



HAL
open science

A double proposal normalized importance sampling estimator

Roland Lamberti, Yohan Petetin, François Septier, François Desbouvries

► **To cite this version:**

Roland Lamberti, Yohan Petetin, François Septier, François Desbouvries. A double proposal normalized importance sampling estimator. SSP 2018: IEEE Statistical Signal Processing Workshop, Jun 2018, Freiburg, Germany. pp.238 - 242, 10.1109/SSP.2018.8450849 . hal-01870199

HAL Id: hal-01870199

<https://hal.science/hal-01870199v1>

Submitted on 1 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A DOUBLE PROPOSAL NORMALIZED IMPORTANCE SAMPLING ESTIMATOR

Roland Lamberti¹, Yohan Petetin¹, François Septier², François Desbouvries¹

(1) Samovar, Telecom Sudparis, CNRS, Université Paris-Saclay, 91011 Evry, France

(2) IMT Lille Douai, Université Lille, CNRS, UMR 9189 - CRISAL, F-59000 Lille, France

ABSTRACT

Monte Carlo methods are widely used in signal processing for computing integrals of interest. Among Monte Carlo methods, Importance Sampling is a variance reduction technique which consists in sampling from an instrumental distribution and reweighting the samples in order to correct the discrepancy between the target and proposal distributions. When either the target or the proposal distribution is known only up to a constant, the moment of interest can be rewritten as a ratio of two expectations, which can be approximated via self-normalized importance sampling. In this paper we show that it is possible to improve the self-normalized importance sampling estimate by approximating the two expectations in this ratio via two importance distributions. In order to tune them we optimize the variance of the final estimate under a reasonable constraint. Our results are validated via simulations.

Index Terms— Monte Carlo integration, (self-normalized) importance sampling, variance minimization

1. INTRODUCTION

1.1. Importance Sampling (IS)

Let x be a random variable (r.v.) with density $\pi(\cdot)$. As far as notations are concerned we do not make a distinction between r.v. and their realizations. In many signal processing problems we are interested in computing the expectation of some function $f(x)$ w.r.t. π :

$$\mu = \int f(x)\pi(x)dx = E_{\pi}(f(x)). \quad (1)$$

In practice (1) can be very difficult to compute, so one often needs to resort to approximations. In this paper we focus on Monte Carlo (MC) integration methods (see e.g. [1]) which aim at approximating $E_{\pi}(f(x))$ with a (possibly weighted) set of random samples. The variance of the estimate depends on the number N of samples, on function f and pdf π , and on the distribution from which the points are sampled from.

As far as this last point is concerned, a natural distribution to sample from is the target distribution π , which results in the crude MC estimate $\hat{\mu}^{\text{MC}} = \frac{1}{N} \sum_{i=1}^N f(x^i)$ with $x^i \stackrel{\text{i.i.d.}}{\sim} \pi$. However i) it is generally difficult to sample directly from π ;

ii) the crude MC estimate can give poor results, particularly when the regions where π is large do not coincide with those where f is large. IS is a well established technique [2] [3] [4, §5.4] for bypassing these difficulties. Rewriting (1) as

$$\mu = \int f(x) \frac{\pi(x)}{q(x)} q(x) dx = E_q(f(x) \frac{\pi(x)}{q(x)}), \quad (2)$$

where q is some importance distribution, leads to the IS estimator (the dependency in pdf q is stressed via the notation $\hat{\mu}^{\text{IS}}(q)$):

$$\hat{\mu}^{\text{IS}}(q) = \frac{1}{N} \sum_{i=1}^N \frac{\pi(x^i)}{q(x^i)} f(x^i), \quad x^i \stackrel{\text{i.i.d.}}{\sim} q. \quad (3)$$

As far as variance reduction is concerned, one can easily show that $N \text{var}(\hat{\mu}^{\text{IS}}(q)) = \int \frac{\pi^2 f^2}{q} - \mu^2$ and that the importance density $q_{\text{opt}}^{\text{IS}}$ which minimizes $\text{var}(\hat{\mu}^{\text{IS}}(q))$, and associated (scaled) minimal variance, are

$$q_{\text{opt}}^{\text{IS}}(x) \propto |f(x)|\pi(x), \quad (4)$$

$$N \text{var}(\hat{\mu}^{\text{IS}}(q_{\text{opt}}^{\text{IS}})) = \left(\int |f(x)|\pi(x) dx \right)^2 - \mu^2. \quad (5)$$

Even if in practice the optimal estimator $\hat{\mu}^{\text{IS}}(q_{\text{opt}}^{\text{IS}})$ cannot be computed (we know $q_{\text{opt}}^{\text{IS}}$ up to the unknown constant $\int |f(x)|\pi(x) dx$), this result tells us that the regions where it is important to sample from (whence the term "importance distribution") are not those where π is large, but rather those where $|f|\pi$ is large (see point ii) above).

1.2. Normalized IS (NIS)

From (3) we see that $\hat{\mu}^{\text{IS}}$ can be computed only if π is known exactly, or π and q are known up to a common constant. If this is not the case one can resort to self-normalized (NIS). More precisely, let $\pi(x) = \frac{p(x)}{\int p(x) dx}$ where p is known but the normalization constant is not. Then (2) can be rewritten as

$$\mu = \frac{\int f(x) \frac{p(x)}{q(x)} q(x) dx}{\int \frac{p(x)}{q(x)} q(x) dx} = \frac{E_q(f(x) \frac{p(x)}{q(x)})}{E_q(\frac{p(x)}{q(x)})}; \quad (6)$$

next each expectation can be approximated by a crude MC estimate from the same set of N i.i.d. samples $\{x^i\}_{i=1}^N$ drawn

from q . The NIS estimate $\hat{\mu}^{\text{NIS}}(q)$ of μ reads

$$\hat{\mu}^{\text{NIS}}(q) = \frac{\sum_{i=1}^N \frac{p(x^i)}{q(x^i)} f(x^i)}{\sum_{j=1}^N \frac{p(x^j)}{q(x^j)}} = \frac{\frac{1}{N} \sum_{i=1}^N \frac{\pi(x^i)}{q(x^i)} f(x^i)}{\frac{1}{N} \sum_{j=1}^N \frac{\pi(x^j)}{q(x^j)}}. \quad (7)$$

The first expression in (7) is used to compute the estimate in practice, while the second one is used to analyze its statistical properties.

The properties of $\hat{\mu}^{\text{NIS}}(q)$ are well understood [5, Chapter 2] [6] [7]. First, it is well-established [6] that under mild assumptions,

$$\hat{\mu}^{\text{NIS}}(q) \xrightarrow{\text{a.s.}} \mu; \quad (8)$$

the bias of $\hat{\mu}^{\text{NIS}}(q)$ is asymptotically null in $1/N$; its variance can be approximated at first-order by applying the delta-method to the ratio (7) [8, p.35]:

$$N \widehat{\text{var}}(\hat{\mu}^{\text{NIS}}(q)) = \text{var}\left(\frac{\pi(x)f(x)}{q(x)}\right) + \mu^2 \text{var}\left(\frac{\pi(x)}{q(x)}\right) - 2\mu \text{cov}\left(\frac{\pi(x)f(x)}{q(x)}, \frac{\pi(x)}{q(x)}\right); \quad (9)$$

finally the following Central Limit Theorem (CLT) [6] (see also [7]) holds:

$$\sqrt{N}(\hat{\mu}^{\text{NIS}}(q) - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma^{2, \text{NIS}}(q)), \quad (10)$$

$$\sigma^{2, \text{NIS}}(q) = \text{E}_q\left[\frac{\pi^2(x)}{q^2(x)}(f(x) - \mu)^2\right]. \quad (11)$$

Note that the first-order approximated variance in (9) and asymptotic variance in (11) coincide:

$$N \widehat{\text{var}}(\hat{\mu}^{\text{NIS}}(q)) = \sigma^{2, \text{NIS}}(q).$$

The choice of the importance distribution $q(\cdot)$ in (7) has been the subject of a vast literature, see e.g. [1, 6, 8–12] (and [13] specifically in the non-normalized case) When the objective is to approximate μ for a large class of functions f , the importance distribution q is often chosen close to π . Now for a specific function f , the IS distribution which minimizes $\sigma^{2, \text{NIS}}(q)$ in (11), and associated minimal asymptotic variance, are [6]

$$q_{\text{opt}}^{\text{NIS}}(x) \propto |f(x) - \mu| \pi(x), \quad (12)$$

$$\sigma^{2, \text{NIS}}(q_{\text{opt}}^{\text{NIS}}) = \left(\int |f(x) - \mu| \pi(x) dx\right)^2. \quad (13)$$

1.3. Scope of the paper

In this paper we start from the observation that the minimum variance $\text{var}(\hat{\mu}^{\text{IS}}(q_{\text{opt}}^{\text{IS}}))$ in (5) is zero for all positive functions f , while in the normalized case $\sigma^{2, \text{NIS}}(q_{\text{opt}}^{\text{NIS}})$ in (13) is strictly positive whatever function f (except of course in the degenerate situation where $f(x)$ is constant). Heuristically, the reason why is that $\hat{\mu}^{\text{NIS}}$ approximates the ratio (6), so both integrals

should be estimated as precisely as possible; but even though $q_{\text{opt}}^{\text{NIS}}(x)$ in (12) is the best compromise between $\frac{|f|p}{\int |f|p}$ and $\frac{p}{\int p}$, it is impossible for *one single* set of points $\{x^i\}_{i=1}^N$ to be both in regions where $|f|\pi$ is large, and in regions where π is large (- unless, again, if f is a constant);

In this paper we thus revisit the NIS mechanism by considering two importance distributions q_1 and q_2 , which leads to the Double Proposal Importance Sampling (DPIS) estimator $\hat{\mu}^{\text{DPIS}}(q_1, q_2)$. We next discuss on how to tune these importance distributions in order to improve the classical estimate $\hat{\mu}^{\text{NIS}}$ in terms of asymptotic variance while keeping its computational attractiveness. The rest of this paper is organized as follows. Section 2.1 is devoted to the DPIS estimator and its properties. In section 2.2 we discuss the choice of importance distributions q_1 and q_2 and we propose a simple solution. Section 3 deals with simulations and we end the paper with a conclusion.

2. DOUBLE PROPOSAL IS (DPIS)

2.1. The DPIS estimator

In this section we start again from $\mu = \frac{\int f p}{\int p}$, and we introduce two importance distributions, q_1 for the numerator and q_2 for the denominator. So μ can be rewritten as

$$\mu = \frac{\int f(x) \frac{p(x)}{q_1(x)} q_1(x) dx}{\int \frac{p(x)}{q_2(x)} q_2(x) dx} = \frac{\text{E}_{q_1}(f(x) \frac{p(x)}{q_1(x)})}{\text{E}_{q_2}(\frac{p(x)}{q_2(x)})}. \quad (14)$$

An MC estimate of μ based on (14) can be obtained with N i.i.d. samples $\{x_1^i, x_2^i\}_{i=1}^N$ where $x_1^i \sim q_1$ and $x_2^i \sim q_2$ (note that for the moment, we do not make any assumption on the dependency for a given i between x_1^i and x_2^i). The resulting estimate $\hat{\mu}^{\text{DPIS}}$ reads

$$\hat{\mu}^{\text{DPIS}}(q_1, q_2) = \frac{\sum_{i=1}^N \frac{p(x_1^i)}{q_1(x_1^i)} f(x_1^i)}{\sum_{j=1}^N \frac{p(x_2^j)}{q_2(x_2^j)}} = \frac{\frac{1}{N} \sum_{i=1}^N \frac{\pi(x_1^i)}{q_1(x_1^i)} f(x_1^i)}{\frac{1}{N} \sum_{j=1}^N \frac{\pi(x_2^j)}{q_2(x_2^j)}}. \quad (15)$$

The following proposition describes the statistical properties of $\hat{\mu}^{\text{DPIS}}(q_1, q_2)$ and generalizes those recalled in Section 1.2 (the proofs are omitted but follow the same line as those of the results presented in Section 1.2).

Proposition 1 (First-order variance and CLT for DPIS)

Let $x_1 \sim q_1$ and $x_2 \sim q_2$. A first-order approximation of $\text{var}(\hat{\mu}^{\text{DPIS}})$ reads

$$N \widehat{\text{var}}(\hat{\mu}^{\text{DPIS}}(q_1, q_2)) = \text{var}\left(\frac{\pi(x_1)f(x_1)}{q_1(x_1)}\right) + \mu^2 \text{var}\left(\frac{\pi(x_2)}{q_2(x_2)}\right) - 2\mu \text{cov}\left(\frac{\pi(x_1)f(x_1)}{q_1(x_1)}, \frac{\pi(x_2)}{q_2(x_2)}\right). \quad (16)$$

In addition, let us assume that

$$\sigma^2(q_1, q_2) = \text{E}\left[\left(\frac{\pi(x_1)}{q_1(x_1)} f(x_1) - \frac{\pi(x_2)}{q_2(x_2)} \mu\right)^2\right] \quad (17)$$

is finite. Then the following CLT holds:

$$\sqrt{N}(\hat{\mu}^{\text{DPIS}} - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma^2(q_1, q_2)) \quad (18)$$

and the (scaled) first-order approximated variance $N \times \widehat{\text{var}}(\hat{\mu}^{\text{DPIS}})(q_1, q_2)$ in (16) coincides with the asymptotic variance $\sigma^2(q_1, q_2)$ in (17).

2.2. A linear sample adjustment solution

Proposition (1) holds whatever q_1, q_2 , and the dependency between the samples $\{x_1^i\}$ drawn from q_1 and $\{x_2^i\}$ drawn from q_2 . In this section let us discuss how to tune the samples x_1^i and x_2^i . We begin with two simple remarks: i) $\hat{\mu}^{\text{DPIS}}(q_1, q_2)$ coincides with $\hat{\mu}^{\text{NIS}}$ when $q_1 = q_2$ and $x_1^i = x_2^i$ for all i , $1 \leq i \leq N$; ii) by contrast with the classical NIS mechanism, the variance of $\hat{\mu}^{\text{DPIS}}(q_1, q_2)$ reduces to 0 when $q_1 = q_{1,\text{opt}}^{\text{DPIS}} \propto |f|p$ and $q_2 = q_{2,\text{opt}}^{\text{DPIS}} \propto p$.

Of course, neither $q_{1,\text{opt}}^{\text{DPIS}}$ nor $q_{2,\text{opt}}^{\text{DPIS}}$ are computable in practice. Techniques to approach such distributions have been proposed in the context of unnormalized IS and could be adapted here, but in this paper we follow another direction; we observe that the presence of two (potentially different) importance densities q_1 and q_2 introduces more flexibility as compared to NIS, which leads us to the problem of minimizing the variance term (16), but at the price of a reasonable computational effort. To that end, remember that a computational advantage of the NIS mechanism is that one only needs to draw N samples from a single q . In this paper we do want to depart from this unique importance density $q = q_1 = q_2$, but as a compromise between sampling diversity and sampling cost we will assume that x_2^i is deterministic given x_1^i .

So let us now consider an importance distribution q and N initial i.i.d. samples $\{x^i\}_{i=1}^N$ drawn from q . Our basic idea is to transform the samples from q into samples $\{x_1^i\}_{i=1}^N$ from q_1 in order to approximate the numerator of (14), and next into samples $\{x_2^i\}_{i=1}^N$ from q_2 in order to approximate the denominator of (14). Here we will focus on linear transformations of the initial samples $\{x^i\}_{i=1}^N$ relying on scaling parameters (α_1, α_2) and translation parameters (β_1, β_2) :

$$x_1^i = \alpha_1 x^i + \beta_1, \quad x_2^i = \alpha_2 x^i + \beta_2. \quad (19)$$

It means that collections $\{x_1^i\}_{i=1}^N$ and $\{x_2^i\}_{i=1}^N$ are drawn i.i.d. from distributions $q_1(\cdot) = |\frac{1}{\alpha_1}|q(\frac{\cdot - \beta_1}{\alpha_1})$ and $q_2(\cdot) = |\frac{1}{\alpha_2}|q(\frac{\cdot - \beta_2}{\alpha_2})$, respectively but are dependent. The linear relationship between x_1^i and x_2^i for all i is given by $x_2^i = \frac{\alpha_2}{\alpha_1}(x_1^i - \beta_1) + \beta_2$.

It remains to compute the optimal parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2)$. Under this constraint (16) can be expressed as

$$N \widehat{\text{var}}(\hat{\mu}^{\text{DPIS}}) = \frac{1}{C^2} (\mathbb{E}_q[\frac{1}{q^2(x)} (|\alpha_1|p(\alpha_1 x + \beta_1) \times f(\alpha_1 x + \beta_1) - |\alpha_2|p(\alpha_2 x + \beta_2)\mu)^2]) \quad (20)$$

where $C = \int p(x)dx$. Note that if the minimization is performed correctly, the resulting estimate $\hat{\mu}^{\text{DPIS}}$ necessarily outperforms $\hat{\mu}^{\text{NIS}}$ (provided N is large enough to satisfy the first-order approximation (16)) which corresponds to the setting $\alpha_1 = \alpha_2 = 1, \beta_1 = \beta_2 = 0$.

Equation (20) cannot be computed exactly because of the μ factor and of the integrals, but can be approximated via the available samples $\{x^i\}_{i=1}^N$ initially drawn from q . An MC estimate of (20) reads

$$N \widehat{\text{var}}(\hat{\mu}^{\text{DPIS}}) = \frac{1}{C^2} \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{q^2(x^i)} \times (|\alpha_1|p(\alpha_1 x^i + \beta_1)f(\alpha_1 x^i + \beta_1) - |\alpha_2|p(\alpha_2 x^i + \beta_2)\hat{\mu}^{\text{NIS}})^2 \right). \quad (21)$$

To minimize this approximated variance, we use a sub-gradient descent method with backtracking line search for determining parameter move steps, using multiple random initializations over a given area in an effort to avoid local minima. Moreover, in order to improve the overall approximation we update our initial estimation of μ (needed in the first-order variance expression) at every gradient iteration using samples obtained with the current set of parameters.

3. SIMULATIONS

In this section we compare, given an importance distribution q , the classical NIS estimate $\hat{\mu}^{\text{NIS}}(q)$ based on N samples $\{x^i\}_{i=1}^N$ from q , to our DPIS estimate $\hat{\mu}^{\text{DPIS}}(q_1, q_2)$ in which the original samples x^i are linearly adjusted by minimizing (20) or (21). The estimates are compared in terms of RMSE over $P = 100$ MC runs.

3.1. A toy example

Let us first compute an MC estimate of μ in (1) where $f(x) = x^2$ and $\pi(x) = \mathcal{N}(x, 0, 1)$ ($\mathcal{N}(x, m, \sigma^2)$ denotes the Gaussian distribution with mean m and variance σ^2). For the initial importance distribution we set $q = \pi$. Here $\widehat{\text{var}}(\hat{\mu}^{\text{DPIS}})$ can be computed in closed form as a function of $(\alpha_1, \beta_1, \alpha_2, \beta_2)$ because $\mu = 1$ is known and the integrals are computable. More precisely, $N \widehat{\text{var}}(\hat{\mu}^{\text{DPIS}})$ reads

$$\begin{aligned} & \frac{\alpha_1^2}{\sqrt{2\alpha_1^2 - 1}} e^{\frac{\beta_1^2}{2\alpha_1^2 - 1}} \left[\frac{\beta_1^4}{(2\alpha_1^2 - 1)^4} + \frac{6\beta_1^2 \alpha_1^2}{(2\alpha_1^2 - 1)^3} + \frac{3\alpha_1^4}{(2\alpha_1^2 - 1)^2} \right] \\ & + \frac{\alpha_2^2}{\sqrt{2\alpha_2^2 - 1}} e^{\frac{\beta_2^2}{2\alpha_2^2 - 1}} - \frac{2|\alpha_1||\alpha_2|}{\sqrt{\alpha_1^2 + \alpha_2^2 - 1}} \times \\ & \exp\left(\frac{\beta_1^2(1 - \alpha_2^2) + \beta_2^2(1 - \alpha_1^2) + 2\beta_1\beta_2\alpha_1\alpha_2}{2(\alpha_1^2 + \alpha_2^2 - 1)} \right) \\ & \times \left[\frac{\alpha_1^2}{\alpha_1^2 + \alpha_2^2 - 1} + \left(\frac{\beta_2\alpha_1\alpha_2 - \beta_1\alpha_2^2 + \beta_1}{\alpha_1^2 + \alpha_2^2 - 1} \right)^2 \right]. \end{aligned} \quad (22)$$

The optimal parameters are ($\alpha_1^{opt} \approx 2.2, \beta_1^{opt} \approx 0, \alpha_2^{opt} \approx 1.25, \beta_2^{opt} \approx 0$) which means that only the variance of q is adjusted.

We compute two estimates based on a double proposal distribution: the first one uses the optimal parameters above while the second one uses the parameters ($\alpha_1, \alpha_2, \beta_1, \beta_2$) estimated from (21) via a sub-gradient method (the parameters are estimated for each MC run). We compare these two DPIS estimates to an NIS estimate optimized using the same method (i.e., only one density q is optimized, which corresponds to setting $\alpha_2 = \alpha_1$ and $\beta_2 = \beta_1$), and finally to an NIS estimate with density q .

Fig. 1 displays the RMSE of the four estimates as a function of the number of samples N . We first observe that our first-order variance minimization produces translated and scaled samples $\{x_1^i, x_2^i\}_{i=1}^N$ that are better adapted to the estimation of μ than either the original samples $\{x^i\}_{i=1}^N$ drawn from the importance distribution q , or the samples $\{\tilde{x}^i\}_{i=1}^N$ adjusted from $\{x^i\}_{i=1}^N$ in order to fit simultaneously the numerator and the denominator; in other words, adjusting two densities, one for the numerator and one for the denominator, is significantly better than adjusting one single density to both. Next, even if the parameters ($\alpha_1, \alpha_2, \beta_1, \beta_2$) are deduced from an MC approximation of the first-order variance of $\hat{\mu}^{DPIS}$ (see (21)), the corresponding estimate gives almost the same performance as that which uses the optimal parameters. This shows that our technique has the potential to remain relevant in more complex models where the integrals are not available in closed-form. In order to illustrate this claim, Section 3.2 presents a simple example of such a model.

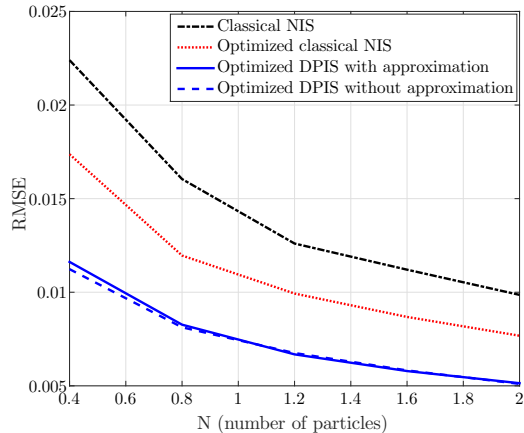


Fig. 1: Non-linear Gaussian model, $f(x) = x^2, \pi(x) = q(x) = \mathcal{N}(x, 0, 1)$. Simulation with 100 MC runs comparing DPIS estimators, using either exact or approximated integrals in the first-order variance expressions for minimization, with classical NIS and optimized classical NIS.

3.2. Bayesian, non-linear Gaussian setting

We now consider a Bayesian problem where x is hidden, y is observed and the target distribution $\pi(x) = p(x|y) \propto p(x)p(y|x)$ with $p(x) = \mathcal{N}(x, 0, 1), p(y|x) = \mathcal{N}(y, x^2, 1)$ and we want to estimate the moment of $f(x) = \mathcal{N}(x, 3, 1)$. In this setting, in order to compute RMSE values we take as a reference a classical NIS estimate using 10^7 samples. Fig. 2 displays the classical NIS estimate based on i.i.d. samples drawn from $q, \hat{\mu}^{DPIS}$ where the parameters have been estimated by the optimization method described in §2.2, and an NIS estimate optimized over only two parameters ($\alpha_2 = \alpha_1$ and $\beta_2 = \beta_1$). The parameters obtained in this scenario are ($\alpha_1 \approx 1.25, \beta_1 \approx 1, \alpha_2 \approx 2.45, \beta_2 \approx -0.1$), on average.

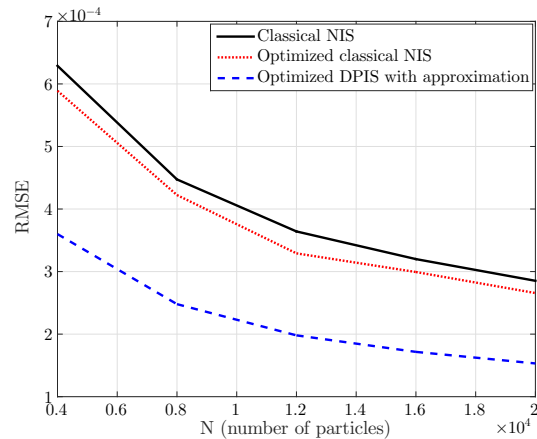


Fig. 2: Bayesian non-linear Gaussian model, $f(x) = \mathcal{N}(x, 3, 1), \pi(x) = q(x) = \mathcal{N}(x, 1, 1), p(y|x) = \mathcal{N}(y, x^2, 1)$. Simulation with 100 MC runs comparing the DPIS estimator using approximated variance integrals with classical NIS and optimized classical NIS.

4. CONCLUSIONS

In this paper we observed that NIS approximates the ratio of two expectations with one single set of random samples, which by construction cannot fit simultaneously the numerator and the denominator. We thus proposed an NIS technique based on the use of two instrumental distributions. In order to tune these proposals in practice we adjusted an initial set of samples drawn from some initial distribution q via two linear transforms, one targeting the numerator and the other the denominator. These linear transforms are optimized by minimizing the first-order variance of our DPIS estimate. Our simulations show that the resulting DPIS estimate outperforms NIS estimates which take into account only one set of samples, even if this set is adjusted according to the DPIS procedure. Possible perspectives for future work include improving the variance approximation provided by the delta method, which is only of the first order and may thus fail in certain models if N is insufficiently high, and assessing the empirical relevance of our optimized proposal density adjustment in more complex, high-dimensional, models.

5. REFERENCES

- [1] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods (Springer Texts in Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [2] H. Kahn and A.W. Marshall, "Methods of reducing sample size in Monte Carlo computations," *Journal of the Operations Research Society of America*, vol. 1, no. 5, pp. 263–278, 1953.
- [3] A. W. Marshall, "The use of multi-stage sampling schemes in Monte Carlo computations," in *Symposium on Monte Carlo Methods*, M. Meyer, Ed., New York, 1956, pp. 123–140.
- [4] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, Methuen and Co., London, 1964.
- [5] T. Hesterberg, *Advances in Importance Sampling*, Ph.D. thesis, Stanford University, 1988.
- [6] J. Geweke, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, vol. 57, no. 6, pp. 1317–39, 1989.
- [7] O. Cappé, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models (Springer Series in Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [8] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer Publishing Company, Incorporated, 2008.
- [9] E. Veach and L.J. Guibas, "Optimally combining sampling techniques for monte carlo rendering," in *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA, 1995, SIGGRAPH '95, pp. 419–428, ACM.
- [10] O. Cappé, A. Guillin, J.-M. Marin, and C.P. Robert, "Population Monte Carlo," *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 907–929, 2004.
- [11] R. Douc, A. Guillin, J.-M. Marin, and C.P. Robert, "Convergence of adaptive mixtures of importance sampling schemes," *The Annals of Statistics*, vol. 35, no. 1, pp. 420–448, February 2007.
- [12] Art B. Owen, *Monte Carlo theory, methods and examples*, 2013.
- [13] Dirk P. Kroese, Reuven Y. Rubinstein, and Peter W. Glynn, "Chapter 2 - the cross-entropy method for estimation," in *Handbook of Statistics*, C.R. Rao and V. Govindaraju, Eds., vol. 31 of *Handbook of Statistics*, pp. 19 – 34. Elsevier, 2013.