k satisfies (K1) and (K2). For $i \leq j$ denote

$$s_{i,j} := x_i + \cdots + x_j.$$

Denote $a := a_n$ omitting the index n for clarity.

We make use of the following property which states the invariance of conditional densities under tilting: For $1 \le i \le j \le n$, for all a in the range of $\mathbf{X}$, for all u and s

$$p(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s) = \pi^a(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s), \tag{2}$$

where $\mathbf{S}_{i,j} := \mathbf{X}_i + \cdots + \mathbf{X}_j$ together with $\mathbf{S}_{1,0} = s_{1,0} = 0$. By the Bayes formula it holds that

$$\begin{aligned} p_{na}(x_1^k) &= \prod_{i=0}^{k-1} p(\mathbf{X}_{i+1} = x_{i+1} | \mathbf{S}_{i+1,n} = na - s_{1,i}) \qquad (3)\\ &= \prod_{i=0}^{k-1} \pi^a(\mathbf{X}_{i+1} = x_{i+1}) \frac{\pi^a(\mathbf{S}_{i+2,n} = na - s_{1,i+1})}{\pi^a(\mathbf{S}_{i+1,n} = na - s_{1,i})} \\ &= \left[\prod_{i=0}^{k-1} \pi^a(\mathbf{X}_{i+1} = x_{i+1})\right] \frac{\pi^a(\mathbf{S}_{k+1,n} = na - s_{1,k})}{\pi^a(\mathbf{S}_{1,n} = na)}. \qquad (4) \end{aligned}$$

Denote $\overline{\mathbf{S}_{k+1,n}}$ and $\overline{\mathbf{S}_{1,n}}$ the normalized versions of $\mathbf{S}_{k+1,n}$ and $\mathbf{S}_{1,n}$ under the sampling distribution Π^a. By (4)

$$\begin{aligned} p_{na}(x_1^k) &= \left[\prod_{i=0}^{k-1} \pi^a(\mathbf{X}_{i+1} = x_{i+1})\right] \\ &\quad \times \frac{\sqrt{n}}{\sqrt{n-k}} \frac{\pi^a(\overline{\mathbf{S}_{k+1,n}} = (ka - s_{1,k})/(s_a\sqrt{n-k}))}{\pi^a(\overline{\mathbf{S}_{1,n}} = 0)}. \end{aligned}$$

A first order Edgeworth expansion is performed in both terms of the ratio in the above display; see Remark 5 below. This yields, assuming (K1) and (K2), the following:

PROPOSITION 1. *For all x_1^k in $\mathbb{R}^k$*

$$\begin{aligned} p_{na}(x_1^k) &= \left[\prod_{i=0}^{k-1} \pi^a(\mathbf{X}_{i+1} = x_{i+1})\right] \\ &\quad \times \left[\frac{\mathfrak{n}((ka - s_{1,k})/(s(t^a)\sqrt{n-k}))}{\mathfrak{n}(0)} \sqrt{\frac{n}{n-k}} \right. \qquad (5)\\ &\quad \left. \times \left(1 + \frac{\mu_3(t^a)}{6s^3(t^a)\sqrt{n-k}} H_3\left(\frac{ka - s_{1,k}}{s(t^a)\sqrt{n-k}}\right)\right) + O\left(\frac{1}{\sqrt{n}}\right)\right], \end{aligned}$$

where $H_3(x) := x^3 - 3x$. The value of t^a is defined through $m(t^a) = a$.

Despite its appealing aspect, (5) is of poor value for applications, since it does not yield an explicit way to simulate samples under a proxy of p_{na} for large values of k. The other way is to construct the approximation of p_{na} step by step, approximating the terms in (3) one by one and using the invariance under the tilting at each step, which introduces a product of different tilted densities in (4). This method produces a valid approximation of p_{na} on subsets of $\mathbb{R}^k$ which contain the trajectories of the conditioning random walk with larger and larger probability, going to 1 as n tends to infinity.

This introduces the main focus of this paper.

2.3. *A recursive approximation scheme.* We introduce a positive sequence ε_n which satisfies

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

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

It will be shown that $\varepsilon_n (\log n)^2$ is the rate of accuracy of the approximating scheme.

We denote a the generic term of the convergent sequence $(a_n)_{n\geq 1}$. For clarity the dependence on n of all quantities involved in the subsequent development is omitted in the notation.

2.3.1. *Approximation of the density of the runs.* Define a density $g_{na}(y_1^k)$ on $\mathbb{R}^k$ as follows. Set

$$g_0(y_1|y_0) := \pi^a(y_1)$$

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

Set t_i to be the unique solution of the equation

$$m_i := m(t_i) = \frac{n}{n-i}\left(a - \frac{s_{1,i}}{n}\right), \tag{6}$$

where $s_{1,i} := y_1 + \cdots + y_i$. The tilted adaptive family of densities π^{m_i} is the basic ingredient of the derivation of approximating scheme. Let

$$s_i^2 := \frac{d^2}{dt^2}\big(\log E_{\pi^{m_i}} \exp(t\mathbf{X})\big)(0)$$

and

$$\mu_j^i := \frac{d^j}{dt^j}\big(\log E_{\pi^{m_i}} \exp(t\mathbf{X})\big)(0), \qquad j = 3, 4,$$

which are the second, third and fourth cumulants of π^{m_i}. Let

$$g\big(y_{i+1}|y_1^i\big) = C_i p_{\mathbf{X}}(y_{i+1}) \mathfrak{n}(\alpha\beta + a, \beta, y_{i+1}) \tag{7}$$

be a density where

$$\alpha = t_i + \frac{\mu_3^i}{2s_i^2(n-i-1)}, \tag{8}$$

$$\beta = s_i^2(n-i-1) \tag{9}$$

and C_i is a normalizing constant.

Define

$$g_{na}(y_1^k) := g_0(y_1|y_0) \prod_{i=1}^{k-1} g(y_{i+1}|y_1^i). \tag{10}$$

We then have:

THEOREM 2. *Assume* (K1) *and* (K2) *together with* (E1) *and* (E2). *Let* Y_1^n *be a sample from density* p_{na}. *Then*

$$\begin{aligned} p_{na}(Y_1^k) &:= p(\mathbf{X}_1^k = Y_1^k | \mathbf{S}_{1,n} = na) \\ &= g_{na}(Y_1^k)\left(1 + o_{P_{na}}(\varepsilon_n (\log n)^2)\right). \end{aligned} \tag{11}$$

PROOF. The proof uses Bayes's formula to write $p(\mathbf{X}_1^k = Y_1^k|\mathbf{S}_{1,n} = na)$ as a product of k conditional densities of the individual terms of the trajectory evaluated at Y_1^k. Each term of this product is approximated by an Edgeworth expansion which together with the properties of Y_1^k under P_{na} completes the proof. This proof is rather long, and we have deferred its technical steps to the Appendix.

Denote $S_{1,0} = 0$ and $S_{1,i} := S_{1,i-1} + Y_i$. It holds that

$$\begin{aligned} p(\mathbf{X}_1^k = Y_1^k|\mathbf{S}_{1,n} = na) &= p(\mathbf{X}_1 = Y_1|\mathbf{S}_{1,n} = na), \\ \prod_{i=1}^{k-1} p(\mathbf{X}_{i+1} &= Y_{i+1}|\mathbf{X}_1^i = Y_1^i, \mathbf{S}_{1,n} = na) \\ &= \prod_{i=0}^{k-1} p(\mathbf{X}_{i+1} = Y_{i+1}|\mathbf{S}_{i+1,n} = na - S_{1,i}) \end{aligned} \tag{12}$$

by independence of the r.v.'s $\mathbf{X}_i$'s.

Define t_i through

$$m(t_i) = \frac{n}{n-i}\left(a - \frac{S_{1,i}}{n}\right)$$

a function of the past r.v.'s Y_1^i, and set $m_i := m(t_i)$ and $s_i^2 := s^2(t_i)$. By (2)

$$\begin{aligned} p(\mathbf{X}_{i+1} &= Y_{i+1}|\mathbf{S}_{i+1,n} = na - S_{1,i}) \\ &= \pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1}|\mathbf{S}_{i+1}^n = na - S_{1,i}) \\ &= \pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1}) \frac{\pi^{m_i}(\mathbf{S}_{i+2,n} = na - S_{1,i+1})}{\pi^{m_i}(\mathbf{S}_{i+1,n} = na - S_{1,i})}, \end{aligned}$$

where we used the independence of the $\mathbf{X}_j$'s under π^{m_i}. A precise evaluation of the dominating terms in this latest expression is needed in order to handle the product (12).

Under the sequence of densities π^{m_i} the i.i.d. r.v.'s $\mathbf{X}_{i+1}, \ldots, \mathbf{X}_n$ define a triangular array which satisfies a local central limit theorem, and an Edgeworth expansion. Under π^{m_i}, $\mathbf{X}_{i+1}$ has expectation m_i and variance s_i^2. Center and normalize both the numerator and denominator in the fraction which appear in the last display. Denote $\overline{\pi_{n-i-1}}$ the density of the normalized sum $(\mathbf{S}_{i+2,n} - (n-i-1)m_i)/(s_i\sqrt{n-i-1})$ when the summands are i.i.d. with common density π^{m_i}. Accordingly $\overline{\pi_{n-i}}$ is the density of the normalized sum $(\mathbf{S}_{i+1,n} - (n-i)m_i)/(s_i\sqrt{n-i})$ under i.i.d. π^{m_i} sampling. Hence, evaluating both $\overline{\pi_{n-i-1}}$ and its normal approximation at point Y_{i+1},

$$
\begin{aligned}
p(&\mathbf{X}_{i+1} = Y_{i+1} | \mathbf{S}_{i+1,n} = na - S_{1,i}) \\
&= \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1}) \frac{\overline{\pi_{n-i-1}}((m_i - Y_{i+1})/s_i\sqrt{n-i-1})}{\overline{\pi_{n-i}}(0)} \qquad (13)\\
&:= \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1}) \frac{N_i}{D_i}.
\end{aligned}
$$

The sequence of densities $\overline{\pi_{n-i-1}}$ converges pointwise to the standard normal density under (E1) which implies that $n-i$ tends to infinity for all $1 \le i \le k$, and an Edgeworth expansion to order 5 is performed for the numerator and the denominator. The main arguments used in order to obtain the order of magnitude of the involved quantities are (i) a maximal inequality which controls the magnitude of m_i for all i between 0 and $k-1$ (Lemma 22), (ii) the order of the maximum of the Y_i's (Lemma 23). As proved in the Appendix,

$$
N_i = \mathfrak{n}(-Y_{i+1}/s_i\sqrt{n-i-1}) \cdot A \cdot B + O_{P_{na}}\left(\frac{1}{(n-i-1)^{3/2}}\right), \qquad (14)
$$

where

$$
A := \left(1 + \frac{aY_{i+1}}{s_i^2(n-i-1)} - \frac{a^2}{2s_i^2(n-i-1)} + \frac{o_{P_{na}}(\varepsilon_n \log n)}{n-i-1}\right) \qquad (15)
$$

and

$$
B := \begin{pmatrix} 1 - \dfrac{\mu_3^i}{2s_i^4(n-i-1)}(a - Y_{i+1}) \\ -\dfrac{\mu_3^i - s_i^4}{8s_i^4(n-i-1)} - \dfrac{15(\mu_3^i)^2}{72s_i^6(n-i-1)} + \dfrac{O_{P_{na}}((\log n)^2)}{(n-i-1)^2} \end{pmatrix}. \qquad (16)
$$

The $O_{P_{na}}(\frac{1}{(n-i-1)^{3/2}})$ term in (14) is uniform on $(m_i - Y_{i+1})/s_i\sqrt{n-i-1}$. Turn back to (13) and perform the same Edgeworth expansion in the denominator, which is written as

$$D_i = \mathfrak{n}(0)\left(1 - \frac{\mu_4^i - 3s_i^4}{8s_i^4(n-i)} - \frac{15(\mu_3^i)^2}{72s_i^6(n-i)}\right) + O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right). \tag{17}$$

The terms in $g(Y_{i+1}|Y_1^i)$ follow from an expansion in the ratio of the two expressions (14) and (17) above. The Gaussian contribution is explicit in (14) while the term $\exp(\frac{\mu_3^i}{2s_i^4(n-i-1)}Y_{i+1})$ is the dominant term in B. Turning to (13) and comparing with (11) it appears that the normalizing factor C_i in $g(Y_{i+1}|Y_1^i)$ compensates the term $\frac{\sqrt{n-i}}{\Phi(t_i)\sqrt{n-i-1}}\exp(\frac{-a\mu_3^i}{2s_i^2(n-i-1)})$, where the term $\Phi(t_i)$ comes from $\pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1})$. Furthermore the product of the remaining terms in the above approximations in (14) and (17) form the $1 + o_{P_{na}}(\varepsilon_n(\log n)^2)$ approximation rate, as claimed. Details are deferred to the Appendix. This yields

$$p(\mathbf{X}_1^k = Y_1^k|\mathbf{S}_{1,n} = na) = (1 + o_{P_{na}}(\varepsilon_n(\log n)^2))g_0(Y_1|Y_0)\prod_{i=1}^{k-1} g(Y_{i+1}|Y_1^i),$$

which completes the proof of the theorem. □

That the variation distance between P_{na_n} and G_{na_n} tends to 0 as $n \to \infty$ is stated in Section 3.

REMARK 3. When the $\mathbf{X}_i$'s are i.i.d. with a standard normal density, then the result in the above approximation theorem holds with $k = n-1$ implying that $p(\mathbf{X}_1^{n-1} = x_1^{n-1}|\mathbf{S}_{1,n} = na) = g_{na}(x_1^{n-1})$ for all x_1^{n-1} in $\mathbb{R}^{n-1}$. This extends to the case when they have an infinitely divisible distribution. However, formula (11) holds true without the error term only in the Gaussian case. Similar exact formulas can be obtained for infinitely divisible distributions using (12) where no use of tilting is made. Such formulas are used to produce Figures 1, 2, 3 and 4 in order to assess the validity of the selection rule for k in the exponential case.

REMARK 4. The density in (7) is a slight modification of π^{m_i}. The modification from $\pi^{m_i}(y_{i+1})$ to $g(y_{i+1}|y_1^i)$ is a small shift in the location parameter depending both on a and on the skewness of p, and a change in the variance: large values of $\mathbf{X}_{i+1}$ have smaller weight for large i, so that the distribution of $\mathbf{X}_{i+1}$ tends to concentrate around m_i as i approaches k.

REMARK 5. In Theorem 2, as in Proposition 1, Theorem 8 or Lemma 23, we use an Edgeworth expansion for the density of the normalized sum of the $(n-i)$th row of some triangular array of row-wise independent r.v.'s with a common density. Consider the i.i.d. r.v.'s $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with common density $\pi^a(x)$ where a

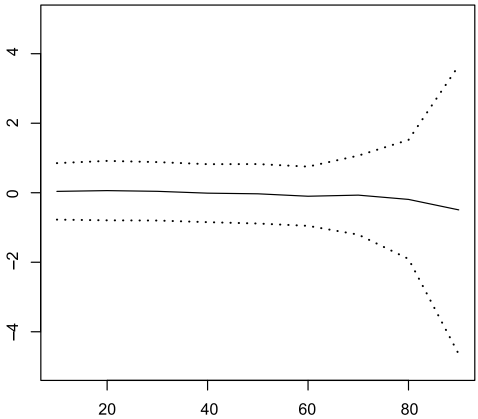

FIG. 1. $\overline{\text{ERE}}(k)$ (*solid line*) *along with upper and lower bound of* $\overline{\text{CI}}(k)$ (*dotted line*) *as a function of* k *with* $n = 100$ *and* a *such that* $P_n \simeq 10^{-8}$.

may depend on n but remains bounded. The Edgeworth expansion with respect to the normalized density of $\mathbf{S}_{1,n}$ under π^a can be derived following closely the proof given, for example, in [16], page 532 ff., by substituting the cumulants of p by those of π^a. Denote $\varphi_a(z)$ the characteristic function of $\pi^a(x)$. Clearly for any $\delta > 0$ there exists $q_{a,\delta} < 1$ such that $|\varphi_a(z)| < q_{a,\delta}$ and since a is bounded, $\sup_n q_{a,\delta} < 1$. Therefore inequality (2.5) in [16], page 533 holds. With ψ_n defined as in [16], (2.6) holds with φ replaced by φ_a and σ by $s(t^a)$; (2.9) holds, which completes the proof of the Edgeworth expansion in the simple case. The proof is analogous for higher order expansions.

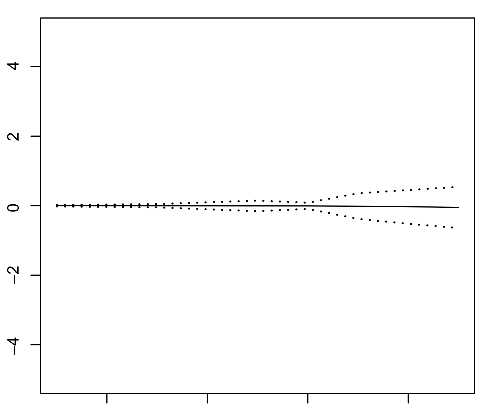

FIG. 2. $\text{ERE}(k)$ (*solid line*) *along with upper and lower bound of* $\text{CI}(k)$ (*dotted line*) *as a function of* k *with* $n = 100$ *and* a *such that* $P_n \simeq 10^{-8}$.

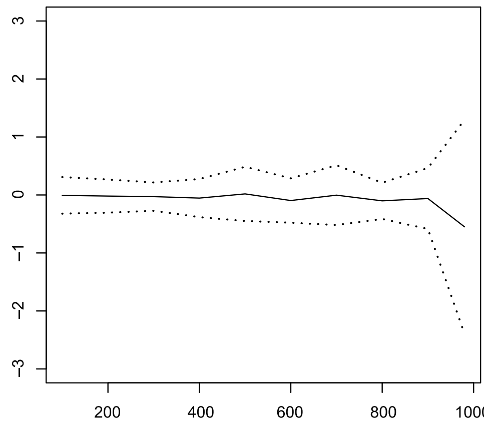

FIG. 3. $\overline{\mathrm{ERE}}(k)$ (*solid line*) *along with upper and lower bound of* $\overline{\mathrm{CI}}(k)$ (*dotted line*) *as a function of k with* $n = 1000$ *and a such that* $P_n \simeq 10^{-8}$.

2.3.2. *Sampling under the approximation*. Applications of Theorem 2 in importance sampling procedures and in Statistics require a reverse result. So assume that Y_1^k is a random vector generated under G_{na} with density g_{na}. Can we state that $g_{na}(Y_1^k)$ is a good approximation for $p_{na}(Y_1^k)$? This holds true. We state a simple lemma in this direction.

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

$$\mathfrak{r}_n(Y_1^n) = \mathfrak{s}_n(Y_1^n)\big(1 + o_{\mathfrak{R}_n}(\varepsilon_n)\big) \tag{18}$$

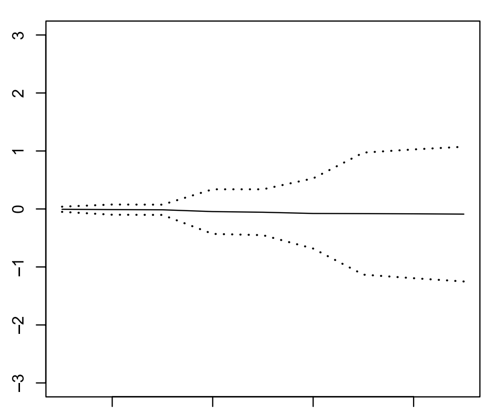

FIG. 4. ERE(k) (*solid line*) *along with upper and lower bound of* CI(k) (*dotted line*) *as a function of k with* $n = 1000$ *and a such that* $P_n \simeq 10^{-8}$.

as n tends to ∞. *Then*

$$\mathfrak{s}_n(Y_1^n) = \mathfrak{r}_n(Y_1^n)(1 + o_{\mathfrak{S}_n}(\varepsilon_n)). \tag{19}$$

PROOF. Denote

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

It holds for all positive δ,

$$\lim_{n\to\infty} \mathfrak{R}_n(A_{n,\delta\varepsilon_n}) = 1.$$

Write

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

Since

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

it follows that

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

which proves the claim. □

As a direct by-product of Theorem 2 and Lemma 6 we obtain:

THEOREM 7. *Assume* (K1) *and* (K2) *together with* (E1) *and* (E2). *Let* Y_1^k *be a sample with density* g_{na}. *It holds that*

$$p_{na}(Y_1^k) = g_{na}(Y_1^k)(1 + o_{G_{na}}(\varepsilon_n(\log n)^2)).$$

3. Random walks conditioned by a function of their summands. This section extends the above results to the case when the conditioning event is written as

$$\mathbf{U}_{1,n} := u_{1,n} \tag{20}$$

with

$$\mathbf{U}_{1,n} := u(\mathbf{X}_1) + \cdots + u(\mathbf{X}_n),$$

where the function u is real valued and the sequence $u_{1,n}/n$ converges. The characteristic function of the random variable $u(\mathbf{X})$ is assumed to belong to L^r for some $r \ge 1$. Let $p_{\mathbf{U}}$ denote the density of $\mathbf{U} = u(\mathbf{X})$ and denote $p_{\mathbf{X}}$ the density of $\mathbf{X}$.

Assume

$$\phi_{\mathbf{U}}(t) := E[\exp(t\mathbf{U})] < \infty \tag{21}$$

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

Denote

$$\pi_{\mathbf{U}}^{\alpha}(u) := \frac{\exp(tu)}{\phi_{\mathbf{U}}(t)} p_{\mathbf{U}}(u) \tag{22}$$

with $m(t) = \alpha$, and α belongs to the support of $P_{\mathbf{U}}$, the distribution of $\mathbf{U}$.

We also introduce the family of densities

$$\pi_{u}^{\alpha}(x) := \frac{\exp(tu(x))}{\phi_{\mathbf{U}}(t)} p_{\mathbf{X}}(x). \tag{23}$$

3.1. *Approximation of the density of the runs*. Assume that the sequence ε_n satisfies (E1) and (E2).

Define a density $g_{u_{1,n}}(y_1^k)$ on $\mathbb{R}^k$ as follows. Set

$$m_0 := u_{1,n}/n$$

and

$$g_0(y_1|y_0) := \pi_u^{m_0}(y_1) \tag{24}$$

with y_0 arbitrary and, for $1 \le i \le k-1$, define $g(y_{i+1}|y_1^i)$ recursively. Denote $u_{1,i} := u(y_1) + \cdots + u(y_i)$.

Set t_i to be the unique solution of the equation

$$m_i := m(t_i) = \frac{u_{1,n} - u_{1,i}}{n-i}, \tag{25}$$

and let

$$s_i^2 := \frac{d^2}{dt^2}\bigl(\log E_{\pi_{\mathbf{U}}^{m_i}} \exp(t\mathbf{U})\bigr)(0)$$

and

$$\mu_j^i := \frac{d^j}{dt^j}\bigl(\log E_{\pi_{\mathbf{U}}^{m_i}} \exp(t\mathbf{U})\bigr)(0), \qquad j = 3, 4,$$

which are the second, third and fourth cumulants of $\pi_{\mathbf{U}}^{m_i}$. A density $g(y_{i+1}|y_1^i)$ is defined as

$$g\bigl(y_{i+1}|y_1^i\bigr) = C_i p_{\mathbf{X}}(y_{i+1}) \mathfrak{n}\bigl(\alpha\beta + m_0, \beta, u(y_{i+1})\bigr). \tag{26}$$

Here

$$\alpha = t_i + \frac{\mu_3^i}{2s_i^4(n-i-1)}, \tag{27}$$

$$\beta = s_i^2(n-i-1), \tag{28}$$

and the C_i is a normalizing constant.

Set

$$g_{u_{1,n}}(y_1^k) := g_0(y_1|y_0) \prod_{i=1}^{k-1} g(y_{i+1}|y_1^i). \tag{29}$$

THEOREM 8. *Assume* (K1) *and* (K2) *together with* (E1) *and* (E2). *Then* (i)

$$p_{u_{1,n}}(Y_1^k) := p(\mathbf{X}_1^k = Y_1^k | \mathbf{U}_{1,n} = u_{1,n}) = g_{u_{1,n}}(Y_1^k)(1 + o_{P_{u_{1,n}}}(\varepsilon_n (\log n)^2))$$

and (ii)

$$p_{u_{1,n}}(Y_1^k) = g_{u_{1,n}}(Y_1^k)(1 + o_{G_{u_{1,n}}}(\varepsilon_n (\log n)^2)).$$

PROOF. We only sketch the initial step of the proof of (i), which rapidly follows the same path as that in Theorem 2.

As in the proof of Theorem 2, evaluate

$$\begin{aligned}
&p(\mathbf{X}_{i+1} = Y_{i+1} | \mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}) \\
&\quad = p_{\mathbf{X}}(\mathbf{X}_{i+1} = Y_{i+1}) \frac{p_{\mathbf{U}}(\mathbf{U}_{i+2,n} = u_{1,n} - U_{1,i+1})}{p_{\mathbf{U}}(\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i})} \\
&\quad = \frac{p_{\mathbf{X}}(\mathbf{X}_{i+1} = Y_{i+1})}{p_{\mathbf{U}}(\mathbf{U}_{i+1} = u(Y_{i+1}))} p_{\mathbf{U}}(\mathbf{U}_{i+1} = u(Y_{i+1})) \frac{p_{\mathbf{U}}(\mathbf{U}_{i+2,n} = u_{1,n} - U_{1,i+1})}{p_{\mathbf{U}}(\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i})}.
\end{aligned}$$

Use the invariance of the conditional density with respect to the change of sampling defined by $\pi_{\mathbf{U}}^{m_i}$ to obtain

$$\begin{aligned}
&p(\mathbf{X}_{i+1} = Y_{i+1} | \mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}) \\
&\quad = \frac{p_{\mathbf{X}}(\mathbf{X}_{i+1} = Y_{i+1})}{p_{\mathbf{U}}(\mathbf{U}_{i+1} = u(Y_{i+1}))} \pi_{\mathbf{U}}^{m_i}(\mathbf{U}_{i+1} = u(Y_{i+1})) \frac{\pi_{\mathbf{U}}^{m_i}(\mathbf{U}_{i+2,n} = u_{1,n} - U_{1,i+1})}{\pi_{\mathbf{U}}^{m_i}(\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i})} \\
&\quad = p_{\mathbf{X}}(\mathbf{X}_{i+1} = Y_{i+1}) \frac{e^{t_i u(Y_{i+1})}}{\phi_{\mathbf{U}}(t_i)} \frac{\pi_{\mathbf{U}}^{m_i}(\mathbf{U}_{i+2,n} = u_{1,n} - U_{1,i+1})}{\pi_{\mathbf{U}}^{m_i}(\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i})},
\end{aligned}$$

and proceed via the Edgeworth expansions in the above expression, following verbatim the proof of Theorem 2. We omit details. The proof of (ii) follows from Lemma 6. □

We turn to a consequence of Theorem 8.

For all $\delta > 0$, let

$$E_{k,\delta} := \left\{ y_1^k \in \mathbb{R}^k : \left| \frac{p_{u_{1,n}}(y_1^k) - g_{u_{1,n}}(y_1^k)}{g_{u_{1,n}}(y_1^k)} \right| < \delta \right\},$$

which by Theorem 8 satisfies

$$\lim_{n\to\infty} P_{u_{1,n}}(E_{k,\delta}) = \lim_{n\to\infty} G_{u_{1,n}}(E_{k,\delta}) = 1. \tag{30}$$

It holds that

$$\sup_{C\in\mathcal{B}(\mathbb{R}^k)} \left|P_{u_{1,n}}(C\cap E_{k,\delta}) - G_{u_{1,n}}(C\cap E_{k,\delta})\right|$$
$$\leq \delta \sup_{C\in\mathcal{B}(\mathbb{R}^k)} \int_{C\cap E_{k,\delta}} g_{u_{1,n}}(y_1^k)\,dy_1^k \leq \delta.$$

By (30)

$$\sup_{C\in\mathcal{B}(\mathbb{R}^k)} \left|P_{u_{1,n}}(C\cap E_{k,\delta}) - P_{u_{1,n}}(C)\right| < \eta_n$$

and

$$\sup_{C\in\mathcal{B}(\mathbb{R}^k)} \left|G_{u_{1,n}}(C\cap E_{k,\delta}) - G_{u_{1,n}}(C)\right| < \eta_n$$

for some sequence $\eta_n \to 0$; hence

$$\sup_{C\in\mathcal{B}(\mathbb{R}^k)} \left|P_{u_{1,n}}(C) - G_{u_{1,n}}(C)\right| < \delta + 2\eta_n$$

for all positive δ. Applying Scheffé's lemma, we have proved:

THEOREM 9. *Under the hypotheses of Theorem 8 the total variation distance between $P_{u_{1,n}}$ and $G_{u_{1,n}}$ goes to 0 as n tends to infinity, and*

$$\lim_{n\to\infty} \int \left|p_{u_{1,n}}(y_1^k) - g_{u_{1,n}}(y_1^k)\right| dy_1^k = 0.$$

REMARK 10. This result is to be compared with Theorem 1.6 in [13] and Theorem 2.15 in [11] which provides a rate for this convergence for small k's under some additional conditions on the moment generating function of $\mathbf{U}$.

3.1.1. *Approximation under other sampling schemes.* In statistical applications the r.v.'s Y_i's in Theorems 2 and 8 may in certain cases be sampled under some other distribution than P_{na} or G_{na}.

Consider the following situation.

The model consists of an exponential family $\mathcal{P} := \{P_{\theta,\eta}, (\theta,\eta) \in \mathcal{N}\}$ defined on $\mathbb{R}$ with canonical parametrization (θ,η) and sufficient statistics (t,u) defined on $\mathbb{R}$ through the densities

$$p_{\theta,\eta}(x) := \frac{dP_{\theta,\eta}(x)}{dx} = \exp(\theta t(x) + \eta u(x) - K(\theta,\eta))h(x). \tag{31}$$

We assume that both θ and η belong to $\mathbb{R}$. The natural parameter space $\mathcal{N}$ is a convex set in $\mathbb{R}^2$ defined as the domain of

$$k(\theta,\eta) := \exp(K(\theta,\eta)) = \int \exp(\theta t(x) + \eta u(x))h(x)\,dx.$$

For the statistician, θ is the parameter of interest whereas η is a nuisance one. The unknown parameter of the i.i.d. sample $\mathbf{X}_1^n := (\mathbf{X}_1, \dots, \mathbf{X}_n)$ observed as $X_1^n := (X_1, \dots, X_n)$ is (θ_T, η_T).

Conditioning on a sufficient statistic for the nuisance parameter produces a new exponential family which is free of η. For any θ denote $\widehat{\eta}_\theta$ the MLE of η_T in model (31) parametrized in η, when θ is fixed. A classical solution for the estimation of θ_T consists in maximizing the likelihood

$$L(\theta|X_1^n) := \prod_{i=1}^n p_{\theta,\widehat{\eta}_\theta}(X_i)$$

with respect to θ. This approach produces satisfactory results when $\widehat{\eta}_\theta$ is a consistent estimator of η_θ. However for curved exponential families, it may happen that for some θ the likelihood

$$L_\theta(\eta|X_1^n) := \prod_{i=1}^n p_{\theta,\eta}(X_i)$$

is multimodal with respect to η which may produce misestimation in $\widehat{\eta}_\theta$, leading in turn to inconsistency in the resulting estimates of θ_T; see [20].

Consider $g_{u_{1,n},(\theta,\eta)}$ defined in (29) for fixed (θ, η), with $u_{1,n} := u(X_1) + \cdots + u(X_n)$. Since $u_{1,n}$ is sufficient for η, $p_{u_{1,n},(\theta,\eta)}$ is independent of η for all k. Assume at present that the density $g_{u_{1,n},(\theta,\eta)}$ on $\mathbb{R}^k$ approximates $p_{u_{1,n},(\theta,\eta)}$ on the sample X_1^n generated under (θ_T, η_T); it follows then that inserting any value η_0 in (29) does not change the value of the resulting likelihood

$$L_{\eta_0}(\theta|X_1^k) := g_{u_{1,n},(\theta,\eta_0)}(X_i).$$

Optimizing $L_{\eta_0}(\theta|X_1^k)$ with respect to θ produces a consistent estimator of θ_T. We refer to [5] for examples and discussion.

Let $\mathbf{Y}_1^n$ be i.i.d. copies of $\mathbf{Z}$ with distribution Q and density q; assume that Q satisfies the Cramér condition $\int(\exp(tx))q(x)\,dx < \infty$ for t in a nonempty neighborhood of 0. Let $\mathbf{V}_{1,n} := u(\mathbf{Y}_1) + \cdots + u(\mathbf{Y}_n)$, and define

$$q_{u_{1,n}}(y_1^k) := q(\mathbf{Y}_1^k = y_1^k | \mathbf{V}_{1,n} = u_{1,n})$$

with distribution $Q_{u_{1,n}}$. The following theorem then holds:

THEOREM 11. *Assume* (K1) *and* (K2) *together with* (E1) *and* (E2). *Then, with the same hypotheses and notation as in Theorem* 8,

$$p(\mathbf{X}_1^k = Y_1^k | \mathbf{U}_{1,n} = u_{1,n}) = g_{u_{1,n}}(Y_1^k)(1 + o_{Q_{u_{1,n}}}(\varepsilon_n (\log n)^2)).$$

Also the total variation distance between $Q_{u_{1,n}}$ *and* $P_{u_{1,n}}$ *goes to* 0 *as* n *tends to infinity.*

PROOF. It is enough to check that Lemmas 21, 22 and 23 hold when $\mathbf{Y}$ satisfies the Cramér condition. $\square$

REMARK 12. In the previous discussion $Q = P_{\theta_T,\eta_T}$ and $\mathbf{X}_1^n$ are independent copies of $\mathbf{X}$ with distribution P_{θ,η_0}.

3.2. *For how long is the approximation valid?* This section provides a rule leading to an effective choice of the crucial parameter k in order to achieve a given accuracy bound for the relative error in Theorem 8(ii). The accuracy of the approximation is measured through

$$\mathrm{ERE}(k) := E_{G_{u_{1,n}}} 1_{D_k}(Y_1^k) \frac{p_{u_{1,n}}(Y_1^k) - g_{u_{1,n}}(Y_1^k)}{p_{u_{1,n}}(Y_1^k)} \tag{32}$$

and

$$\mathrm{VRE}(k) := \mathrm{Var}_{G_{u_{1,n}}} 1_{D_k}(Y_1^k) \frac{p_{u_{1,n}}(Y_1^k) - g_{u_{1,n}}(Y_1^k)}{p_{u_{1,n}}(Y_1^k)} \tag{33}$$

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

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

with $\varepsilon_n(\log n)^2/\delta_n \to 0$ and $\delta_n \to 0$; therefore $G_{u_{1,n}}(D_k) \to 1$. The r.v.'s Y_1^k are sampled under $g_{u_{1,n}}$. Note that the density $p_{u_{1,n}}$ 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

$$\mathrm{CI}(k) := [\mathrm{ERE}(k) - 2\sqrt{\mathrm{VRE}(k)}, \mathrm{ERE}(k) + 2\sqrt{\mathrm{VRE}(k)}].$$

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

The calculation of $\mathrm{VRE}(k)$ and $\mathrm{ERE}(k)$ should be carried out as follows.

Write

$$\begin{aligned}\mathrm{VRE}(k)^2 &= E_{P_{\mathbf{X}}}\left(\frac{g_{u_{1,n}}^3(Y_1^k)}{p_{u_{1,n}}(Y_1^k)^2 p_{\mathbf{X}}(Y_1^k)}\right) \\ &\quad - E_{P_{\mathbf{X}}}\left(\frac{g_{u_{1,n}}^2(Y_1^k)}{p_{u_{1,n}}(Y_1^k) p_{\mathbf{X}}(Y_1^k)}\right)^2 \\ &=: A - B^2.\end{aligned}$$

By the Bayes formula,

$$p_{u_{1,n}}(Y_1^k) = p_{\mathbf{X}}(Y_1^k) \frac{np(\mathbf{U}_{k+1,n}/(n-k) = m(t_k))}{(n-k)p(\mathbf{U}_{1,n}/n = u_{1,n}/n)}. \tag{34}$$

The following lemma holds; see [17] and [19].

LEMMA 13. *Let $\mathbf{U}_1, \dots, \mathbf{U}_n$ be i.i.d. random variables with common density $p_{\mathbf{U}}$ on $\mathbb{R}$ and satisfying the Cramér conditions with m.g.f. $\phi_{\mathbf{U}}$. Then with $m(t) = u$,*

$$p(\mathbf{U}_{1,n}/n = u) = \frac{\sqrt{n}\phi_{\mathbf{U}}^n(t)\exp(-ntu)}{s(t)\sqrt{2\pi}}\bigl(1 + o(1)\bigr)$$

when $|u|$ is bounded.

Introduce

$$D := \left[\frac{\pi_{\mathbf{U}}^{m_0}(m_0)}{p_{\mathbf{U}}(m_0)}\right]^n$$

and

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

with m_k defined in (25) and $m_0 = u_{1,n}/n$. Define t by $m(t) = m_0$. By (34) and Lemma 13 it holds that

$$p_{u_{1,n}}(Y_1^k) = \sqrt{\frac{n}{n-k}} p_{\mathbf{X}}(Y_1^k) \frac{D}{N} \frac{s(t)}{s(t_k)} \bigl(1 + o_{P_{u_{1,n}}}(1)\bigr).$$

The approximation of A is obtained through Monte Carlo simulation. Define

$$A(Y_1^k) := \frac{n-k}{n} \left(\frac{g_{u_{1,n}}(Y_1^k)}{p_{\mathbf{X}}(Y_1^k)}\right)^3 \left(\frac{N}{D}\right)^2 \frac{s^2(t_k)}{s^2(t)}, \tag{35}$$

and simulate L i.i.d. samples $Y_1^k(l)$, each one made of k i.i.d. replicates under $p_{\mathbf{X}}$. Set

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

We use the same approximation for B. Define

$$B(Y_1^k) := \sqrt{\frac{n-k}{n}} \left(\frac{g_{u_{1,n}}(Y_1^k)}{p_{\mathbf{X}}(Y_1^k)}\right)^2 \left(\frac{N}{D}\right) \frac{s(t_k)}{s(t)} \tag{36}$$

and

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

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

Set

$$\overline{\mathrm{VRE}}(k) := \widehat{A} - (\widehat{B})^2, \tag{37}$$

which is a suitable approximation of $\mathrm{VRE}(k)$.

The curve $k \to \overline{\mathrm{ERE}}(k)$ is a proxy for (32) and is obtained through

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

A proxy of $\mathrm{CI}(k)$ can now be defined as

$$\overline{\mathrm{CI}}(k) := \left[\overline{\mathrm{ERE}}(k) - 2\sqrt{\overline{\mathrm{VRE}}(k)}, \overline{\mathrm{ERE}}(k) + 2\sqrt{\overline{\mathrm{VRE}}(k)}\right]. \tag{38}$$

We now check the validity of the above approximation, comparing $\overline{\mathrm{CI}}(k)$ with $\mathrm{CI}(k)$ on a toy case.

Consider $u(x) = x$. The case when $p_{\mathbf{X}}$ is a centered exponential distribution with variance 1 allows for an explicit evaluation of $\mathrm{CI}(k)$ making no use of Lemma 13. The conditional density p_{na} is calculated analytically, the density g_{na} is obtained through (10), hence providing a benchmark for our proposal. The terms $\widehat{A}$ and $\widehat{B}$ are obtained by Monte Carlo simulation following the algorithm presented below. Figures 1, 2 and 3, 4 show the increase in δ w.r.t. k in the large deviation range, with a such that $P(\mathbf{S}_{1,n} > na) \simeq 10^{-8}$. We have considered two cases, when $n = 100$ and when $n = 1000$. These figures show that the approximation scheme is quite accurate, since the relative error is fairly small. Also they show that $\overline{\mathrm{ERE}}$ and $\overline{\mathrm{CI}}$ provide good tools for the assessing the value of k.

Algorithms 1 and 2 produce the curve $k \to \overline{\mathrm{CI}}(k)$. The resulting $k = k_\delta$ is the longest run length for which $g_{u_{1,n}}$ a good proxy for $p_{u_{1,n}}$.

The calculation of $g_{u_{1,n}}(y_1^k)$ above requires the value of

$$C_i = \left(\int p_{\mathbf{X}}(x)\mathfrak{n}\big(\alpha\beta + m_0, \beta, u(x)\big)\,dx\right)^{-1}.$$

This can be done through Monte Carlo simulation.

REMARK 14. Solving $t_i = m^{-1}(m_i)$ might be difficult. It may happen that the inverse function of m is at hand, but even when $p_{\mathbf{X}}$ is the Weibull density and $u(x) = x$, this is not the case. We can replace step $*$ by

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

Indeed since

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

```
Input          : y_1^k, p_X, n, u_{1,n}
Output         : g_{u_{1,n}}(y_1^k)
Initialization:
        t_0 ← m^{-1}(m_0);
        g_0(y_1|y_0) ← (24);
Procedure      :
        for i ← 1 to k − 1 do
            m_i ← (25);
            t_i ← m^{-1}(m_i) ∗;
            α ← (27);
            β ← (28);
            Calculate C_i;
            g(y_{i+1}|y_1^i) ← (26);
        end
        Compute g_{u_{1,n}}(y_1^k) ← (29);
Return         : g_{u_{1,n}}(y_1^k)
```

Algorithm 1: Evaluation of $g_{u_{1,n}}(y_1^k)$.

use a first order approximation to derive that t_{i+1} can be substituted by τ_{i+1} defined as

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

When $\lim_{n\to\infty} u_{1,n}/n = Eu(\mathbf{X})$, the values of the function $s^2(\cdot)$ are close to $\mathrm{Var}[u(\mathbf{X})]$, and the above approximation is appropriate. For the large deviation case, the same argument applies, since $s^2(t_i)$ keeps close to $s^2(t^a)$.

3.2.1. *Simulation of typical paths of a random walk under a conditioning point.* By Theorem 8(ii), $g_{u_{1,n}}$ and the density of $p_{u_{1,n}}$ approach each other on a family of subsets of $\mathbb{R}^k$ which contain the typical paths of the random walk under the conditional density with probability going to 1 as n increases. By Lemma 6 large sets under $P_{u_{1,n}}$ are also large sets under $G_{u_{1,n}}$. It follows that long runs of typical paths under $p_{u_{1,n}}$ can be simulated as typical paths under $g_{u_{1,n}}$ defined in (29) at least for large n.

The simulation of a sample X_1^k with $g_{u_{1,n}}$ can be fast and easy when $\lim_{n\to\infty} u_{1,n}/n = Eu(\mathbf{X})$. Indeed the r.v. $\mathbf{X}_{i+1}$ with density $g(x_{i+1}|x_1^i)$ is obtained through a standard acceptance-rejection algorithm. The values of the parameters which appear in the Gaussian component of $g(x_{i+1}|x_1^i)$ in (7) are easily calculated, and the dominating density can be chosen for all i as $p_{\mathbf{X}}$. The constant in the acceptance rejection algorithm is then $1/\sqrt{2\pi\beta}$. This is in contrast with the

Input : $p_{\mathbf{X}}$, δ, n, $u_{1,n}$, L
Output : k_δ
Initialization: $k = 1$
Procedure :
  while $\delta \notin \overline{\mathrm{CI}}(k)$ **do**
    for $l \leftarrow 1$ **to** L **do**
      Simulate $Y_1^k(l)$ i.i.d. with density $p_{\mathbf{X}}$;
      $A(Y_1^k(l)) :=$ (35) using Algorithm 1;
      $B(Y_1^k(l)) :=$ (36) using Algorithm 1;
    end
    Calculate $\overline{\mathrm{CI}}(k) \leftarrow$ (38);
    $k := k + 1$;
  end
Return : $k_\delta := k$

Algorithm 2: Calculation of k_δ.

case when the conditioning value is in the range of a large deviation event, that is, $\lim_{n\to\infty} u_{1,n}/n \neq Eu(\mathbf{X})$, which appears in a natural way in importance sampling estimation for rare event probabilities; then MCMC techniques can be used.

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

REMARK 15. Simulation of Y_1 can be performed through the method suggested in [1].

Input : p, μ, σ^2
Output : Y
Initialization:
  Select a density f on [0, 1] and a positive constant K such that $p(\mathfrak{N}^{-1}(x)) \leq Kf(x)$ for all x in [0, 1]
Procedure : **while** $Z < p(\mathfrak{N}^{-1}(X))$ **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 3: Simulation of Y with density proportional to $p(x)\mathfrak{n}(\mu, \sigma^2, x)$.

Input : $p_{\mathbf{X}}, \delta, n, u_{1,n}$
Output : Y_1^k
Initialization:
Set $k \leftarrow k_\delta$ with Algorithm 2;
$t_0 \leftarrow= m^{-1}(m_0)$;
Procedure :
Simulate Y_1 with density (24);
$u_{1,1} \leftarrow u(Y_1)$;
for $i \leftarrow 1$ **to** $k-1$ **do**
$m_i \leftarrow (25)$;
$t_i \leftarrow m^{-1}(m_i)$;
$\alpha \leftarrow (27)$;
$\beta \leftarrow (28)$;
Simulate Y_{i+1} with density $g(y_{i+1}|y_1^i)$ using Algorithm 3;
$u_{1,i+1} \leftarrow u_{1,i} + u(Y_{i+1})$;
end
Return : Y_1^k

Algorithm 4: Simulation of a sample Y_1^k with density $g_{u_{1,n}}$.

Figures 5, 6, 7 and 8 present a number of simulations of random walks conditioned on their sum with $n = 1000$ when $u(x) = x$. In the Gaussian case, when the approximating scheme is known to be optimal up to $k = n - 1$, the simulation is performed with $k = 999$ and two cases are considered: the moderate deviation case is assumed to be modeled when $P(\mathbf{S}_{1,n} > na) = 10^{-2}$ (Figure 5); that this range of probability is in the "moderate deviation" range is a commonly assessed statement among statisticians. The large deviation case pertains to $P(\mathbf{S}_{1,n} > na) = 10^{-8}$ (Figure 6). The centered exponential case with $n = 1000$ and $k = 800$ is presented in Figures 7 and 8, under the same events.

In order to check the accuracy of the approximation, Figures 9, 10 (normal case, $n = 1000$, $k = 999$) and Figures 11, 12 (centered exponential case, $n = 1000$, $k = 800$) present the histograms of the simulated $\mathbf{X}_i$'s together with the tilted densities at point a which are known to be the limit density of $\mathbf{X}_1$ conditioned on $\mathcal{E}_n$ in the large deviation case, and to be equivalent to the same density in the moderate deviation case, as can be deduced from [15]. The tilted density in the Gaussian case is the normal with mean a and variance 1; in the centered exponential case the tilted density is an exponential density on $(-1, \infty)$ with parameter $1/(1 + a)$.

Consider now the case when $u(x) = x^2$. Figure 13 presents the case when $\mathbf{X}$ is $N(0, 1)$, $n = 1000, k = 800$, $P(\mathbf{U}_{1,n} = u_{1,n}) \simeq 10^{-2}$. We present the histograms of the X_i's together with the graph of the corresponding tilted density; when $\mathbf{X}$ is $N(0, 1)$, then $\mathbf{X}^2$ is χ^2. It is well known that when $u_{1,n}/n$ is fixed to be larger than 1, then the limit distribution of $\mathbf{X}_1$ conditioned on $(\mathbf{U}_{1,n} = u_{1,n})$

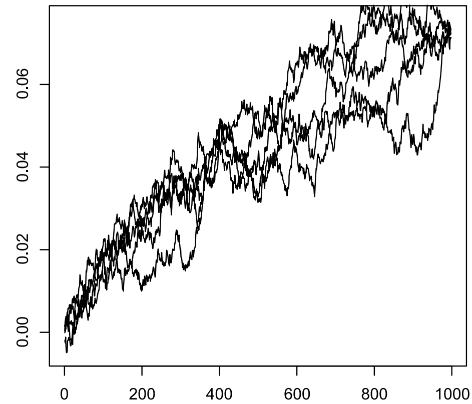

FIG. 5. *Trajectories in the normal case for $P_n = 10^{-2}$.*

tends to $N(0, a)$ which is the Kullback–Leibler projection of $N(0, 1)$ on the set of all probability measures Q on $\mathbb{R}$ with $\int x^2\, dQ(x) = a := \lim_{n\to\infty} u_{1,n}/n$. This distribution is precisely $g_0(y_1|y_0)$ defined above. Also consider (26); the expansion using the definitions (27) and (28) prove that as $n \to \infty$ the dominating term in $g_i(y_{i+1}|y_1^i)$ is precisely $N(0, m_0)$, and the terms including y_{i+1}^4 in the exponential stemming from $\mathfrak{n}(\alpha\beta + m_0, \beta, u(y_{i+1}))$ are of order $O(1/(n-i))$; the terms depending on y_1^i are of smaller order. The fit which is observed in Figure 13 is in accordance with the above statement in the LDP range (when $\lim_{n\to\infty} u_{1,n}/n \neq 1$), and with the MDP approximation when $\lim_{n\to\infty} u_{1,n}/n = 1$ and $\liminf_{n\to\infty}(u_{1,n} - n)/\sqrt{n} \neq 0$, following [15].

4. Conditioning on large sets. The approximation of the density

$$p_{A_n}(\mathbf{X}_1^k = Y_1^k) := p(\mathbf{X}_1^k = Y_1^k | \mathbf{U}_{1,n} \in A_n)$$

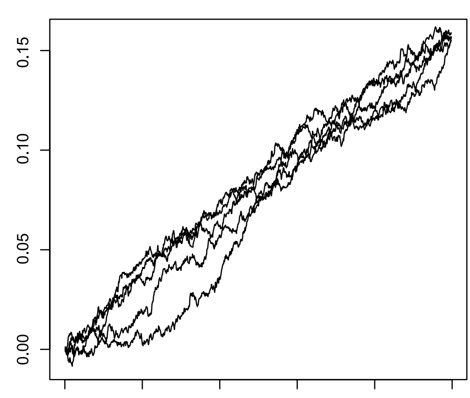

FIG. 6. *Trajectories in the normal case for $P_n = 10^{-8}$.*

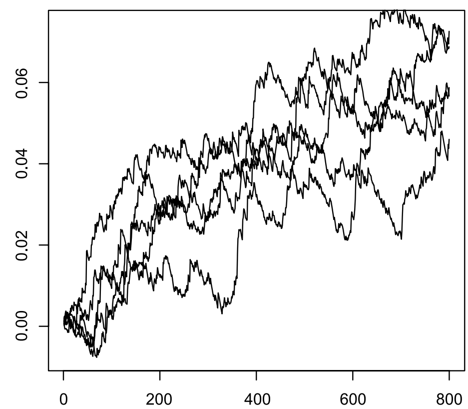

FIG. 7. *Trajectories in the exponential case for* $P_n = 10^{-2}$.

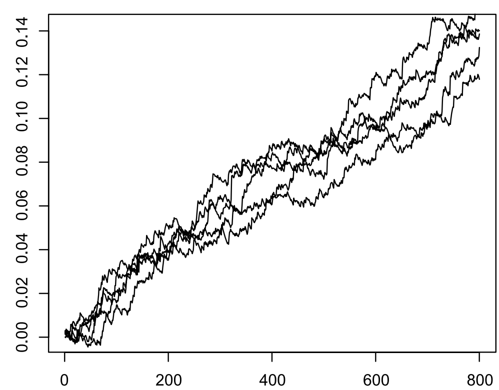

FIG. 8. *Trajectories in the exponential case for* $P_n = 10^{-8}$.

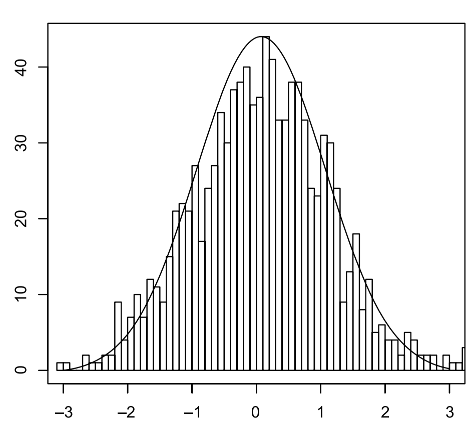

FIG. 9. *Histogram of the* $\mathbf{X}_i$*'s in the normal case with* $n = 1000$ *and* $k = 999$ *for* $P_n = 10^{-2}$. *The curve represents the associated tilted density.*

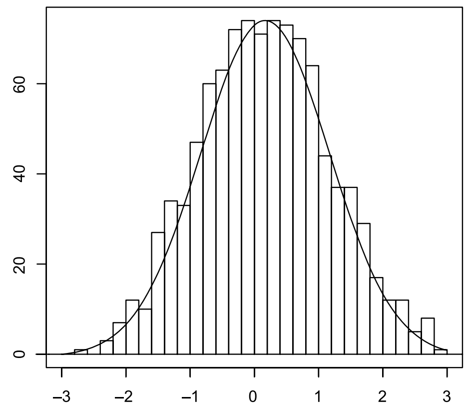

FIG. 10. *Histogram of the* $\mathbf{X}_i$*'s in the normal case with* $n = 1000$ *and* $k = 999$ *for* $P_n = 10^{-8}$. *The curve represents the associated tilted density.*

of the runs $\mathbf{X}_1^k$ under large sets $(\mathbf{U}_{1,n} \in A_n)$ for Borel sets A_n with nonempty interior follows from the above results through integration. Here, in the same vein as previously, Y_1^k is generated under P_{A_n}. An application of this result for the evaluation of rare event probabilities through importance sampling is briefly presented in the next section. The present section pertains to the large deviation case.

4.1. *Conditioning on a large set defined through the density of its dominating point.* We focus on cases when $(\mathbf{U}_{1,n} \in A_n)$ can be expressed as $(\mathbf{U}_{1,n}/n \in A)$ where A is a fixed Borel set (independent of n) with essential infimum α larger than $E\mathbf{U}$ and which can be described as a "thin" or "thick" Borel set according to its local density at point α.

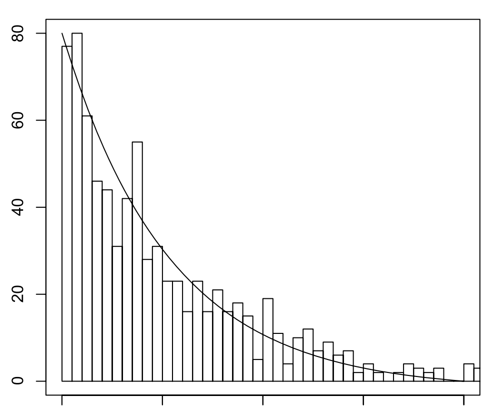

FIG. 11. *Histogram of the* $\mathbf{X}_i$*'s in the exponential case with* $n = 1000$ *and* $k = 800$ *for* $P_n = 10^{-2}$. *The curve represents the associated tilted density.*

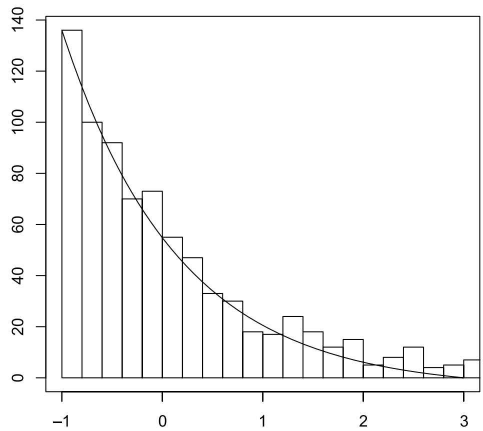

FIG. 12. *Histogram of the* $\mathbf{X}_i$*'s in the exponential case with* $n = 1000$ *and* $k = 800$ *for* $P_n = 10^{-8}$. *The curve represents the associated tilted density.*

The starting point is the approximation of p_{nv} on $\mathbb{R}^k$ for large values of k under the conditioning point

$$\mathbf{U}_{1,n}/n = v$$

when v belongs to A. Denote g_{nv} the corresponding approximation defined in (29). It holds that

$$p_{nA}(x_1^k) = \int_A p_{nv}(\mathbf{X}_1^k = x_1^k) p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} \in nA)\, ds. \tag{39}$$

In contrast with the classical importance sampling approach for this problem we do not consider the dominating point approach, but merely realize a sharp approx-

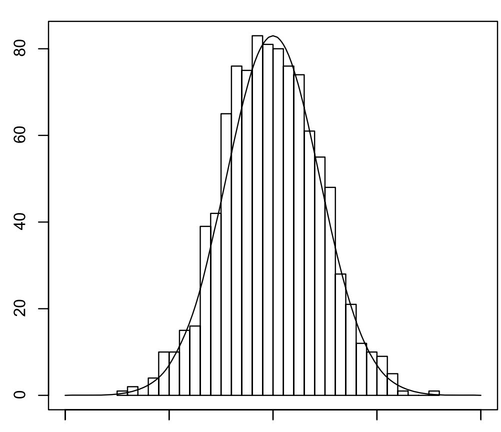

FIG. 13. *Histogram of the* $\mathbf{X}_i$*'s in the normal case with* $n = 1000$, $k = 800$ *and* $u(x) = x^2$ *for* $P_n = 10^{-2}$. *The curve represents the associated tilted density.*

imation 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 [3] for sharp approximations of Laplace-type integrals in $\mathbb{R}^d$.

Turning to (39) it appears that what is needed is a sharp approximation for

$$p(\mathbf{U}_{1,n}/n = v|\mathbf{U}_{1,n} \in nA) = \frac{p(\mathbf{U}_{1,n}/n = v)\mathbb{1}_A(v)}{P(\mathbf{U}_{1,n} \in nA)} \tag{40}$$

with some uniformity for v in A. We will assume that A is bounded above in order to avoid further regularity assumptions on the distribution of $\mathbf{U}$.

Recall that the *essential infimum* $\texttt{essinf}\, A = \alpha$ of the set A with respect to the Lebesgue measure is defined through

$$\alpha := \inf\{x : \text{ for all } \varepsilon > 0, \big|[x, x+\varepsilon] \cap A\big| > 0\}$$

with $\inf \varnothing := -\infty$.

We assume that $\alpha > -\infty$, which is tantamount to saying that we do not consider very thin sets (e.g., not Cantor-type sets).

The density of the point α in A will not be measured in the ordinary way, through

$$d(\alpha) := \lim_{\varepsilon \to 0} \frac{|A \cap [\alpha - \varepsilon, \alpha + \varepsilon]|}{\varepsilon},$$

but through the more appropriate quantity

$$M(t) := t \int_{A-\alpha} e^{-ty}\, dy, \qquad t > 0.$$

For any set A, $0 \le M(t) \le 1$. If there exists an interval $[\alpha, \alpha + \varepsilon] \subset A$, then $\lim_{t\to\infty} M(t) = 1$. As an example, for a self similar set $A := A_p$ defined as $A_p := \bigcup_{n\in\mathbb{Z}} p^n I_p$ where $p > 2$ and $I_p := [(p-1)/p, 1]$, it holds that $0 = \operatorname{essinf} A_p$ and $pA_p = A_p$. Consequently for any $t \ge 0$, $M(tp) = M(t)$ and $M(tp) = M(t)$ for all $t \ge 0$; it follows that

$$\inf_{1 \le u \le p} M(u) = \lim \inf_{t\to\infty} M(t) \le \lim \sup_{t\to\infty} M(t) = \sup_{1\le u\le p} M(u).$$

Define

$$M_n(t) := M(nt)/t = \int_{A-\alpha} e^{-ty}\, dy$$

and

$$\Psi_n(t) := n \log \phi_{\mathbf{U}}(t) + \log M_n(t) - n\alpha t$$

for all $t > 0$ such that $\phi_{\mathbf{U}}(t)$ is finite. We borrow from [2] the following results.

Define $\mu_n(t) := (1/n)\log M_n(t)$ which is for all $n \ge 1$ a decreasing function of t on $(0, \infty)$, and which is negative for large n. Also $\mu_n'(t) = \mu_1'(nt)$ and μ_1' are nondecreasing on $(0, \infty)$.

Let $\overline{\mu} := \lim_{t\to\infty} \mu_1'(t)$ and $\underline{\mu} := \lim_{t\to 0} \mu_1'(t)$. Then according to [2] the following holds:

LEMMA 16. *Under the above notation and hypotheses, the equation* $\Psi_n'(t) = 0$ *has a unique solution* t_n *in* $(0, t_0)$ *for* α *in* $(E\mathbf{U} + \overline{\mu}, \infty)$ *where* $t_0 := \sup\{t : \phi_{\mathbf{U}}(t) < \infty\}$. *Furthermore if* $\alpha > E\mathbf{U} + \underline{\mu}$, *then there exists a compact set* $K \subset (0, t_0)$ *such that* $t_n \in K$ *for all* n.

Assume that $\alpha > E\mathbf{U} + \underline{\mu}$. Define $\psi_n(t) := \Psi_n''(t)$, and suppose that for any $\lambda > 0$,

$$\lim_{n\to\infty} \sup_{|u|<\lambda} \frac{\psi_n(t_n + u/\sqrt{\psi_n(t_n)})}{\psi_n(t_n)} = 1, \tag{41}$$

where t_n is a solution of $\Psi_n'(t) = 0$ in the range $(0, t_0)$. It can be proved that (41) holds, for example, when $t \to \log M(t)/t$ is a regularly varying function at infinity with index $\rho \in (0, 1)$, that is, $\log M(t)/t \in \mathcal{R}_\rho(\infty)$; see [2], Lemma 2.2.

We also assume that

$$\limsup_{t\to\infty} t\big(\log M(t)\big) < \infty, \tag{42}$$

which holds, for example, when $\log(M(t)/t) \in \mathcal{R}_\rho(\infty)$, for $0 \le \rho < 1$.

Theorem 2.1 in [2] provides a general result to be inserted in (40); we take the occasion to correct a misprint in this result.

THEOREM 17. *Assume* (41) *and* (42) *together with the aforementioned conditions on the r.v.* $\mathbf{U}$. *Then for* $\alpha > E\mathbf{U} + \underline{\mu}$,

$$P(\mathbf{U}_{1,n} \in nA) = \frac{\phi_{\mathbf{U}}^n(t_n) M_n(t_n) e^{-nt_n\alpha}}{\sqrt{\psi_n(t_n)}\sqrt{2\pi}}\big(1 + o(1)\big) \qquad \text{as } n \to \infty, \tag{43}$$

with t_n *satisfying* $\Psi_n'(t) = 0$ *provided that the function* $x \to P(\mathbf{U}_{1,n} \in nA + x)$ *is nonincreasing for* n *large enough. In particular, this last condition holds if*

(i) (*Petrov*): $A = (\alpha, \infty)$ *or* $A = [\alpha, \infty)$; *in this case* $M_n(t) = 1/t$; *note that in this case the classical result is slightly different, since*

$$P(\mathbf{U}_{1,n} > na) = \frac{\phi_{\mathbf{U}}^n(t^a) e^{-nt^a a}}{t^a s(t^a)\sqrt{2\pi}}\big(1 + o(1)\big) \qquad \text{as } n \to \infty$$

with $m(t^a) = a$ *and* $a > E\mathbf{U}$; *this is readily seen to be equivalent to* (43) *when* $A = (a, \infty)$.

(ii) $\mathbf{U}$ *has a symmetric unimodal distribution.*

(iii) $\mathbf{U}$ *has a strongly unimodal distribution.*

The shape of A near α is reflected in the behavior of the function $M(t)$ for large values of t. As such, the larger the n, the more relevant is the shape of A near α.

Note further that $M_n(t)e^{-nt\alpha} = \int_A e^{-nty}\,dy$ from which we see that α plays no role in (43). Hence α can be replaced by any number γ such that $\int_{A-\gamma} e^{-ty}\,dy$ converges. Further t_n is independent of α. The so-called dominating point α of A can therefore be defined as

$$\alpha := \lim_{t\to\infty} \log \int_A e^{-ty}\,dy.$$

In order to examine further the role played in (43) by the regularity of A near its essential infimum α, introduce the pointwise Hölder dimension of A at α as

$$\delta(\alpha) := \frac{\log G(\varepsilon)}{-\log \varepsilon},$$

where

$$G(\varepsilon) := \bigl|A \cap [\alpha, \alpha+\varepsilon]\bigr| \qquad \text{for positive } \varepsilon.$$

We refer to Proposition 2.1 in [2] for a set of Abel–Tauber-type results which link the properties of $M(t)$ at infinity with those of G at 0. For example, it follows that $G(\varepsilon) \sim \varepsilon^{\delta(\alpha)}$ (as $\varepsilon \to 0$) if and only if $M(t) \sim ct^{-\delta(\alpha)+1}\Gamma(1+\delta(\alpha))$ (as $t \to \infty$). Consequently if $M_n(t) \to 1$ as $t \to \infty$, then $M(t) \sim t$ as $t \to \infty$ and $G(\varepsilon) \sim \varepsilon$ as $\varepsilon \to 0$.

Asymptotic formulas for the numerator in (40) are well known and have a long history, going back to [19]. It holds that

$$p(\mathbf{U}_{1,n}/n = v) = \frac{\sqrt{n}e^{nvt^v}\phi_{\mathbf{U}}(t^v)}{\sqrt{2\pi}s(t^v)}\bigl(1+o(1)\bigr) \qquad \text{as } n \to \infty \tag{44}$$

with t^v defined as $m(t^v) = v$.

Plugging in (44) and (43) in (39) provides an expression for the density of the runs. For applications the only relevant case is developed in the following paragraph.

4.2. *Conditioning on a thick set.* In the case when $A = (a, \infty)$ or with $a > Eu(\mathbf{X})$ or, more generally, when A is a thick set in a neighborhood of its essential infimum [i.e., when $\lim_{t\to\infty} M(t) = 1$] a simple asymptotic evaluation for (40) when A is unbounded can be obtained. Indeed an expansion of the ratio yields

$$p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} > na) = \bigl(nt \exp\bigl(-nt(v-a)\bigr)\bigr)\mathbb{1}_A(v)\bigl(1+o(1)\bigr) \tag{45}$$

with $m(t) = a$, indicating that $\mathbf{U}_{1,n}/n$ is roughly exponentially distributed on A with expectation $a + 1/nt$. This result is used in Section 5 in order to derive estimators of some rare event probabilities through importance sampling.

In order to obtain a sharp approximation for $p_{nA}(\mathbf{X}_1^k = Y_1^k)$ it is necessary to introduce an interval $(a, a+c_n)$ which contains the principal part of the integral (39).

Let c_n denote a positive sequence such that the following condition (C) holds:

$$\lim_{n\to\infty} nc_n = \infty,$$

$$\sup_{n\geq 1} \frac{nc_n}{(n-k)} < \infty$$

and denote c the current term c_n.

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

$$g_{nA}(y_1^k) \tag{46}$$
$$:= \frac{nm^{-1}(a)\int_a^{a+c} g_{nv}(y_1^k)(\exp(-nm^{-1}(a)(v-a)))\,dv}{1-\exp(-nm^{-1}(a)c)}.$$

The density

$$\frac{nm^{-1}(a)(\exp(-nm^{-1}(a)(v-a)))\mathbb{1}_{(a,a+c)}(v)}{1-\exp(-nm^{-1}(a)c)}, \tag{47}$$

which appears in (46) approximates $p(\mathbf{U}_{1,n}/n = v | a < \mathbf{U}_{1,n}/n < a+c)$. Furthermore due to Theorem 8 $g_{nv}(Y_1^k)$ approximates $p_{nv}(Y_1^k)$ when Y_1^k results from sampling under P_{nA}. For a discussion on the maximal value of k for which a given relative accuracy is attained, see [6].

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

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

Denote (V) the condition

$$\sup_{n\geq 1} \sqrt{n}\int_a^\infty V'(v)\bigl(\exp\bigl(-nm^{-1}(a)(v-a)\bigr)\bigr)\,dv < \infty.$$

THEOREM 18. *Assume* (E1), (E2), (C), (V). *Then for any positive* $\delta < 1$:
(i)

$$p_{nA}\bigl(\mathbf{X}_1^k = Y_1^k\bigr) = g_{nA}\bigl(Y_1^k\bigr)\bigl(1 + o_{P_{nA}}(\delta_n)\bigr) \tag{48}$$

and (ii)

$$p_{nA}\bigl(\mathbf{X}_1^k = Y_1^k\bigr) = g_{nA}\bigl(Y_1^k\bigr)\bigl(1 + o_{G_{nA}}(\delta_n)\bigr), \tag{49}$$

where

$$\delta_n := \max\bigl(\varepsilon_n(\log n)^2, \bigl(\exp(-nc)\bigr)^\delta\bigr). \tag{50}$$

PROOF. See the Appendix. $\square$

REMARK. Most distributions used in statistics satisfy (V); numerous papers have focused on the properties of variance functions and classification of distributions; see, for example, [18] and references therein.

COROLLARY 19. *Under the hypotheses of Theorem 18 the total variation distance between P_{nA} and G_{nA} goes to 0 as n tends to infinity, that is,*

$$\lim_{n\to\infty}\int |p_{nA}(y_1^k)-g_{nA}(y_1^k)|\,dy_1^k=0.$$

5. Applications.

5.1. *Rao–Blackwellization of estimators.* This example illustrates the role of Theorem 8 in statistical inference; the conditioning event is local, in the range where $\lim_{n\to\infty} u_{1,n}/n=Eu(\mathbf{X})$.

In statistics the following situation is often encountered. A model $\mathcal{P}$ consists of a family of densities p_θ where the parameter θ is assumed to belong to $\mathbb{R}^d$, and a sample of i.i.d. r.v.'s $\mathbf{X}_1^n$ is observed, with each of the $\mathbf{X}_i$'s having density p_{θ_T} where θ_T is unknown; denote $X_1,\dots,X_n$ the observed data set. Let $\mathbf{U}_{1,n}:=u(\mathbf{X}_1)+\cdots+u(\mathbf{X}_n)$ and let $u_{1,n}:=u(X_1)+\cdots+u(X_n)$, which usually satisfies $\lim_{n\to\infty}u_{1,n}/n=Eu(\mathbf{X})$. A preliminary estimator $\widehat{\theta}(\mathbf{X}_1^n)$ is chosen, which may have the advantage of being easily computable, at the cost of having poor efficiency, approaching θ_T loosely in terms of the MSE. The famous Rao–Blackwell theorem asserts that the MSE of the conditional expectation of $\widehat{\theta}(\mathbf{X}_1^n)$ given the observed value $u_{1,n}$ of any statistic improves on the MSE of $\widehat{\theta}(\mathbf{X}_1^n)$. When $u_{1,n}$ is sufficient for θ the reduction is maximal, leading to the unbiased minimal variance estimator for θ_T when $\widehat{\theta}(\mathbf{X}_1^n)$ is unbiased (Lehmann–Scheffé theorem).

The conditional density $p_{u_{1,n}}(x_1^n):=p(\mathbf{X}_1^n=x_1^n|\mathbf{U}_{1,n}=u_{1,n})$ is usually unknown, and Rao–Blackwellization of estimators cannot be performed in many cases. Simulations of long runs of length $k=k_n$ under a proxy of $p_{u_{1,n}}(x_1^k)$ provide an easy way to improve the preliminary estimator, averaging values of $\widehat{\theta}((X_1^k)(l))_{1\le l\le L}$ where the samples $(X_1^k(l))$'s are obtained under the approximation of $p_{u_{1,n}}(x_1^k)$ and L runs are performed.

Consider the Gamma density

$$f_{\rho,\theta}(x):=\frac{\theta^{-\rho}}{\Gamma(\rho)}x^{\rho-1}\exp\left(-\frac{x}{\theta}\right)\qquad \text{for } x>0. \tag{51}$$

As ρ varies in $\mathbb{R}^+$ and θ is positive, the density belongs to an exponential family $\gamma_{r,\theta}$ with parameters $r:=\rho-1$ and θ, and sufficient statistics are $t(x):=\log x$ and $u(x):=x$, respectively, for r and θ. Given an i.i.d. sample $X_1^n:=(X_1,\dots,X_n)$ with density γ_{r_T,θ_T} the resulting sufficient statistics are, respectively, $t_{1,n}:=\log X_1+\cdots+\log X_n$ and $u_{1,n}:=X_1+\cdots+X_n$. We consider the parametic model $(\gamma_{r_T,\theta},\theta\ge 0)$ assuming r_T known.

Definition (29) shows that $g_{u_{1,n}}$ depends on the unknown parameter θ_T. It can be seen that $u_{1,n}$ is nearly sufficient for θ in $g_{u_{1,n}}$ in the sense that the value of

$g_{u_{1,n}}(X_1^k)$ does not vary when θ_T is substituted by any other value θ of the parameter and the X_i's are generated under any density $\gamma_{r_T,\theta'}$ (see [5]) this is indeed in agreement with the statement of Theorem 11. Hence on one hand , $u_{1,n}$ can be used to obtain improved estimators of θ_T and on the other hand, $g_{u_{1,n}}$ can be used to simulate samples distributed under a proxy of $p_{u_{1,n}}$ using any θ in lieu of θ_T in (29), as is done in the following procedure:

A first unbiased estimator of θ_T is chosen as

$$\widehat{\theta}_2 := \frac{X_1 + X_2}{2r_T}.$$

Given an i.i.d. sample X_1^n with density γ_{r_T,θ_T} the Rao–Blackwellized estimator of $\widehat{\theta}$ is defined as

$$\theta_{RB,2} := E(\widehat{\theta}_2 | \mathbf{U}_{1,n})$$

whose variance is less than $\operatorname{Var}\widehat{\theta}_2$.

Consider $k = 2$ in $g_{u_{1,n}}(y_1^k)$, and let (Y_1, Y_2) be distributed according to $g_{u_{1,n}}(y_1^2)$. Replicates of (Y_1, Y_2) induce an estimator of $\theta_{RB,2}$ for fixed $u_{1,n}$. Iterating on the simulation of the runs X_1^n produces for $n = 100$ an i.i.d. sample of $\theta_{RB,2}$'s from which $\operatorname{Var}\theta_{RB,2}$ is estimated. The resulting variance shows a net improvement with respect to the estimated variance of $\widehat{\theta}_2$. It is of some interest to investigate this gain in efficiency as the number of terms involved in $\widehat{\theta}_k$ increases together with k. As k approaches n the variance of $\widehat{\theta}_k$ approaches the Cramér–Rao bound. Figure 14 shows the decay of the variance of $\widehat{\theta}_k$. We note that whatever the value of k the estimated value of the variance of $\theta_{RB,k}$ is constant, and is quite close to the Cramér–Rao bound. This is indeed an illustration of Lehmann–Scheffé's theorem.

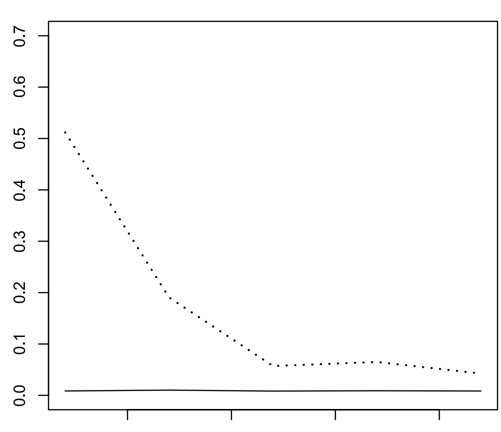

FIG. 14. *Variance of $\widehat{\theta}_k$, the initial estimator (dotted line), along with the variance of $\theta_{RB,k}$, the Rao–Blackwellized estimator (solid line) with $n = 100$ as a function of k.*

5.2. *Importance sampling for rare event probabilities.* Here we consider the application of the approximating scheme under a conditioning event defined through a large set, where this event is also on the large deviation scale. A development of the present section is presented in [6] and in Section 3 of [9]; see also [7]. Consider the estimation of the large deviation probability for the mean of n i.i.d. r.v.'s $u(\mathbf{X}_i)$ satisfying the conditions of this paper. This is a benchmark problem in the study of rare events; we refer to [8] for the background of this section.

Let $u_{1,n} := na$ for fixed a larger than $Eu(\mathbf{X})$. The probability to be estimated is

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

The importance sampling procedure substitutes the empirical estimator

$$\begin{aligned}\widehat{P}_n &:= \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}\big(\mathbf{U}_{1,n}(l) > u_{1,n}\big)\\ &= \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}\left(\sum_{i=1}^{n} u\big(\mathbf{X}_i(l)\big) > u_{1,n}\right)\end{aligned}\tag{52}$$

by

$$P_n^{\mathrm{IS},g} := \frac{1}{L}\sum_{l=1}^{L} \frac{p(u(\mathbf{X}_1(l)))\cdots p(u(\mathbf{X}_n(l)))}{g(u(\mathbf{X}_1(l))\cdots u(\mathbf{X}_n(l)))} \mathbb{1}\left(\sum_{i=1}^{n} u\big(\mathbf{X}_i(l)\big) > u_{1,n}\right).\tag{53}$$

In the above display (53) the sample $\mathbf{X}_1^n(l)$ is generated under i.i.d. sampling with distribution $P_{\mathbf{X}}$ and the L samples are i.i.d. In display (53) the sample $\mathbf{X}_1^n(l)$ is generated under the density g on $\mathbb{R}^n$ (under which the $\mathbf{X}_i$'s may not be independent). The L samples $\mathbf{X}_1^n(l)$ are i.i.d.

It is well known that the optimal sampling density is

$$p_{\mathrm{opt}}\big(x_1^n\big) := p\big(\mathbf{X}_1^n = x_1^n | \mathbf{U}_{1,n} > u_{1,n}\big),$$

which is not achievable since it presumes a known P_n. This optimal sampling density produces the zero variance estimator P_n itself with $L = 1$. However approximating $p_{\mathrm{opt}}(x_1^n)$ sharply at least on the first k coordinates for large k produces a large hit rate for the importance sampling procedure, and pushes the importance factor toward 1.

Define the sampling density g on $\mathbb{R}^n$ as

$$g\big(x_1^n\big) := g_{nA}\big(x_1^k\big) \prod_{i=k+1}^{n} \pi_u^a(x_i),$$

where g_{nA} is defined in (46), and π_u^a is the density defined in (23). The approximating density g_{nA} has been used to simulate the k first $\mathbf{X}_i$'s and the remaining $n-k$'s are i.i.d. with the classical tilted density. The classical IS scheme coincides

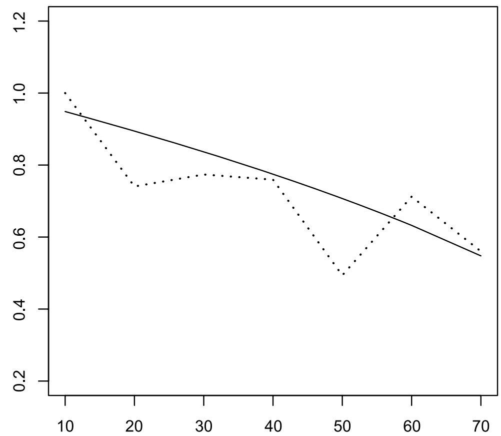

FIG. 15. *Ratio of the empirical value of the MSE of the adaptive estimate w.r.t. the empirical MSE of the i.i.d. twisted one* (*dotted line*) *along with the true value of this ratio* (*solid line*) *as a function of* k.

with the present one with the difference that $k = 1$ and $g_{A_n}(x_1) = \pi_u^a(x_1)$, that is, simulating under an i.i.d. sampling scheme with common density π_u^a.

Simulation under g_{nA} is performed through a double step procedure: In the first step, randomize the value of $\mathbf{U}_{1,n}/n$ on $(a, +\infty)$ according to a proxy of its distribution conditioned on $\mathbf{U}_{1,n} > na$; hence simulate a random variable S on $(a, +\infty)$ with density

$$p_S(s) := nm^{-1}(a_n)\bigl(\exp\bigl(-nm^{-1}(a)(s-a)\bigr)\bigr)\mathbb{1}_{(a,+\infty)}(s). \tag{54}$$

Then plug in nS in lieu of $u_{1,n}$ in (29) and iterate. This is equivalent to considering each point in the target set as a dominating point, weighted by its conditional density under $(\mathbf{U}_{1,n} > na)$. Simulation of S under (54) instead of (47) is slightly suboptimal but much simpler. It can be proved that the MSE of the estimate of P_n in this new IS sampling scheme is reduced by a factor $\sqrt{(n-k)/n}$ with respect to the classical scheme when calculated on large subsets of $\mathbb{R}^k$; see [6]. Figure 15 shows, in a simple case, 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 grows from $k = 0$ (i.i.d. twisted sample) to $k = 70$ (according to the rule presented in [6]). This ratio stabilizes to $\sqrt{n-k}/\sqrt{n}$ for $L = 2000$. The abscissa is k and the solid line is $k \to \sqrt{n-k}/\sqrt{n}$.

REMARK 20. In the present context, Dupuis and Wang [14] have shown that i.i.d. sampling schemes can produce "rogue paths" which may alter the properties of the estimate, and the estimation of its variance. They consider an i.i.d. random sample X_1^n where X_1 has a normal distribution $N(1, 1)$ and

$$\mathcal{E}_n := \left\{x_1^n : \frac{x_1 + \cdots + x_n}{n} \in A\right\},$$

where $A = (-\infty, a) \cup (b, +\infty)$ with $a < 1 < b$. The quantity to be estimated is $P(\mathcal{E}_n)$.

Assuming that $a + b < 2$, the standard i.i.d. IS scheme introduces the dominating point b and the family of i.i.d. tilted r.v.'s with common $N(b, 1)$ distribution. "Rogue paths" generated under $N(b, 1)$ may hit the set $(-\infty, a)$ with small probability under the sampling scheme, hence producing a very large importance factor. The resulting variance of the estimate is very sensitive with respect to these values, as exemplified in their Table 1, page 24. Simulation of paths according to G_{nS} with S defined in (54) produces their constructive samples which yield both a hit rate close to 100% and an importance factor close to $P(\mathcal{E}_n)$. We refer to [6] for discussion and examples. We also note that Dupis and Wang [14] propose an adaptive tilting scheme, based on the product of the π^{m_i}, $1 \le i \le n$, which yields an efficient IS algorithm.

APPENDIX

For clarity the current term a_n is denoted a in all proofs.

A.1. Three lemmas pertaining to the partial sum under its final value. We state three lemmas which describe some functions of the random vector $\mathbf{X}_1^n$ conditioned on $\mathcal{E}_n$. The r.v. $\mathbf{X}$ is assumed to have expectation 0 and variance 1.

LEMMA 21. *It holds that* $E_{P_{na}}(\mathbf{X}_1) = a, E_{P_{na}}(\mathbf{X}_1\mathbf{X}_2) = a^2 + O(\frac{1}{n})$, $E_{P_{na}}(\mathbf{X}_1^2) = s^2(t) + a^2 + O(\frac{1}{n})$ *where* $m(t) = a$.

PROOF. Using

$$p_{na}(\mathbf{X}_1 = x) = \frac{p_{\mathbf{S}_{2,n}}(na - x)p_{\mathbf{X}_1}(x)}{p_{\mathbf{S}_{1,n}}(na)} = \frac{\pi^a_{\mathbf{S}_{2,n}}(na - x)\pi^a_{\mathbf{X}_1}(x)}{\pi^a_{\mathbf{S}_{1,n}}(na)},$$

normalizing both $\pi^a_{\mathbf{S}_{2,n}}(na - x)$ and $\pi^a_{\mathbf{S}_{1,n}}(na)$ and making use of a first order Edgeworth expansion in those expressions yields $E_{P_{na}}(\mathbf{X}_1^2) = s^2(t) + a^2 + O(\frac{1}{n})$. A similar expansion for the joint density $p_{na}(\mathbf{X}_1 = x, \mathbf{X}_2 = y)$, with the same tilted distribution π^a produces the limit expression of $E_{P_{na}}(\mathbf{X}_1\mathbf{X}_2)$. $\square$

LEMMA 22. *Assume* (E1). *Then* (i) $\max_{1\le i\le k} |m_i| = a + o_{P_{na}}(\varepsilon_n)$. *Also* (ii) $\max_{1\le i\le k} s_i^2$, $\max_{1\le i\le k} \mu_3^i$ *and* $\max_{1\le i\le k} \mu_4^i$ *tend in* P_{na} *probability to the variance, skewness and kurtosis of* $\pi^{\underline{a}}$ *where* $\underline{a} := \lim_{n\to\infty} a_n$.

PROOF. (i) Define

$$\begin{aligned} V_{i+1} &:= m(t_i) - a \\ &= \frac{S_{i+1,n}}{n-i} - a. \end{aligned}$$

We state that

$$\max_{0\le i\le k-1} |V_{i+1}| = o_{P_{na}}(\varepsilon_n), \tag{55}$$

namely for all positive δ

$$\lim_{n\to\infty} P_{na}\Big(\max_{0\le i\le k-1} |V_{i+1}| > \delta\varepsilon_n\Big) = 0,$$

which we obtain following the proof of Kolmogorov maximal inequality. Define

$$A_i := \big(\big(|V_{i+1}| \ge \delta\varepsilon_n\big) \text{ and } \big(|V_j| < \delta\varepsilon_n \text{ for all } j < i+1\big)\big)$$

from which

$$\Big(\max_{0\le i\le k-1} |V_{i+1}| > \delta\varepsilon_n\Big) = \bigcup_{i=0}^{k-1} A_i.$$

It holds that

$$\begin{aligned}
E_{P_{na}} V_k^2 &= \int_{\cup A_i} V_k^2\, dP_{na} + \int_{(\cup A_i)^c} V_k^2\, dP_{na} \\
&\ge \int_{\cup A_i} \big(V_i^2 + 2(V_k - V_i)V_i\big)\, dP_{na} + \int_{(\cup A_i)^c} \big(V_i^2 + 2(V_k - V_i)V_i\big)\, dP_{na} \\
&\ge \int_{\cup A_i} V_i^2\, dP_{na} \\
&\ge \delta^2\varepsilon_n^2 \sum_{j=0}^{k-1} P_{na}(A_j) \\
&= \delta^2\varepsilon_n^2 P_{na}\Big(\max_{0\le i\le k-1} |V_{i+1}| > \delta\varepsilon_n\Big).
\end{aligned}$$

The third line above follows from $EV_i(V_k - V_i) = 0$ which is proved below. Hence

$$P_{na}\Big(\max_{0\le i\le k-1} |V_{i+1}| > \delta\varepsilon_n\Big) \le \frac{\mathrm{Var}_{P_{na}}(V_k)}{\delta^2\varepsilon_n^2} = \frac{1}{\delta^2\varepsilon_n^2(n-k)}\big(1 + o(1)\big),$$

where we used Lemma 21; therefore (55) holds under (E1). Direct calculation yields $E_{P_{na}}(V_i(V_k - V_i)) = 0$, which completes the proof of (i).

(ii) follows from (i) since $\lim_{n\to\infty} \max_{1\le i\le k} m(t_i) = \underline{a}$. $\square$

We also need the order of magnitude of $\max(|\mathbf{X}_1|, \dots, |\mathbf{X}_k|)$ under P_{na} which is stated in the following result.

LEMMA 23. *It holds that* $\max(|\mathbf{X}_1|, \dots, |\mathbf{X}_n|) = O_{P_{na}}(\log n)$.

PROOF. Set $|\mathbf{X}_1| := \mathbf{X}_i^- + \mathbf{X}_i^+$ with $\mathbf{X}_i^- := -\min(0, \mathbf{X}_i)$, $\mathbf{X}_i^+ := \max(0, \mathbf{X}_i)$; it is enough to prove that $\max_i \mathbf{X}_i^- = O_{P_{na}}(\log n)$ and $\max_i \mathbf{X}_i^+ = O_{P_{na}}(\log n)$. Since $E[\exp(t\mathbf{X})]$ is finite in a nonempty neighborhood of 0 so are $E[\exp(t\mathbf{X}^-)]$ and $E[\exp(t\mathbf{X}^+)]$. We hence prove the lemma for positive r.v.'s $\mathbf{X}_i$'s only.

Denote a the current term of the sequence a_n. For all t it holds that

$$
\begin{aligned}
P_{na}\bigl(\max(\mathbf{X}_1, \ldots, \mathbf{X}_n) > t\bigr) &\le n P_{na}(\mathbf{X}_n > t) \\
&= n \int_t^\infty \pi^a(\mathbf{X}_n = u) \frac{\pi^a(\mathbf{S}_{1,n-1} = na - u)}{\pi^a(\mathbf{S}_{1,n} = na)}\, du.
\end{aligned}
$$

Let τ be such that $m(\tau) = a$. Denote $s := s(\tau)$. Center and normalize both $\mathbf{S}_{1,n}$ and $\mathbf{S}_{1,n-1}$with respect to the density π^a in the last line above, denoting $\overline{\pi_n^a}$ the density of $\overline{\mathbf{S}_{1,n}} := (\mathbf{S}_{1,n} - na)/s\sqrt{n}$ when $\mathbf{X}$ has density π^a with mean a and variance s^2, we obtain

$$
\begin{aligned}
&P_{na}\bigl(\max(\mathbf{X}_1, \ldots, \mathbf{X}_n) > t\bigr) \\
&\quad\le n \frac{\sqrt{n}}{\sqrt{n-1}} \\
&\qquad \times \int_t^\infty \pi^a(\mathbf{X}_n = u) \frac{\overline{\pi_{n-1}^a}(\overline{\mathbf{S}_{1,n-1}} = (na - u - (n-1)a)/(s\sqrt{n-1}))}{\overline{\pi_n^a}(\overline{\mathbf{S}_{1,n}} = 0)}\, du.
\end{aligned}
$$

Under the sequence of densities π^a the triangular array $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ obeys a first order Edgeworth expansion

$$
\begin{aligned}
&P_{na}\bigl(\max(\mathbf{X}_1, \ldots, \mathbf{X}_n) > t\bigr) \\
&\quad\le n \frac{\sqrt{n}}{\sqrt{n-1}} \int_t^\infty \pi^a(\mathbf{X}_n = u) \frac{\mathfrak{n}((a-u)/s\sqrt{n-1})\mathbf{P}(u, i, n) + o(1)}{\mathfrak{n}(0) + o(1)}\, du \\
&\quad\le n Cst \int_t^\infty \pi^a(\mathbf{X}_n = u)\, du
\end{aligned}
$$

for some constant Cst independent of n and τ and

$$
\mathbf{P}(u, i, n) := 1 + P_3\bigl((a-u)/s\sqrt{n-1}\bigr),
$$

where $P_3(x) = \frac{\mu_3}{6s^3}(x^3 - 3x)$ is the third Hermite polynomial; s^2 and μ_3 are the second and third centered moments of π^a. We have used the fact that the sequence a converges to bound all moments of the tilted densities π^a. We used uniformity on u in the remaining term of the Edgeworth expansions. Making use of the Chernoff inequality to bound $\Pi^a(\mathbf{X}_n > t)$,

$$
P_{na}\bigl(\max(\mathbf{X}_1, \ldots, \mathbf{X}_n) > t\bigr) \le n Cst \frac{\Phi(t+\lambda)}{\Phi(t)} e^{-\lambda t}
$$

for any λ such that $\phi(t+\lambda)$ is finite. For t such that

$$
t/\log n \to \infty
$$

it holds that

$$P_{na}\big(\max(\mathbf{X}_1,\dots,\mathbf{X}_n) < t\big) \to 1,$$

which proves the lemma. □

A.2. Proof of the approximations resulting from Edgeworth expansions in Theorem 2. We complete the calculation leading to (15) and (16).

Set $Z_{i+1} := (m_i - Y_{i+1})/s_i\sqrt{n-i-1}$.

It then holds that

$$\begin{aligned}
&\overline{\pi_{n-i-1}}(Z_{i+1}) \\
&\quad = \mathfrak{n}(Z_{i+1})\left[\begin{array}{c} 1 + \dfrac{1}{\sqrt{n-i-1}}P_3(Z_{i+1}) + \dfrac{1}{n-i-1}P_4(Z_{i+1}) \\ + \dfrac{1}{(n-i-1)^{3/2}}P_5(Z_{i+1}) \end{array}\right] \\
&\qquad + O_{P_{na}}\left(\frac{P_5(Z_{i+1})}{(n-i-1)^{3/2}}\right).
\end{aligned} \tag{56}$$

We perform an expansion in $\mathfrak{n}(Z_{i+1})$ up to order 3, with a first order term $\mathfrak{n}(-Y_{i+1}/(s_i\sqrt{n-i-1}))$, namely

$$\begin{aligned}
&\mathfrak{n}(Z_{i+1}) \\
&\quad = \mathfrak{n}\big(-Y_{i+1}/(s_i\sqrt{n-i-1})\big) \\
&\qquad \times \left(\begin{array}{c} 1 + \dfrac{Y_{i+1}m_i}{s_i^2(n-i-1)} + \dfrac{m_i^2}{2s_i^2(n-i-1)}\left(\dfrac{Y_{i+1}^2}{s_i^2(n-i-1)} - 1\right) \\ + \dfrac{m_i^3}{6s_i^3(n-i-1)^{3/2}}\dfrac{\mathfrak{n}^{(3)}(Y^*/(s_i\sqrt{n-i-1}))}{\mathfrak{n}(-Y_{i+1}/(s_i\sqrt{n-i-1}))} \end{array}\right),
\end{aligned} \tag{57}$$

where $Y^* = \frac{1}{s_i\sqrt{n-i-1}}(-Y_{i+1} + \theta m_i)$ with $|\theta| < 1$.

Lemmas 22 and 23 provide the orders of magnitude of the random terms in the above displays when sampling under P_{na}.

Use those lemmas to obtain

$$\frac{Y_{i+1}m_i}{s_i^2(n-i-1)} = \frac{Y_{i+1}}{n-i-1}\big(a + o_{P_{na}}(\varepsilon_n)\big) \tag{58}$$

and

$$\frac{m_i^2}{s_i^2(n-i-1)} = \frac{1}{n-i-1}\big(a + o_{P_{na}}(\varepsilon_n)\big)^2.$$

Also when (E1) and (E2) holds, then the dominant terms in the bracket in (57) are precisely those in the two displays just above. This yields

$$\mathfrak{n}(Z_{i+1}) = \mathfrak{n}\left(\frac{-Y_{i+1}}{s_i\sqrt{n-i-1}}\right)\left(\begin{array}{c} 1 + \dfrac{aY_{i+1}}{s_i^2(n-i-1)} - \dfrac{a^2}{2s_i^2(n-i-1)} \\ + \dfrac{o_{P_{na}}(\varepsilon_n \log n)}{n-i-1} \end{array}\right).$$

We now need a precise evaluation of the terms in the Hermite polynomials in (56). This is achieved using Lemmas 22 and 23 which provide uniformity on i between 1 and $k = k_n$ in all terms depending on the sample path Y_1^k. The Hermite polynomials depend upon the moments of the underlying density π^{m_i}. Since $\overline{\pi_1^{m_i}}$ has expectation 0 and variance 1 the terms corresponding to P_1 and P_2 vanish. For up to order 4 polynomials, write $P_3(x) = \frac{\mu_3^{(i)}}{6(s_i)^3} H_3(x)$, $P_4(x) = \frac{(\mu_3^i)^2}{72(s_i)^6} H_6(x) + \frac{\mu_4^{(i,n)} - 3(s_i)^4}{24(s_i)^4} H_4(x)$ with $H_3(x) := x^3 - 3x$, $H_4(x) := x^4 - 6x^2 + 3$ and $H_6(x) := x^6 - 15x^4 + 45x^2 - 15$.

Using Lemma 22 it appears that the terms in x^j, $j \geq 3$ in P_3 and P_4 will play no role in the asymptotic behavior in (56) with respect to the constant term in P_4 and the term in x from P_3. Indeed substituting x by Z_{i+1} and dividing by $n-i-1$, the term in x^2 in P_4 is $O_{P_{na}}(\log n)^2/(n-i)^2$ where we have used Lemma 22. These terms are of smaller order than the term $-3x$ in P_3 which is $-\frac{\mu_3^i}{2s_i^4(n-i-1)}(a - Y_{i+1}) = \frac{1}{n-i-1} O_{P_{na}}(\log n)$.

It holds that

$$\frac{P_3(Z_{i+1})}{\sqrt{n-i-1}} = -\frac{\mu_3^i}{2s_i^4(n-i-1)}(m_i - Y_{i+1}) + \frac{\mu_3^i(m_i - Y_{i+1})^3}{6(s_i)^6(n-i-1)^2},$$

which yields

$$\frac{P_3(Z_{i+1})}{\sqrt{n-i-1}} = -\frac{\mu_3^i}{2s_i^4(n-i-1)}(a - Y_{i+1}) + \frac{O_{P_{na}}(\log n)^3}{(n-i-1)^2}. \tag{59}$$

For the term of order 4 it holds that

$$\frac{P_4(Z_{i+1})}{n-i-1} = \frac{1}{n-i-1}\left(\frac{(\mu_3^i)^2}{72s_i^6} H_6(Z_{i+1}) + \frac{\mu_4^i - 3s_i^4}{24s_i^4} H_4(Z_{i+1})\right),$$

which yields

$$\frac{P_4(Z_{i+1})}{n-i-1} = \frac{\mu_4^i - 3s_i^4}{8s_i^4(n-i-1)} - \frac{15(\mu_3^i)^2}{72s_i^6(n-i-1)} + \frac{O_{P_{na}}((\log n)^2)}{(n-i-1)^2}. \tag{60}$$

The fifth term in the expansion plays no role in the asymptotics.

In summary, comparing the remainder terms in (59) and (60), we obtain

$$\overline{\pi_{n-i-1}}(Z_{i+1}) = \mathfrak{n}\big(-Y_{i+1}/(s_i\sqrt{n-i-1})\big)\cdot A\cdot B + O_{P_{na}}\left(\frac{P_5(Z_{i+1})}{(n-i-1)^{3/2}}\right),$$

where A and B are given in (15) and (16).

A.3. Final step of the proof of Theorem 2. We make use of the following version of the law of large numbers for triangular arrays; see [21] Theorem 3.1.3.

THEOREM 24. *Let $X_{i,n}$, $1 \le i \le k$ denote an array of row-wise real exchangeable r.v.'s and $\lim_{n\to\infty} k = \infty$. Let $\rho_n := EX_{1,n}X_{2,n}$. Assume that for some finite Γ, $EX_{1,n}^2 \le \Gamma$. If for some doubly indexed sequence $(a_{i,n})$ such that $\lim_{n\to\infty}\sum_{i=1}^k a_{i,n}^2 = 0$ it holds that*

$$\lim_{n\to\infty} \rho_n \left(\sum_{i=1}^k a_{i,n}^2\right)^2 = 0$$

and then

$$\lim_{n\to\infty} \sum_{i=1}^k a_{i,n} X_{i,n} = 0$$

in probability.

Denote

$$\kappa_1^i := \frac{\mu_3^i}{2s_i^4}, \qquad \kappa_2^i := \frac{\mu_4^i - 3s_i^4}{8s_i^4} + \frac{15(\mu_3^i)^2}{72s_i^6},$$
$$\mu_1^* := \kappa_1^i + \frac{a}{s_i^2}, \qquad \mu_2^* := \kappa_1^i - \frac{a}{2s_i^2}.$$

By (13), (14) and (17)

$$\begin{aligned} &p(\mathbf{X}_{i+1} = Y_{i+1} | S_{i+1,n} = na - S_{1,i}) \\ &\quad = \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1}) \frac{\mathfrak{n}(-Y_{i+1}/(s_i\sqrt{n-i-1}))}{\mathfrak{n}(0)} A(i) \end{aligned}$$

with

$$\begin{aligned} A(i) &:= \left(1 + \frac{\mu_1^* Y_{i+1}}{n-i-1} - \frac{\mu_2^* a}{n-i-1} - \frac{\kappa_2^i}{n-i-1} + \frac{o_{P_{na}}(\varepsilon_n \log n)}{n-i-1}\right) \\ &\qquad \Big/ \left(1 - \frac{\kappa_2^i}{n-i} + O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right)\right). \end{aligned}$$

We perform a second order expansion in both the numerator and the denominator of the above expression, which yields

$$A(i) = \exp\left(\frac{\mu_1^* Y_{i+1}}{n-i-1} - \frac{a}{2s_i^2(n-i-1)} - \frac{a\kappa_1^i}{n-i-1} + \frac{o_{P_{na}}(\varepsilon_n \log n)}{n-i-1}\right) A'(i). \tag{61}$$

The term $\exp(\frac{\mu_1^* Y_{i+1}}{n-i-1} + \frac{a}{2s_i^2(n-i-1)})$ in (61) is captured in $g(Y_{i+1}|Y_1^i)$.

The term $A'(i)$ in (61) is expressed as

$$A'(i) := Q_1^i \cdot Q_2^i$$

with

$$Q_1^i := \exp\left(-\left(\frac{\kappa_2^i}{(n-i-1)(n-i)} + \frac{(\kappa_2^i)^2}{2(n-i)^2} + \frac{1}{2}\left(\frac{\mu_1^* Y_{i+1}}{n-i-1} - \frac{a\mu_2^*}{n-i-1} - \frac{\kappa_2^i}{n-i-1}\right)^2\right)\right)$$

and

$$Q_2^i := \frac{\exp(B_1)}{\exp(B_2)},$$

where

$$B_1 := \frac{o_{P_{na}}(\varepsilon_n^2(\log n)^2)}{(n-i-1)^2} + \frac{\mu_1^* Y_{i+1}}{(n-i-1)^2} o_{P_{na}}(\varepsilon_n \log n) + \frac{\mu_2^* a}{(n-i-1)^2} o_{P_{na}}(\varepsilon_n \log n) + \frac{o_{P_{na}}(\varepsilon_n^2(\log n))^2}{(n-i-1)^2} + o(u_1^2),$$

$$B_2 := \frac{\kappa_2^i}{n-i} O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right) + O_{P_{na}}\left(\frac{1}{(n-i)^3}\right) + O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right) + o\left(\left(\frac{\kappa_2^i}{n-i} + O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right)\right)^2\right)$$

with

$$u_1 = \frac{\mu_1^* Y_{i+1}}{n-i-1} - \frac{\mu_2^* a}{n-i-1} - \frac{\kappa_2^i}{n-i-1} + \frac{o_{P_{na}}(\varepsilon_n \log n)}{n-i-1}.$$

We first prove that

$$\prod_{i=0}^{k-1} A'(i) = 1 + o_{P_{na}}(\varepsilon_n (\log n)^2) \tag{62}$$

as n tends to infinity.

Since

$$p(\mathbf{X}_1^k = Y_1^k | S_{i+1}^n = na) = g_0(Y_1|Y_0) \prod_{i=0}^{k-1} g(Y_{i+1}|Y_1^i) \prod_{i=0}^{k-1} A'(i) \prod_{i=0}^{k-1} L_i,$$

where

$$L_i := \frac{C_i^{-1}}{\Phi(t_i)} \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \exp\left(-\frac{a\kappa_1^i}{n-i-1}\right),$$

the completion of the proof will follow from

$$\prod_{i=0}^{k-1} L_i = 1 + o_{P_{na}}(\varepsilon_n (\log n)^2). \tag{63}$$

The proof of (62) is achieved in two steps.

CLAIM 25. $\prod_{i=0}^{k-1} Q_1^i = 1 + o_{P_{na}}(\varepsilon_n (\log n)^2)$.

By Lemma 22 the random terms μ_j^i deriving from π^{m_i} satisfy

$$\max_{1 \le i \le k} |\mu_j^i - \mu_j| = o_{P_{na}}(1)$$

as n tends to ∞, where μ_j is the jth cumulant of $\pi^{\underline{a}}$ where $\underline{a} := \lim_{n\to\infty} a$ is finite. Therefore we may substitute μ_j^i by μ_j in order to check the convergence of all subsequent series.

Expanding Q_1 define, for any positive β_1, β_2, β_3 and β_4

$$A_n^1 := \left\{ \frac{1}{\varepsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{\kappa_2^i}{(n-i-1)(n-i)} \right| < \beta_1 \right\},$$

$$A_n^2 := \left\{ \frac{1}{\varepsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{(\kappa_2^i)^2}{(n-i-1)^2} \right| < \beta_2 \right\},$$

$$A_n^3 := \left\{ \frac{1}{\varepsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{(\mu_2^* a)^2}{(n-i-1)^2} \right| < \beta_3 \right\}$$

and

$$A_n^4 := \left\{ \frac{1}{\varepsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{\mu_2^* \kappa_2^i a}{(n-i-1)^2} \right| < \beta_4 \right\}.$$

It clearly holds that

$$\lim_{n\to\infty} P_{na}(A_n^j) = 1, \qquad j = 1, \dots, 4.$$

Let for any positive β_5,

$$A_n^5 := \left\{ \frac{1}{\varepsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{\kappa_1^i \kappa_2^i Y_{i+1}}{(n-i-1)^2} \right| < \beta_5 \right\}.$$

If $\lim_{n\to\infty} P_{na}(A_n^5) = 1$, then $\lim_{n\to\infty} P_{na}(A_n^j)$, $j = 6, 7$ where

$$A_n^6 := \left\{ \frac{1}{\varepsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{\mu_1^* \kappa_2^i Y_{i+1}}{(n-i-1)^2} \right| < \beta_6 \right\},$$

$$A_n^7 := \left\{ \frac{1}{\varepsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{\mu_1^* \mu_2^* a Y_{i+1}}{(n-i-1)^2} \right| < \beta_7 \right\}.$$

Apply Theorem 24 with $X_{i,n} = Y_{i+1}$ and $a_{i,n} = \frac{1}{\varepsilon_n (\log n)^2 (n-i-1)^2}$. By Lemma 21,

$$E_{P_{na}} Y_1^2 = s^2(0) + a + O\left(\frac{1}{n}\right).$$

Hence $E_{P_{na}}[Y_1^2] \le \Gamma$ for some finite Γ. Furthermore $\rho_n = a^2 + O(\frac{1}{n})$. Both conditions in Theorem 24 are fullfilled. Indeed,

$$\lim_{n\to\infty} \sum_{i=1}^{k} a_{n,i}^2 = \lim_{n\to\infty} \frac{1}{\varepsilon_n^2 (\log n)^4 (n-k)^3} = 0,$$

which holds under (E1), as holds

$$\lim_{n\to\infty} \rho_n \left(\sum_{i=1}^{k} a_{n,i} \right)^2 = \lim_{n\to\infty} \frac{a^2}{\varepsilon_n^2 (\log n)^4 (n-k)^2} = 0.$$

Therefore, for $i = 5, 6, 7$

$$\lim_{n\to\infty} P_{na}(A_n^i) = 1.$$

Define for any positive β_8,

$$A_n^8 := \left\{ \frac{1}{\varepsilon_n (\log n)^2} \sum_{i=0}^{k-1} \frac{(\mu_1^*)^2 Y_{i+1}^2}{(n-i-1)^2} < \beta_8 \right\}.$$

Apply Theorem 24 with $X_{i,n} = Y_{i+1}^2$ and $a_{i,n} = \frac{1}{\varepsilon_n (\log n)^2 (n-i-1)^2}$.
The following holds:

$$\lim_{n\to\infty} \sum_{i=1}^{k} a_{n,i}^2 = 0$$

when (E1) holds.

By Lemma 21,

$$E_{P_{na}} Y_1^4 = E_{\pi^a} Y_1^4 + O\left(\frac{1}{n}\right),$$

which entails that such that $EY_1^4 \le \Gamma < \infty$ for some Γ. Also

$$E_{P_{na}}(Y_1^2 Y_2^2) = (s^2(0)+a)(s^2(0)+a) + O\left(\frac{1}{n}\right)$$

and

$$\lim_{n\to\infty} \rho_n \left(\frac{1}{\varepsilon_n(\log n)^2} \sum_{i=0}^{k-1} \frac{1}{(n-i-1)^2}\right)^2 = 0$$

under (E1). Hence

$$\lim_{n\to\infty} P_{na}(A_n^8) = 1.$$

It follows that, noting that A_n is the intersection of the events A_n^i, $j = 1, \dots, 8$

$$\lim_{n\to\infty} P_{na}(A_n) = 1.$$

To summarize, we have proved that, under (E1),

$$Q_1 = 1 + o_{P_{na}}(\varepsilon_n(\log n)^2).$$

CLAIM 26. $\prod_{i=0}^{k-1} Q_2^i = 1 + o_{P_{na}}(\varepsilon_n(\log n)^2)$.

This is equivalent to proving that the sum of the terms in B_1 (resp., in B_2) is of order $o_{P_{na}}(\varepsilon_n(\log n)^2)$.

The four terms in the sum of the terms in B_1 are, respectively, of order $o_{P_{na}}(\varepsilon_n^2(\log n)^4)/(n-k)$, $o_{P_{na}}(\varepsilon_n(\log n)^3)/(n-k)$, $o_{P_{na}}(a\varepsilon_n(\log n)^2)/(n-k)$ and $o_{P_{na}}(\varepsilon_n(\log n)^2)/(n-k)$ using Lemma 22. The sum of the terms $o(u_1^2)$ is of order less than these. Assuming (E1) all these terms are $o_{P_{na}}(\varepsilon_n(\log n)^2)$.

For the sum of terms in B_2, by uniformity of the Edgeworth expansion with respect to Y_1^k it holds that $\sum_{i=1}^k B_2 = O_{P_{na}}((n-k)^{-1/2}) = o_{P_{na}}(\varepsilon_n(\log n)^2)$ by (E1).

We now turn to the proof of (63).

Define

$$u := -x \frac{\mu_3^i}{2s_i^4(n-i-1)} + \frac{(x-a)^2}{2s_i^2(n-i-1)}.$$

Use the classical bounds

$$1 - u + \frac{u^2}{2} - \frac{u^3}{6} \le e^{-u} \le 1 - u + \frac{u^2}{2}$$

to obtain on both sides of the above inequalities the second order approximation of C_i^{-1} through integration with respect to p. The upper bound yields

$$C_i^{-1} \leq \Phi(t_i) + \frac{\kappa_1^i}{n-i-1}\Phi'(t_i) + \frac{1}{s_i^2(n-i-1)}(\Phi''(t_i) - 2a\Phi'(t_i) + a^2) + O_{P_{na}}\left(\frac{1}{(n-i-1)^2}\right)$$

from which

$$L_i \leq \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \exp\left(-\frac{a\kappa_1^i}{n-i-1}\right) \times \left(\begin{array}{c} 1 + \frac{\kappa_1^i}{n-i-1} m_i \\ -\frac{s_i^2 + m_i^2 - 2am_i + a^2}{2s_i^2(n-i-1)} + O_{P_{na}}\left(\frac{1}{(n-i-1)^2}\right) \end{array}\right),$$

where the approximation term is uniform on the Y_1^k.

Substituting $\frac{\sqrt{n-i}}{\sqrt{n-i-1}}$ and $\exp(-\frac{a\kappa_1^i}{n-i-1})$ by their expansions $1 + \frac{1}{2(n-i-1)} + O(\frac{1}{(n-i-1)^2})$ and $1 - \frac{a\kappa_1^i}{n-i-1} + \frac{(a\kappa_1^i)^2}{(n-i-1)^2} + O(\frac{a^2}{(n-i-1)^2})$ in the upper bound of L_i above yields

$$L_i \leq \left(1 + \frac{1}{2(n-i-1)} - \frac{a\kappa_1^i}{n-i-1} + \frac{(a\kappa_1^i)^2}{2(n-i-1)^2} + o\left(\frac{1}{(n-i-1)^2}\right)\right) \times \left(1 + \frac{\kappa_1^i m_i}{n-i-1} - \frac{s_i^2 + m_i^2 - 2am_i + a^2}{2s_i^2(n-i-1)} + O_{P_{na}}\left(\frac{1}{(n-i-1)^2}\right)\right).$$

Using Lemma 22, $m_i^2 - 2am_i + a^2 = o_{P_{na}}(a\varepsilon_n)$ and therefore

$$L_i \leq \left(1 + \frac{1}{2(n-i-1)} - \frac{a\kappa_1^i}{n-i-1} + \frac{(a\kappa_1^i)^2}{(n-i-1)^2} + o\left(\frac{1}{(n-i-1)^2}\right)\right) \times \left(1 + \frac{\kappa_1^i a}{n-i-1} - \frac{1}{2(n-i-1)} + \frac{o_{P_{na}}(a\varepsilon_n)}{n-i-1}\right).$$

Write

$$\prod_{i=1}^k L_i \leq \prod_{i=1}^k (1 + M_i)$$

with

$$M_i = \frac{(a\kappa_1^i)^2}{(n-i-1)^2} + \frac{o_{P_{na}}(a\varepsilon_n)}{n-i-1}.$$

Under (E1), $\sum_{i=0}^{k-1} M_i$ is $o_{P_{na}}(\varepsilon_n(\log n)^2)$. This completes the proof of the theorem.

A.4. Proof of Theorem 18. The following lemma (see [17], Corollary 6.4.1) provides an asymptotic formula for the tail probability of $\mathbf{U}_{1,n}$ under the hypotheses and notation of Section 3. Define

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

LEMMA 27. *Under the same hypotheses as above,*

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

where $\psi(a) := t^a s(t^a)$.

LEMMA 28. *Suppose that* (V) *holds. Then* (i) $E_{P_{nA}}\mathbf{U}_1 = a + o(1)$, (ii) $E_{P_{nA}}\mathbf{U}_1^2 = 1 + s^2(t) + o(1)$ *and* (iii) $E_{P_{nA}}\mathbf{U}_1\mathbf{U}_2 = a^2 + o(1)$ *where* $m(t) = a$.

PROOF. It holds that

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

Integration by parts yields

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

Using Lemma 27 and the Chernoff inequality,

$$\begin{aligned}
&\int_a^\infty P(\mathbf{U}_{1,n}/n > v|\mathbf{U}_{1,n} > na)\, dv \\
&\qquad \le \sqrt{2\pi}\psi(a)\sqrt{n}\int_a^\infty \exp\big(n\big(I_{\mathbf{U}}(a) - I_{\mathbf{U}}(v)\big)\big)\, dv,
\end{aligned}$$

where $\psi(a) = ts(t)$.

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

$$\int_a^\infty P(\mathbf{U}_{1,n}/n > v|\mathbf{U}_{1,n} > na)\, dv \le \frac{\sqrt{2\pi}\psi(a)}{\sqrt{n}I_{\mathbf{U}}^{'}(a)}.$$

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

Insert $E_{P_{nv}}\mathbf{U}_1^2 = v^2 + s_{\mathbf{U}}^2(t) + O(\frac{1}{n})$ into

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

First, via integration by parts, Lemma 13 and the Chernoff inequality,

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

Second,

$$\int_a^\infty V(v)p(\mathbf{U}_{1,n}/n = v|\mathbf{U}_{1,n} > na)\,dv$$
$$= s^2(t) + 2\int_a^\infty V'(v)P(\mathbf{U}_{1,n}/n > v|\mathbf{U}_{1,n} > na)\,dv,$$

which tends to $s^2(t)$ as $n \to \infty$ using again the Chernoff inequality, condition (V) and Lemma 13.

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

The proof of (iii) is similar to the above. □

Lemma 28 yields the maximal inequality stated in Lemma 22 under the condition $(\mathbf{U}_{1,n} > na)$. We also need the order of magnitude of the maximum of $(|\mathbf{U}_1|, \dots, |\mathbf{U}_k|)$ under P_{nA} which is stated in the following result.

LEMMA 29. *It holds that*

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

PROOF. Using the same argument as in Lemma 23 we consider the case when the r.v.'s $\mathbf{U}_i$ take nonnegative values. We prove that

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

when

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

For fixed d it holds that

$$\begin{aligned}
&P_{nA}\big(\max(\mathbf{U}_1, \dots, \mathbf{U}_n) > t_n\big)\\
&\quad= \int_a^{a+d} P\big(\max(\mathbf{U}_1, \dots, \mathbf{U}_n) > t_n|\mathbf{U}_{1,n}/n = v\big)\\
&\qquad\qquad \times p(\mathbf{U}_{1,n}/n = v|\mathbf{U}_{1,n}/n > a)\,dv\\
&\qquad + \int_{a+d}^\infty P\big(\max(\mathbf{U}_1, \dots, \mathbf{U}_n) > t_n|\mathbf{U}_{1,n}/n = v\big)\\
&\qquad\qquad \times p(\mathbf{U}_{1,n}/n = v|\mathbf{U}_{1,n}/n > a)\,dv\\
&\quad =: I + II.
\end{aligned}$$

Now

$$II \le \frac{P(\mathbf{U}_{1,n}/n > a + d)}{P(\mathbf{U}_{1,n}/n > a)},$$

which tends to 0 by Lemma 27.

Furthermore by Lemma 23, $\lim_{n\to\infty} P(\max(\mathbf{U}_1,\dots,\mathbf{U}_n) > t_n|\mathbf{U}_{1,n}/n = v) =: \lim_{n\to\infty} r_n = 0$ when $v \in (a, a+d)$. Hence

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

This proves the lemma. □

We now prove (48).

Step 1. We first prove that the integral (39) can be reduced to its principal part, namely that

$$\begin{aligned} p_{nA}\bigl(Y_1^k\bigr) &= \bigl(1+o_{P_{nA}}(1)\bigr) \\ &\quad\times \int_a^{a+c} p\bigl(\mathbf{X}_1^k = Y_1^k|\mathbf{U}_{1,n}/n = v\bigr) p(\mathbf{U}_{1,n}/n = v|\mathbf{U}_{1,n} > na)\,dv \end{aligned} \tag{64}$$

holds for any fixed $c > 0$.

Apply Bayes's formula to obtain

$$\begin{aligned} p_{nA}\bigl(Y_1^k\bigr) &= \frac{np_{\mathbf{X}}(Y_1^k)}{(n-k)} \\ &\quad\times \frac{\int_a^\infty p(\mathbf{U}_{k+1,n}/(n-k) = n/(n-k)(t - k\overline{U_{1,k}}/n))\,dt}{P(\mathbf{U}_{1,n} > na)}, \end{aligned}$$

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

Denote

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

with

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

Then (64) holds whenever $I \to 0$ (under P_{nA}).

Under P_{nA} it holds that

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

A similar result as Lemma 22 holds under condition $(\mathbf{U}_{1,n} > na)$, using Lemma 28; namely it holds that

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

Using both results

$$m_k = a + O_{P_{nA}}(v_n) \tag{65}$$

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

We now prove that $I \to 0$. Using once more Lemma 27 yields

$$I = \frac{m^{-1}(m_k)s(m^{-1}(m_k))}{m^{-1}(m_k + nc/(n-k))s(m^{-1}(m_k + nc/(n-k)))} \times \exp\biggl(-(n-k)\biggl(I_{\mathbf{U}}\biggl(m_k + \frac{nc}{n-k}\biggr) - I_{\mathbf{U}}(m_k)\biggr)\biggr).$$

Now by convexity of the function $I_{\mathbf{U}}$

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

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

We have proved that

$$I = O_{P_{nA}}(\exp(-nc)).$$

Step 2. We claim that (48) holds uniformly in v in $(a, a+c)$ when Y_1^k is generated under P_{nA}. This result follows from a similar argument as used in Theorem 8 where (48) is proved under the local sampling P_{nv}. A close look at the proof shows that (48) holds whenever Lemmas 22 and 23, stated for the variables $\mathbf{U}_i$'s instead of $\mathbf{X}_i$'s hold under P_{nA}. Those lemmas are substituted by Lemmas 28 and 29 here above.

Inserting (48) in (64) yields

$$p_{nA}(Y_1^k) = \biggl(\int_a^{a+c} g_{nv}(Y_1^k) p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} > na)\, dv\biggr) \times (1 + o_{p_{nA}}(\max(\varepsilon_n(\log n)^2, (\exp(-nc))^{\delta})))$$

for some $\delta < 1$.

The conditional density of $\mathbf{U}_{1,n}/n$ given $(\mathbf{U}_{1,n} > na)$ is stated in (45) which holds uniformly in v on $(a, a+c)$.

In summary we have proved

$$p_{nA}(Y_1^k) = \biggl(nm^{-1}(a) \int_a^{a+c} g_{nv}(Y_1^k) \exp(-nm^{-1}(a)(v-a))\, dv\biggr) \times (1 + o_{p_{nA}}(\max(\varepsilon_n(\log n)^2, (\exp(-nc))^{\delta})))$$

as $n \to \infty$ for any positive $\delta < 1$.

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

$$nm^{-1}(a)\int_a^{a+c} g_{nv}(Y_1^k)\exp(-nm^{-1}(a)(v-a))\,dv$$
$$=1+o_{P_{nA}}(\exp(-nc))$$

as $n\to\infty$ which completes the proof of (48). The proof of (49) follows from (48) and Lemma 6.

Acknowledgements. The authors thank the referee for his careful reading of the paper and for comments that considerably improved the presentation of this work. Also the authors thank Dr. Tarn Duong for his help and discussions.

REFERENCES

[1] BARBE, P. and BRONIATOWSKI, M. (1999). Simulation in exponential families. *ACM Trans. Model. Comput. Simul.* **9** 203–223.

[2] BARBE, P. and BRONIATOWSKI, M. (2000). Large-deviation probability and the local dimension of sets. *J. Math. Sci.* **99** 1225–1233.

[3] BARBE, P. and BRONIATOWSKI, M. (2004). On sharp large deviations for sums of random vectors and multidimensional Laplace approximation. *Teor. Veroyatn. Primen.* **49** 743–774. MR2142565

[4] BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. MR0489333

[5] BRONIATOWSKI, M. and CARON, V. (2012). Conditional inference in parametric models. In *Statistical Models and Methods for Reliability and Survival Analysis* (L. Gerville-Reache, C. Huber, N. Limnios, M. Mesbah and V. Couallier, eds.). Wiley, New York.

[6] BRONIATOWSKI, M. and CARON, V. (2013). Small variance estimators for rare event probabilities. *ACM Trans. Model. Comput. Simul.* **23** Art. 7, 23. MR3034217

[7] BRONIATOWSKI, M. and RITOV, Y. (2009). Importance sampling for rare events and conditioned random walks. Available at arXiv:0910.1819.

[8] BUCKLEW, J. A. (2004). *Introduction to Rare Event Simulation*. Springer, New York. MR2045385

[9] CARON, V., GUYADER, A., MUNOZ ZUNIGA, M. and TUFFIN, B. (2014). Some recent results in rare event estimation. *ESAIM Proc.* **44** 239–259.

[10] CSISZÁR, I. (1984). Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Probab.* **12** 768–793. MR0744233

[11] DEMBO, A. and ZEITOUNI, O. (1996). Refinements of the Gibbs conditioning principle. *Probab. Theory Related Fields* **104** 1–14. MR1367663

[12] DEN HOLLANDER, W. T. F. and WEISS, G. H. (1988). On the range of a constrained random walk. *J. Appl. Probab.* **25** 451–463. MR0954494

[13] DIACONIS, P. and FREEDMAN, D. A. (1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. *J. Theoret. Probab.* **1** 381–410. MR0958245

[14] DUPUIS, P. and WANG, H. (2004). Importance sampling, large deviations, and differential games. *Stoch. Stoch. Rep.* **76** 481–508. MR2100018

[15] ERMAKOV, M. S. (2006). The importance sampling method for modeling the probabilities of moderate and large deviations of estimates and tests. *Teor. Veroyatn. Primen.* **51** 319–332. MR2324204

[16] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*. *Vol*. *II*, 2nd ed. Wiley, New York. MR0270403

[17] JENSEN, J. L. (1995). *Saddlepoint Approximations*. *Oxford Statistical Science Series* **16**. Oxford Univ. Press, New York. MR1354837

[18] LETAC, G. and MORA, M. (1990). Natural real exponential families with cubic variance functions. *Ann*. *Statist*. **18** 1–37. MR1041384

[19] RIHTER, V. (1957). Local limit theorems for large deviations. *Dokl*. *Akad*. *Nauk SSSR* (*N.S.*) **115** 53–56. MR0093816

[20] SUNDBERG, R. (2010). Flat and multimodal likelihoods and model lack of fit in curved exponential families. *Scand*. *J*. *Stat*. **37** 632–643. MR2779640

[21] TAYLOR, R. L., DAFFER, P. Z. and PATTERSON, R. F. (1985). *Limit Theorems for Sums of Exchangeable Random Variables*. Rowman & Allanheld, Totowa, NJ. MR0860208

[22] VAN CAMPENHOUT, J. M. and COVER, T. M. (1981). Maximum entropy and conditional probability. *IEEE Trans*. *Inform*. *Theory* **27** 483–489. MR0635527

[23] WEISS, G. H. and DEN HOLLANDER, W. T. F. (1988). A note on configurational properties of constrained random walks. *J*. *Phys*. *A* **21** 2405–2415. MR0953218

LSTA
UNIVERSITÉ PIERRE ET MARIE CURIE—PARIS 6
75005 PARIS
FRANCE
E-MAIL: michel.broniatowski@upmc.fr
virgile.caron@upmc.fr