Rare Event Probability Estimation Using Information Projection

Anh Thu Phan Ho, Wadih Sawaya, Patrick Bas

Rare Event Probability Estimation Using Information Projection

Anh Thu Phan Ho [§*1], Wadih Sawaya [*2], Patrick Bas [#3]

[#] *Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France*

[1] anh-thu.phan-ho@gipsa-lab.grenoble-inp.fr

[2] Wadih.Sawaya@telecom-lille.fr

[3] Patrick.Bas@ec-lille.fr

[§] *Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble, France CNRS, GIPSA-Lab, F-38000 Grenoble, France*

[*] *Télécom Lille, Univ. Lille, CNRS, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France*

***Abstract*—This paper proposes an Importance Sampling (IS) simulation to estimate the rare event probabilities which satisfy a large deviation principle. The sampling distribution of the IS is constructed using the information projection. With this construction of the sampling distribution, the importance sampling simulation is showed to be asymptotically efficient. A specific example is presented to illustrate the IS technique.**

I. Introduction

Estimating rare event probabilities with a reasonable degree of accuracy has become very important due to its applications in various areas such as financial risk, insurance, telecommunication, computational physics [4]. It is yet still a challenging problem since the occurrence of the rare event is extremely small. The use of a standard Monte Carlo (MC) is completely inadequate as it would require a very large number of trials. Importance sampling (IS) is one of MC simulation methods to reduce this computational burden. The basic idea of IS is to sample from an alternative sampling distribution which is the so-called change of measure so that the considered event occurs more frequently.

The most difficult problem in IS is to find a good sampling distribution in the sense that it minimizes the variance of the estimator. It is worth noting that the IS could also backfire if the variance of the estimator of IS is greater than that of the standard MC. Thus it is crucial to find an efficient and better IS comparing to the standard MC. There are two efficiency criteria for IS MC simulation algorithms: *bounded relative error* criterion and *logarithmic asymptotic efficiency* or *asymptotic optimality* criterion [1], [5]. The asymptotic optimality criterion is particularly convenient when the rare event probabilities satisfy the large deviation asymptotic which decays exponentially. In [5], the authors considered the estimation probability of a rare event known as the probability of large deviation. They proposed the IS by constructing the "exponentially twisted" distribution as a sampling function to obtain the asymptotically efficient simulation.

In this short paper, we also utilize IS to estimate the probability of rare event which satisfies a large deviation principle. However, the sampling distribution is chosen by using information projection. Using the information projection makes our proves simpler since we can use information theoretic quanties and apply theorems of information theory such as Pythagorean like theorem, conditional limit theorem. The efficiency of IS with this sampling distribution is showed in the sense of asymptotic optimality criterion. The paper is organized as follows. In Section II, we state our problem with the large deviation formulation and the IS estimation. In Section III, we propose the sampling distribution and prove the efficiency of the IS. This idea of IS is illustrated in Section IV with a specific example. Finally, the conclusion and the future work are drawn.

II. Problem statement

A. Notations

Throughout this paper, $\mathcal{P}(\mathcal{X})$ denotes the space of all probability measures on the finite alphabet $\mathcal{X}$. Here $\mathcal{P}(\mathcal{X})$ is identified with the probability simplex in $\mathbb{R}^{|\mathcal{X}|}$, i.e., the set of all $|\mathcal{X}|$-dimensional real vectors with non-negative components that sum to 1. Therefore, the topology on $\mathcal{P}(\mathcal{X})$ is inherited as the subspace topology from the ordinary topology on $\mathbb{R}^{|\mathcal{X}|}$.

Definition 1. (Type.) The type $\hat{P}_{x^n}$ of a sequence $x^n = (x_1, ..., x_n) \in \mathcal{X}^n$ is its empirical distribution. More specifically, $\hat{P}_{x^n} = \left(\hat{P}_{x^n}(a_1), ..., \hat{P}_{x^n}\left(a_{|\mathcal{X}|}\right)\right)$ is an element of the set $\mathcal{P}(\mathcal{X})$ and

$$\hat{P}_{x^n}(a) = \frac{n(a|x^n)}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{a\}}(x_i) \text{ for all } a \in \mathcal{X}.$$

where $n(a \mid x^n)$ is the number of times the symbol a occurs in the sequence $x^n \in \mathcal{X}^n$ and $\mathbb{I}_{\{a\}}$is the indicator function of the set $\{a\}$.

The set of all possible types on $\mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{X})$. It is therefore obvious that $\mathcal{P}_n(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$.

We denote $\hat{P}_{X^n}$ the random element associated to it, and taking values in the set $\mathcal{P}_n(\mathcal{X})$:

$$\hat{P}_{X^n}(a) = \frac{n(a \mid X^n)}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{a\}}(X_i) \ \text{ for all } a \in \mathcal{X}.$$

B. Large deviation formulation

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables defined on the alphabet $\mathcal{X}$ and drawn according to the probability mass function (PMF) Q. Let $E \subseteq \mathcal{P}_n(\mathcal{X})$ and $H_E = \left\{x^n : \hat{P}_{x^n} \in E\right\}$, it is straightforward that $Q^n(H_E) = \mathbb{P}(X^n \in H_E) = \mathbb{P}\left(\hat{P}_{X^n} \in E\right)$. If $Q \in E$, we know from the weak law of large numbers, that the sequence of empirical distributions $\hat{P}_{x^n}$ converges in probability to the true distribution Q for all $x \in \mathcal{X}$, (Theorem 11.2.1, [2]). In contrast, for $Q \notin E$ the family of probabilities $\mathbb{P}\left(\hat{P}_{X^n} \in E\right)$ decays exponentially fast as $n \to \infty$.

Let $Q \notin E$, we want to estimate the rare event probability $\alpha_n = \mathbb{P}\left(\hat{P}_{X^n} \in E\right)$ which is characterized by the rarity parameter n.

Sanov's theorem [3] states that the family of probabilities $\mathbb{P}\left(\hat{P}_{X^n} \in E\right)$ satisfies the large deviation principle with rate function $D(P \parallel Q)$ where $P \in \mathcal{P}(\mathcal{X})$. Particularly, for any set $E \subset \mathcal{P}(\mathcal{X})$, the limsup of $\frac{1}{n}\log \mathbb{P}\left(\hat{P}_{X^n} \in E\right)$ is upper bounded by $\inf_{P \in E} D(P \parallel Q)$ and the liminf of $\frac{1}{n}\log \mathbb{P}\left(\hat{P}_{X^n} \in E\right)$ is lower bounded by $\inf_{P \in E^\circ} D(P \parallel Q)$. For specific sets E such that $E \subseteq \overline{E^\circ}$, where E° denotes the interior and $\overline{E}$ denotes the closure of E, these upper and lower bounds are identical thereby the large deviation limit exists as follows:

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left(\hat{P}_{X^n} \in E\right) = -D(Q^* \parallel Q), \tag{1}$$

where

$$Q^* = \arg \inf_{P \in E} D(P \parallel Q), \tag{2}$$

and $D(P \parallel Q)$ is the relative entropy or Kullback Leibler divergence of probability vector P relative to Q, where $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$.

In many practical applications, it is usually sufficient to consider the set E convex since with the convexity of E we can show that E is contained in the closure of its interior, i.e., $E \subseteq \overline{E^\circ}$ where we have the large deviation limit as in (1). More interestingly, with this assumption, the probability vector Q^* is unique [3]. It should be noted that for sets E satisfying $E \subseteq \overline{E^\circ}$, we can show that $\inf_{P \in E^\circ} D(P \parallel Q) = \inf_{P \in \overline{E^\circ}} D(P \parallel Q) = \inf_{P \in E} D(P \parallel Q)$. Since $\overline{E^\circ}$ is a closed set and $D(P \parallel Q)$ is a continuous function of P, then the minimum exists. Therefore Q^* can be rewritten as

$$Q^* = \arg \min_{P \in \overline{E^\circ}} D(P \parallel Q). \tag{3}$$

In this case Q^* is called an entropic projection which has a similar meaning to the dominating point mentioned in [5].

C. Importance sampling estimation

As mentioned above, we desire to estimate rare event probabilities α_n which is characterized by the rarity parameter n such that $\alpha_n \to 0$ as $n^{-1} \to 0$ and satisfy the large deviation principle. Even though the analytic behavior is well expressed with respect to asymptotic upper and lower bounds of the large deviation statement (which are identical in the case of convex sets E), the exact value α_n remains far from its bounds for limited values of n. In the rest of this work we will deal with convex sets E so that the limit in (1) is verified.

A straightforward way to estimate α_n is to express it as $\alpha_n = \mathbb{E}_{Q^n}\left[\mathbb{I}_{H_E}(X^n)\right]$ and use a crude Monte Carlo simulation by generating N independent outcomes (x^n) according to Q^n. The estimator $\hat{\alpha}_{MC}$ is as follows:

$$\hat{\alpha}_{MC} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_{H_E}\left((x^n)^i\right), \tag{4}$$

where each $(x)^i$ is generated from Q. The MC simulation is inadequate for rare event estimation as the number of trials N needed to hit one rare event is in average $1/\alpha_n$ (e.g. $N = 10^{10}$ for $\alpha_n = 10^{-10}$). Thus as $n \to \infty$, the number of trials grows exponentially fast since the probabilities α_n satisfy the large deviation principle. It is also worth considering the relative error of the estimator κ, i.e. the ratio of its standard deviation to its mean:

$$\kappa^2 = \frac{\mathrm{var}(\mathbb{I}_{H_E}(X^n))}{N\alpha_n^2}. \tag{5}$$

Since $\mathrm{var}(\mathbb{I}_{H_E}(X^n)) = \alpha_n - \alpha_n^2$, we have:

$$\kappa^2 = \frac{1}{N}\left(\frac{1}{\alpha_n} - 1\right). \tag{6}$$

The smaller is α_n, the larger is the number of simulation we need to run in order to obtain a confidence interval with some prescribed accuracy. In other words, given a constrained and fixed number of trials N, the relative error of the crude MC estimation is unbounded with respect to n.

In the subsequent sections we focus on Importance Sampling method and propose an alternative sampling distribution, let say $\tilde{Q}$ to speed up rare events occurrence with a more efficient simulation. Let us express first the estimator $\hat{\alpha}_{IS}$ of IS:

$$\begin{aligned} \alpha_n &= \mathbb{E}_{Q^n}\left[\mathbb{I}_{H_E}(X^n)\right], \\ &= \mathbb{E}_{\tilde{Q}^n}\left[\mathbb{I}_{H_E}(X^n)\frac{Q^n(X^n)}{\tilde{Q}^n(X^n)}\right], \end{aligned}$$

$$\hat{\alpha}_{IS} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_{H_E}\left((x^n)^i\right) \times \left[\frac{Q^n\left((x^n)^i\right)}{\tilde{Q}^n\left((x^n)^i\right)}\right].$$

where each $(x)^i$ is generated from $\tilde{Q}$. The variance of $\hat{\alpha}_{IS}$ is then computed as:

$$\mathrm{var}(\hat{\alpha}_{IS}) = \frac{1}{N}\left(\eta_n\left(\tilde{Q}^n\right) - \alpha_n^2\right), \tag{7}$$

where

$$\eta_n\left(\tilde{Q}^n\right) = \mathbb{E}_{\tilde{Q}^n}\left[\mathbb{I}_{H_E}\left(X^n\right) \times \left(\frac{Q^n\left(X^n\right)}{\tilde{Q}^n\left(X^n\right)}\right)^2\right]. \quad (8)$$

In the next section, we will construct a sampling distribution $\tilde{Q}$ such that $\mathrm{var}\left(\hat{\alpha}_{IS}\right) \leq \mathrm{var}\left(\hat{\alpha}_{MC}\right)$ for every n. We then show that this IS estimator is asymptotically efficient [1], [5] with respect to rarity parameter n, i.e.,

$$\lim_{n\to\infty} \frac{\log \eta_n\left(\tilde{Q}^n\right)}{\log \alpha_n^2} = 1, \quad (9)$$

or equivalently:

$$\lim_{n\to\infty} \frac{1}{n} \log \eta_n\left(\tilde{Q}^n\right) = -2D\left(Q^* \parallel Q\right), \quad (10)$$

where Q^* is defined by (12). Asymptotic efficiency means that the second moment decays exponentially as fast as α_n^2 and the number of simulation runs required for a prescribed accuracy grows sub-exponentially with n.

III. Importance sampling and Information projection

In order to estimate the family of rare event probabilities α_n , we consider the sampling distribution Q^* which gives the infinimum KL-divergence of distributions P in E relative to Q. We recall that the set E is a convex set of non empty interior so that the large deviation limit (1) holds. Recall that for convex set E, we can have

$$\inf_{P\in E} D\left(P \parallel Q\right) = \inf_{P\in \overline{E^\circ}} D\left(P \parallel Q\right) = \min_{P\in \overline{E^\circ}} D\left(P \parallel Q\right). \quad (11)$$

The sampling distribution is then chosen as:

$$Q^* = \arg\min_{P\in \overline{E^\circ}} D\left(P \parallel Q\right). \quad (12)$$

With these hypotheses, we shall show in next subsections that the variance of IS estimator is lower than for standard MC and more desirably, that it satisfies the asymptotic efficiency property.

A. The efficiency of IS

Proposition 2. $\mathrm{var}\left(\hat{\alpha}_{IS}\right) \leq \mathrm{var}\left(\hat{\alpha}_{MC}\right)$

Proof: In order to show that $\mathrm{var}\left(\hat{\alpha}_{IS}\right) \leq \mathrm{var}\left(\hat{\alpha}_{MC}\right)$, we only need to establish that:

$$\eta_n\left(Q^{*n}\right) \quad \leq \quad \alpha_n. \quad (13)$$

We have

$$\begin{aligned}\eta_n\left(Q^{*n}\right) &= \mathbb{E}_{Q^{*n}}\left[\mathbb{I}_{H_E}\left(X^n\right) \times \left(\frac{Q^n\left(X^n\right)}{Q^{*n}\left(X^n\right)}\right)^2\right], \\ &= \mathbb{E}_{Q^n}\left[\mathbb{I}_{H_E}\left(X^n\right) \times \left(\frac{Q^n\left(X^n\right)}{Q^{*n}\left(X^n\right)}\right)\right]. \quad (14)\end{aligned}$$

Since we consider a sequence X^n of i.i.d. random variables, the second term in the expectation may be developped as:

$$\begin{aligned}&\frac{Q^n(X^n)}{Q^{*n}(X^n)} = \prod_{i=1}^n \frac{Q(X_i)}{Q^*(X_i)}, \\ &= \prod_{a\in\mathcal{X}} \left(\frac{Q(a)}{Q^*(a)}\right)^{n(a|X^n)}, \\ &= \exp\left[n \sum_{a\in\mathcal{X}} \frac{n(a|X^n)}{n} \log \frac{Q(a)}{Q^*(a)}\right], \\ &= \exp\left[n \sum_{a\in\mathcal{X}} \hat{P}_{X^n}(a) \log \frac{\hat{P}_{X^n}(a)}{Q^*(a)} \frac{Q(a)}{\hat{P}_{X^n}(a)}\right], \\ &= \exp\left[n \sum_{a\in\mathcal{X}} \hat{P}_{X^n}(a) \left(\log \frac{\hat{P}_{X^n}(a)}{Q^*(a)} - \log \frac{\hat{P}_{X^n}(a)}{Q(a)}\right)\right], \\ &= \exp\left\{n\left[\mathbb{D}\left(\hat{P}_{X^n} \parallel Q^*\right) - \mathbb{D}\left(\hat{P}_{X^n} \parallel Q\right)\right]\right\}. \quad (15)\end{aligned}$$

where $\mathbb{D}\left(\hat{P}_{X^n} \parallel Q\right)$ is a non negative random element. Plugging (15) into the right hand side of (14), we have:

$$\begin{aligned}&\eta_n\left(Q^{*n}\right) \\ &= \mathbb{E}_{Q^n}\left\{I_{H_E}\left(X^n\right) \times \exp\left[n\left(\mathbb{D}\left(\hat{P}_{X^n} \parallel Q^*\right) - \right.\right.\right. \\ &\left.\left.\left. \mathbb{D}\left(\hat{P}_{X^n} \parallel Q\right)\right)\right]\right\}, \\ &= \sum_{x^n\in\mathcal{X}^n} Q^n\left(x^n\right) \mathbb{I}_{H_E}\left(x^n\right) \exp\left[n\left(\mathbb{D}\left(\hat{P}_{X^n} \parallel Q^*\right) - \right.\right. \\ &\left.\left. \mathbb{D}\left(\hat{P}_{x^n} \parallel Q\right)\right)\right], \\ &= \sum_{x^n\in H_E} Q^n\left(x^n\right) \exp\left\{n\left[D\left(\hat{P}_{x^n} \parallel Q^*\right) - D\left(\hat{P}_{x^n} \parallel Q\right)\right]\right\}. \quad (16)\end{aligned}$$

Since $\overline{E^\circ}$ is a closed convex set and Q^* is a distribution achieving the minimum KL divergence of distributions $\hat{P}_{x^n}$ in $\overline{E^\circ}$ to Q as defined in Eq.(12), using a Pythagorean like theorem (cf. Theorem 11.6.1, [2]), we have for all $\hat{P}_{x^n} \in \overline{E^\circ}$:

$$D\left(\hat{P}_{x^n} \parallel Q\right) \geq D\left(\hat{P}_{x^n} \parallel Q^*\right) + D\left(Q^* \parallel Q\right).$$

Equivalently,

$$-D\left(Q^* \parallel Q\right) \geq D\left(\hat{P}_{x^n} \parallel Q^*\right) - D\left(\hat{P}_{x^n} \parallel Q\right). \quad (17)$$

From (16) and (17) we have

$$\begin{aligned}\eta_n\left(Q^{*n}\right) &= \mathbb{E}_{Q^n}\left[\mathbb{I}_{H_E}\left(X^n\right) \times \left(\frac{Q^n(X^n)}{Q^{*n}(X^n)}\right)\right], \\ &\leq \sum_{x^n\in H_E} Q^n\left(x^n\right) \exp\left\{-nD\left(Q^* \parallel Q\right)\right\}, \quad (18) \\ &\leq \sum_{x^n\in H_E} Q^n\left(x^n\right) = \alpha.\end{aligned}$$

The last inequality is easy to see because the relative entropy is non negative thus $\exp\{-nD(Q^* \parallel Q)\} \leq 1$. ■

Proposition 3. *(Asymptotic optimality) Show that*

$$\lim_{n\to\infty} \frac{1}{n} \log \eta_n (Q^{*n}) = -2D(Q^* \parallel Q). \tag{19}$$

Proof: Conditioned to $\hat{P}_{x^n} \in E$, by the conditional limit theorem (cf. Theorem 11.6.2, [2]) we have

$$\hat{P}_{x^n} \overset{n\to\infty}{\to} Q^* \qquad \text{in probability.} \tag{20}$$

By the continuity of relative entropy, we have

$$D\left(\hat{P}_{x^n} \parallel Q\right) \overset{n\to\infty}{\to} D(Q^* \parallel Q) \ . \tag{21}$$

Plugging (21) into (16) we obtain:

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log \eta_n (Q^{*n}) \\
&= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{Q^n} \left[\mathbb{I}_{H_E}(X^n) \times \left(\tfrac{Q^n(X^n)}{Q^{*n}(X^n)} \right) \right], \\
&= \lim_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in H_E} Q^n(x^n) \exp\left\{ n \left[D\left(\hat{P}_{x^n} \parallel Q^*\right) - \right.\right. \\
&\left.\left. D\left(\hat{P}_{x^n} \parallel Q\right) \right] \right\}, \\
&= \lim_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in H_E} Q^n(x^n) \exp\{ n [-D(Q^* \parallel Q)] \}, \\
&= \lim_{n\to\infty} \left[-D(Q^* \parallel Q) + \frac{1}{n} \log \alpha_n \right], \\
&\overset{(a)}{=} -2D(Q^* \parallel Q).
\end{aligned}$$

where (a) comes from the large deviation limit defined in Eq.(1), i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_n = -D(Q^* \parallel Q). \tag{22}$$

■

B. Discussion

In this paper, we estimate the rare event probability $\mathbb{P}\left(\hat{P}_{X^n} \in E\right)$ where $\hat{P}_{X^n}$ is the type of an i.i.d. sequence $\{X_i\}$ drawn according to some PMF Q . We consider convex sets E as subsets of $\mathcal{P}(\mathcal{X})$, the probability simplex in $\mathbb{R}^{|\mathcal{X}|}$. We analyse the use of importance sampling to numerically estimate the family of probabilities $\alpha_n = \mathbb{P}\left(\hat{P}_{X^n} \in E\right)$ satisfying the large deviation principle with rate $D(Q^* \parallel Q)$. Importance sampling with Q^* as sampling distribution performs efficient estimation when crude MC simulations are inadequate for rare event. In [5] the authors applied also importance sampling to estimate a family of probabilities satisfying large deviation principle with a twisted version of the original distribution as sampling distribution. This tilted operation is formed with the exponential of the Legendre-Fenchel transform of the log moment generating function of the original distribution. It is already well known [3] that this transform is a rate function for probabilities satisfying the large deviation principle. The Contraction Principle allows to establish the relation between this transform and the Sanov's rate. Thus in our paper we choose the information projection instead. The authors in [5] first showed that the best sampling distribution is the one associated with the *dominating point* if it exists which in our case is obtained by the sufficient condition on the convexity of E, and under this assumption this probability vector Q^* is unique. Performing simulation runs in the probability domain by counting the frequency of occurence of each value of the random variable is particulary easier to implement.

IV. Practical result

In this section, we assess the use of Q^* as an alternative sampling distribution in IS estimation of a binomial sequence with parameters $(n,\, p = 0.1)$. We wish to estimate the family of probabilities $\alpha_n = \mathbb{P}\left(\sum_{k=1}^{n} X_k \in [0, \lambda)\right)$ where λ is a given threshold far from the expected value np of the sum. From large deviation principle the sampling distribution can easily be computed as $Q^* = \left\{\frac{\lambda}{n},\, 1 - \frac{\lambda}{n}\right\}$. Fig. 1 and Fig. 2 show the estimated probability $\hat{\alpha}_{IS}$ using importance sampling with sampling distribution Q^*. In Fig. 1 the estimated value is represented (stars points) as $\frac{1}{n}\log\hat{\alpha}_{IS}$ and compared to the true value $\frac{1}{n}\log\alpha_n$ (solid line) and to the rate function $D(Q^* \parallel Q)$, as functions of the rarity parameter n . We see that for sequences of length less than 10^4 the limit of the large deviation is far from the exact value so it is expected that the analytic expression is poor for a limited length n of sequences.

In Fig. 2, the length of the sequence is limited to $n = 2000$. As the threshold λ decays, rarity increases and the family of probabilities α_n goes to zero as $\lambda \to 0$. Here again expected values obtained with importance sampling employing Q^* as sampling distribution are very close to the true ones. For $\lambda = 100$, $\alpha_n = 2,3.10^{-16}$, and the estimated variance S of $\hat{\alpha}_{IS}$ is evaluated over $N = 40000$ simulation runs. By the law of large numbers, $S \to \mathrm{var}(\hat{\alpha}_{IS})$ for large N. The relative error is then evaluated as in (5) but for the importance sampling estimation:

$$\kappa = \frac{\sqrt{\mathrm{var}(\hat{\alpha}_{IS})}}{\alpha_n}.$$

We obtain $\kappa = 0.0143$ which means that the 95% confidence interval gives has an accuracy of $\pm 0.028\alpha_n$ when crude MC needs more than 10^{18} simulation runs to insure this accuracy. A factor of 10^{13} may then be saved.

V. Conclusion and Future works

We have shown that the sampling distribution based on information projection made our IS simulation efficient to estimate the rare event probabilities which satisfy a large deviation principle. This efficiency is evaluated in the sense of

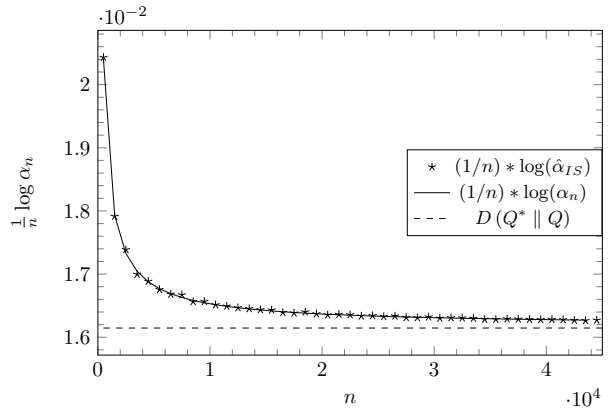

Figure 1. Values of $\frac{1}{n}\log\hat{\alpha}_{IS}$ and of $\frac{1}{n}\log\alpha_n$ compared to $D(Q^* \parallel Q)$.

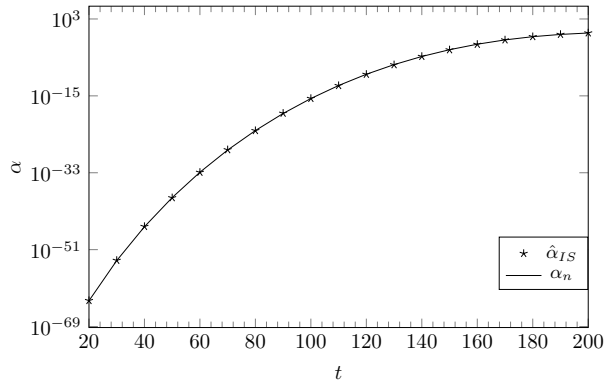

Figure 2. Values of α_n and $\hat{\alpha}_{IS}$ with respect to the threshold.

asymptotic optimality. We have also put this IS into practice with a specific example to compute the heavy tail of binomial.

The future work shall be devoted to employing the IS to estimate the rare event with another rarity parameter. More specifically, in this paper, we have used n as a rarity parameter to index the rare event probability α_n. In the future, we want to estimate a positive quantity $\alpha(\epsilon) = \mathbb{P}(E_\epsilon)$ that depends on a rarity parameter $\epsilon > 0$ where $E_\epsilon \subseteq \mathcal{P}(\mathcal{X})$ such that $\mathbb{P}(E_\epsilon)$ is a monotone strictly decreasing function of ϵ and that $\lim_{\epsilon \to 0} \mathbb{P}(E_\epsilon) = 0$. We plan to find a sampling distribution based on information projection and to check if the IS with the considered sampling distribution is efficient.

ACKNOWLEDGMENT

The authors thank the Centre National de la Recherche Scientifique (CNRS) for specific support on this topic.

REFERENCES

[1] Soren Asmussen, Paul Dupuis, Reuven Rubinstein, and Hui Wang. Importance sampling for rare events. *Aarhus Univ., Aarhus, Denmark, Tech. Rep., YEAR*, 2011.

[2] Thomas M Cover and Joy A Thomas. Elements of information theory 2nd edition. 2006.

[3] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 2. Springer, 1998.

[4] Sandeep Juneja and Perwez Shahabuddin. Rare-event simulation techniques: an introduction and recent advances. *Handbooks in operations research and management science*, 13:291–350, 2006.

[5] John S Sadowsky and James A Bucklew. On large deviations theory and asymptotically efficient monte carlo estimation. *Information Theory, IEEE Transactions on*, 36(3):579–588, 1990.