

Estimation of ratios of normalizing constants using stochastic approximation : the SARIS algorithm

Tom Guédon^{1*}, Charlotte Baey² and Estelle Kuhn¹

^{1*}Université Paris-Saclay, INRAE, MaIAGE, Domaine de Vilvert,
Jouy-en-Josas, 78350, France.

²Laboratoire Paul Painlevé, Université de Lille, Cité scientifique,
Villeneuve-d'Ascq, 59655, France.

*Corresponding author(s). E-mail(s): tom.guedon@inrae.fr;
Contributing authors: charlotte.baey@univ-lille.fr; estelle.kuhn@inrae.fr;

Abstract

Computing ratios of normalizing constants plays an important role in statistical modeling. Two important examples are hypothesis testing in latent variables models, and model comparison in Bayesian statistics. In both examples, the likelihood ratio and the Bayes factor are defined as the ratio of the normalizing constants of posterior distributions. We propose in this article a novel methodology that estimates this ratio using stochastic approximation principle. Our estimator is consistent and asymptotically Gaussian. Its asymptotic variance is smaller than the one of the popular optimal bridge sampling estimator. Furthermore, it is much more robust to little overlap between the two unnormalized distributions considered. Thanks to its online definition, our procedure can be integrated in an estimation process in latent variables model, and therefore reduce the computational effort. The performances of the estimator are illustrated through a simulation study and compared to two other estimators : the ratio importance sampling and the optimal bridge sampling estimators.

Keywords: ratio of normalizing constants, Monte Carlo methods, Stochastic approximation, marginal likelihood estimation

1 Introduction

In statistical modeling, comparing models often hinges on estimating ratios of integrals, which frequently serve as normalizing constants of posterior distributions. For example in latent variables models, such ratios emerge when choosing between two nested models via the likelihood ratio test. Each marginal likelihood represents the normalizing constant of the posterior distribution density of the latent variables given the data. However, computing marginal likelihoods defined as integrals becomes infeasible when the relationship between observed data and latent variables is complex. Similarly, in Bayesian statistics, model selection can be performed by comparing evidences (or marginal likelihoods) between competing models through the Bayes factor. In this setting, marginal likelihoods serve as the normalizing constants of the posterior distribution of the parameters given the data. Therefore, being able to efficiently compute ratios of normalizing constants is of significant practical interest.

This topic has motivated many fields of applications such as phylogenetics (Lartillot and Philippe, 2006), astrophysics (Russel et al., 2018), psychology (Annis et al., 2019) or chemical physics (Shirts and Chodera, 2008). To tackle this task one can either separately estimate the two likelihoods, or directly compute the ratio. As far as marginal likelihood estimation is concerned, several classical methods exist such as importance sampling (Robert and Casella, 1999), the harmonic mean estimator (Newton and Raftery, 1994) or the generalized harmonic mean estimator (Gelfand and Dey, 1994), using samples from the posterior distributions. It has been shown that those methods are often particular cases of estimators used to directly compute ratios of normalizing constants such as Bridge sampling (Meng and Wong, 1996) and ratio importance sampling (Chen and Shao, 1997b), both first introduced in the physics literature by respectively Bennett (1976) and Torrie and Valleau (1977). Gronau et al. (2017) highlights that the Bridge sampling estimator is superior to both the importance sampling and the general harmonic mean estimators, as it is more robust to the choice of the proposal distribution. However, its performance deteriorates as the overlap between the two densities associated with the normalizing constants decreases. To circumvent this issue, refinements were proposed in Meng and Schilling (2002) and Wang et al. (2022). They rely on a modification of the samples and their associated densities in order to increase the overlap between the densities. However, these refinements require some knowledge of the distributions under consideration, making the approach less attractive for practitioners. The optimal ratio importance sampling estimator presents the smallest asymptotic variance among the estimators mentioned here. However, the optimal scheme is not tractable in practice, which might explain why it is not often considered in the literature. Several other refined methods have been developed in the last decades, based on intermediate distributions that create a path between the two distributions. We can mention for example annealed importance sampling (Neal, 2001), sequential Monte Carlo (Del Moral et al., 2006), path sampling (Gelman and Meng, 1998) (or thermodynamic integration), stepping stone sampling (Xie et al., 2011) and generalized stepping stone sampling (Fan et al., 2011). These methods received a lot of interest and are presented and reviewed for instance in Friel and Wyse (2012) and more exhaustively in Llorente et al. (2023). However, most of these schemes can be seen as refinements of elementary methods to compute

ratios of normalizing constants presented in [Chen and Shao \(1997b\)](#). For example, path sampling is an extension of Bridge sampling ([Gelman and Meng, 1998](#)), and stepping stone sampling is an extension of the pre-umbrella identity using intermediate power posteriors distributions.

We propose here a new approach to calculate ratios of normalizing constants based on a stochastic approximation algorithm. Our procedure is also related to ratio importance sampling, and is therefore called Stochastic Approximation of Ratio Importance Sampling (SARIS). It benefits from the different refinements available in these two fields. After showing that estimating a ratio of normalizing constants is equivalent to finding the root of a function defined as an expectation, we develop an iterative stochastic approximation scheme to compute this root. We show that the sequence generated by the proposed algorithm converges almost surely towards the targeted ratio. The main advantage of our approach is that, thanks to its iterative construction, there is no need to fix the computational effort ahead since the procedure can be stopped once a convergence criterion has been reached. We also express the optimal proposal distribution in terms of asymptotic variance using convergence results from stochastic approximation theory. Moreover our method allows to reach the same asymptotic variance as the theoretical one of the optimal ratio importance sampling estimator.

This paper is organized as follows: the next section describes the context and the objective and gives a quick review of existing methods. [Section 3](#) presents the proposed SARIS procedures and studies their theoretical properties. [Section 4](#) is dedicated to algorithms using solely samples from both distributions involved in the targeted ratio. An extension for the simultaneous estimation of model parameters and likelihood ratio test statistic in latent variables models is also proposed. [Section 5](#) is dedicated to numerical experiments and practical guidelines. We conclude and discuss the perspectives in [Section 6](#). The proofs are postponed to the appendix.

2 Ratios of normalizing constants

In this section we introduce ratios of normalizing constants, the notations and their uses in statistics. We illustrate and motivate our purpose through two concrete examples: the likelihood ratio test statistic in latent variables models and the Bayes factor in Bayesian statistics.

2.1 Statistical setting and objective

Let d be a positive integer and μ a σ -finite positive Borel measure on a subspace \mathcal{Z} of \mathbb{R}^d . Assume that f_0 and f_1 are two positive integrable Borel functions on \mathcal{Z} such that $c_0 = \int_{\mathcal{Z}} f_0(z)\mu(dz) > 0$ and $c_1 = \int_{\mathcal{Z}} f_1(z)\mu(dz) > 0$. We assume that these normalizing constants c_0 and c_1 are unknown and introduce the two probability densities with respect to μ denoted by p_0 and p_1 defined for $i \in \{0, 1\}$ and for all z in \mathcal{Z} by

$$p_i(z) = \frac{f_i(z)}{c_i}.$$

The objective is to estimate the ratio r^* of normalizing constants defined as:

$$r^* = \frac{c_0}{c_1} = \frac{\int_{\mathcal{Z}} f_0(z) \mu(dz)}{\int_{\mathcal{Z}} f_1(z) \mu(dz)}. \quad (1)$$

We first motivate our contribution by two practical examples which require the computation of such ratios.

Example 1 (computation of likelihood ratio test statistic in latent variables model). *Let us consider a general latent variables model where the observed variable is given by the random variable Y , taking values in \mathcal{Y} and the latent variable by Z , taking values in \mathcal{Z} . We denote by y the observation of Y . We assume that the random vector (Y, Z) follows a parametric distribution with density f_θ parameterized by $\theta \in \Theta$. The objective is to test the hypotheses:*

$$H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \Theta_1,$$

where $\Theta_0 \subset \Theta_1 \subset \Theta$. A natural popular test is the likelihood ratio test ([Van der Vaart, 2000](#)) which statistic is defined by:

$$LR = -2 \log \left(\frac{L(\hat{\theta}_0; y)}{L(\hat{\theta}_1; y)} \right),$$

and where the marginal likelihood $L(\theta; y)$ and the maximum likelihood estimates $\hat{\theta}_i$, for $i \in \{0, 1\}$, are defined respectively by

$$L(\theta; y) = \int_{\mathcal{Z}} f_\theta(y; z) \mu(dz)$$

and

$$\hat{\theta}_i = \arg \max_{\theta \in \Theta_i} L(\theta; y).$$

Accurately estimating LR is crucial as its value determines whether H_0 is rejected or not.

Example 2 (computation of Bayes factor for Bayesian model choice). *In Bayesian statistics, the parameter $\theta \in \Theta$ is considered as a random variable with a known prior distribution $p(\theta)$. The posterior distribution of θ given a dataset D is defined as the product of the prior density and the likelihood of the model. To compare two models M_1 and M_2 and choose the one that better fits the data, the Bayes factor B_{12} ([Gelfand and Dey, 1994](#)) is a powerful tool. It is defined as:*

$$B_{12} = \frac{p(D | M_1)}{p(D | M_2)}$$

where $p(D | M_i) = \int_{\Theta} p(D | \theta, M_i) p(\theta) d\theta$ is the marginal likelihood of model M_i , for $i \in \{1, 2\}$.

2.2 State of the art

When the ratio (1) has no explicit expression, its computation can be performed by evaluating separately the numerator and the denominator. This can be done, for instance, using the harmonic mean estimator of [Newton and Raftery \(1994\)](#). In a Bayesian context, this estimator uses draws from the posterior distribution to compute the inverse of the marginal likelihood. However, it is known to overestimate the marginal likelihood and can have infinite variance. Another solution is to use importance sampling ([Robert and Casella, 1999](#)), which requires the introduction of a proposal distribution close to the integrand. However, it may be challenging to build such a proposal in complex settings. Furthermore, importance sampling is very sensitive to a misspecification of the proposal with respect to the density of interest ([Gronau et al., 2017](#)). These methods are specific cases of more general ones that aim at estimating ratios of normalizing constants.

In this section we focus on two existing methods that will serve as comparison for the proposed methodology: i) the Bridge sampling which is particularly popular ([Meng and Schilling, 2002](#); [Frühwirth-Schnatter, 2004](#); [Gronau et al., 2017, 2020](#)), and ii) the ratio importance sampling ([Chen and Shao, 1997b](#)) which is strongly linked to our approach. For a more precise and exhaustive review we refer the reader to [Llorente et al. \(2023, sections 4.1 and 4.2\)](#), and to [Chen and Shao \(1997b\)](#). In the sequel, we denote by Z any random variable defined on a probability space (Ω, \mathcal{A}, P) taking values in \mathcal{Z} , and by E_i the expectation with respect to density p_i for $i \in \{0, 1\}$.

Bridge sampling

First introduced in [Bennett \(1976\)](#) and later reintroduced in [Meng and Wong \(1996\)](#), the Bridge sampling is based on the following identity:

$$r^* = \frac{E_1 [f_0(Z)\alpha(Z)]}{E_0 [f_1(Z)\alpha(Z)]},$$

where α is a non-negative function defined on \mathcal{Z} verifying $0 < \int_{\mathcal{Z}} \alpha(z)p_0(z)p_1(z)\mu(dz) < +\infty$. Let K be a fixed positive integer. Then the Bridge sampling estimator of r^* is obtained using two K -samples $(Z_k^0)_{1 \leq k \leq K}$ and $(Z_k^1)_{1 \leq k \leq K}$ from p_0 and p_1 respectively as follows:

$$\hat{r}_K^{BS} = \frac{\sum_{k=1}^K f_0(Z_k^1)\alpha(Z_k^1)}{\sum_{k=1}^K f_1(Z_k^0)\alpha(Z_k^0)}. \quad (2)$$

This approach is particularly popular since in most statistical contexts it is straightforward to apply once a first inference step has been performed. This is the case for instance in the two examples introduced in the previous section. In the first example dealing with hypotheses testing in latent variables models, p_i is the posterior distribution of the latent variables given the data under hypothesis H_i . In the second example of Bayesian model choice, p_i is the posterior distribution of the parameter

given the data under model M_i . In both contexts, sampling from these distributions is part of the entire estimation process, therefore no additional work is required regardless of the complexity of the distributions.

Meng and Wong (1996) showed that the optimal choice of α , that minimizes the mean square error of the estimator and its asymptotic variance, is given by:

$$\alpha_{bridge}^{opt}(z) \propto \frac{1}{p_1(z) + p_0(z)} \propto \frac{1}{r^* f_1(z) + f_0(z)}. \quad (3)$$

It reaches the following optimal normalized asymptotic variance:

$$V_{bridge}^{opt} = 4r^{*2} \left[\left(\int_{\mathcal{Z}} \frac{2p_0(z)p_1(z)}{p_0(z) + p_1(z)} \mu(dz) \right)^{-1} - 1 \right]. \quad (4)$$

As α_{bridge}^{opt} depends on the unknown ratio r^* , it is not possible to use it directly in practice. Therefore the authors propose an iterative scheme to reach the optimal asymptotic variance. Starting from an initial guess $\hat{r}_K^{(0)}$, and using the two K -samples $(Z_k^0)_{1 \leq k \leq K}$ and $(Z_k^1)_{1 \leq k \leq K}$ from p_0 and p_1 defined above, we get:

$$\hat{r}_K^{(t+1)} = \left(\sum_{k=1}^K \frac{f_0(Z_k^1)}{\hat{r}_K^{(t)} f_1(Z_k^1) + f_0(Z_k^1)} \right) / \left(\sum_{k=1}^K \frac{f_1(Z_k^0)}{\hat{r}_K^{(t)} f_1(Z_k^0) + f_0(Z_k^0)} \right). \quad (5)$$

As t grows to infinity this estimator converges towards \tilde{r}_K^{BS} defined as:

$$\tilde{r}_K^{BS} = \left(\sum_{k=1}^K \frac{f_0(Z_k^1)}{\tilde{r}_K^{BS} f_1(Z_k^1) + f_0(Z_k^1)} \right) / \left(\sum_{k=1}^K \frac{f_1(Z_k^0)}{\tilde{r}_K^{BS} f_1(Z_k^0) + f_0(Z_k^0)} \right)$$

which can be rewritten as:

$$\sum_{k=1}^K \frac{f_0(Z_k^1)}{\tilde{r}_K^{BS} f_1(Z_k^1) + f_0(Z_k^1)} - \sum_{k=1}^K \frac{\tilde{r}_K^{BS} f_1(Z_k^0)}{\tilde{r}_K^{BS} f_1(Z_k^0) + f_0(Z_k^0)} = 0.$$

This last equation shows that the optimal Bridge sampling estimator can also be defined as the root of a function. Note that the solution to this equation nullifies the score function of Geyer's likelihood described in Geyer (1994), leading to the reverse logistic regression estimator.

Even if Bridge sampling is more robust than other methods as mentioned above, it still suffers from a too small overlap between distributions p_0 and p_1 . Indeed, when this overlap vanishes, the optimal variance grows to infinity. In such cases, more refined methods have been developed (see Meng and Schilling (2002); Wang et al. (2022)) that modify the two distributions considered without changing the normalizing constant. However, these methods are more involved since they require some additional effort to work properly.

Ratio importance sampling

The ratio importance sampling (RIS) estimator was first introduced in the physics literature in [Torrie and Valleau \(1977\)](#) as umbrella sampling and then rediscovered in [Chen and Shao \(1997b\)](#). It generalizes the importance sampling estimator to compute a ratio of normalizing constants. Considering a positive density function π on \mathcal{Z} , dominated by p_0 and p_1 , equation (1) can be written as follows:

$$r^* = \frac{\int_{\mathcal{Z}} f_0(z)\mu(dz)}{\int_{\mathcal{Z}} f_1(z)\mu(dz)} = \frac{\mathbb{E}_{\pi} \left[\frac{f_0(Z)}{\pi(Z)} \right]}{\mathbb{E}_{\pi} \left[\frac{f_1(Z)}{\pi(Z)} \right]}. \quad (6)$$

Equation (6) is called the *ratio importance sampling identity*, which will also be the basis of the methodology proposed in this paper. The ratio importance sampling estimator can be obtained using a $2K$ -sample $(Z_k)_{1 \leq k \leq 2K}$ from π as follows:

$$\hat{r}_K^{RIS} = \frac{\sum_{k=1}^{2K} f_0(Z_k)/\pi(Z_k)}{\sum_{k=1}^{2K} f_1(Z_k)/\pi(Z_k)}. \quad (7)$$

Remark 1. *Contrary to the Bridge sampling estimator that uses two samples, the Ratio Importance sampling estimator only requires one. That is why it is presented here using a sample of size $2K$.*

Note that identity (6) is very interesting and very general. For example, by taking $\pi = p_1$ it leads to $r^* = \mathbb{E}_1 \left[\frac{f_0(Z)}{f_1(Z)} \right]$ which gives an unbiased estimator of r^* . However, this estimator does not reach the optimal asymptotic variance. Indeed, [Chen and Shao \(1997b\)](#) showed that the optimal proposal density that minimizes the asymptotic variance of the estimator is given by:

$$\pi_{ris}^{opt}(z) \propto |p_1(z) - p_0(z)| \propto |f_1(z)r^* - f_0(z)|. \quad (8)$$

It reaches the following optimal asymptotic variance:

$$V_{ris}^{opt} = r^{*2} \left(\int_{\mathcal{Z}} |p_1(z) - p_0(z)|\mu(dz) \right)^2. \quad (9)$$

[Chen and Shao \(1997b\)](#), Theorem 3.3) showed that the variance of the optimal ratio importance sampling estimator is smaller than the variance of the optimal Bridge sampling estimator. Furthermore, when the overlap between p_0 and p_1 goes to 0, the optimal variance V_{ris}^{opt} converges to $4r^*$ which is bounded. This is a clear advantage compared to the previous methodology. However, contrary to the Bridge sampling setting, there is no straightforward procedure to approximate the optimal scheme and reach the optimal asymptotic variance. [Chen and Shao \(1997b\)](#) suggested a two-stage approach consisting in building a first consistent estimator \hat{r} of r^* based on a chosen method, and then use it as a plug-in estimator in (8). Since the first step can be difficult to achieve, the optimal scheme can be difficult to implement in practice. This might partly explain why RIS is not as popular as Bridge sampling.

3 Stochastic approximation procedures to compute ratio of normalizing constants

Given the limitations raised by the two approaches presented in the previous section, namely the sensitivity to a small overlap of the two distributions p_0 and p_1 for Bridge sampling, and the little practical applicability for RIS, there is a need for a new method that could address these issues.

In this section we propose an approach based on stochastic approximation principles. Our procedure convert the ratio importance sampling identity (6) into a root finding program, which brings several advantages. First, and contrary to usual Monte Carlo computation, there is no need to fix the sample size ahead of the procedure. Second, the stochastic approximation framework enables the use of sampling distributions that depend on the current estimate of the unknown ratio r^* , circumventing the main obstacle of RIS, but still enjoying its good theoretical properties. Indeed, the obtained sequence of estimates is almost surely convergent and asymptotically Gaussian. It reaches the same asymptotic variance as the one of the optimal RIS estimator, with an applicable scheme. As a consequence, it is much more robust to little overlap between p_0 and p_1 , avoiding the major drawback of Bridge sampling.

3.1 Description of the SARIS algorithm

Let π be a positive density function on \mathcal{Z} . Starting from equation (1), rewritten as:

$$\int_{\mathcal{Z}} (f_0(z) - r^* f_1(z)) \mu(dz) = 0, \quad (10)$$

we can write:

$$\mathbb{E}_{\pi} \left[\frac{f_0(Z) - r^* f_1(Z)}{\pi(Z)} \right] = 0, \quad (11)$$

where \mathbb{E}_{π} stands for the expectation with respect to the density π .

Calculating r^* is now equivalent to finding the root of a function defined as an expectation, and can therefore be solved using stochastic approximation algorithms. Assuming that r_0 is given in \mathbb{R} and that one can sample independent draws from π , we thus consider the sequence $(r_k)_{k \geq 0}$ of estimators of r^* , defined by the following recursion for every positive integer k :

$$r_{k+1} = r_k + \gamma_{k+1} \frac{f_0(Z_{k+1}) - r_k f_1(Z_{k+1})}{\pi(Z_{k+1})}, \quad \text{with } Z_{k+1} \sim \pi, \quad (12)$$

and where $(\gamma_k)_{k \geq 0}$ is a sequence of positive decreasing step sizes.

The main task in the construction of the sequence is the choice of the proposal distribution π . Most of the relevant choices for π might depend on the true ratio r^* . For example the optimal choice for the Bridge sampling involves the quantity $p_0 + p_1$, depending on r^* which can not be evaluated. This is also the main drawback to the use of optimal ratio importance sampling. Thanks to its iterative nature, our methodology allows to consider proposal distributions which might depend on r^* . More precisely, let us consider a positive density function π_r on \mathcal{Z} which depends on r . Equation (11) can be written as:

$$\mathbb{E}_{\pi_{r^*}} \left[\frac{f_0(Z) - r^* f_1(Z)}{\pi_{r^*}(Z)} \right] = 0. \quad (13)$$

Such equations can be solved using the Robbins-Monro algorithm (Robbins and Monro, 1951) that is based on the following stochastic recursion defined for every positive integer k :

$$r_{k+1} = r_k + \gamma_{k+1} \frac{f_0(Z_{k+1}) - r_k f_1(Z_{k+1})}{\pi_{r_k}(Z_{k+1})}, \quad \text{with } Z_{k+1} \sim \pi_{r_k} \quad (14)$$

The general algorithm is called SARIS (Stochastic Approximation Ratio Importance Sampling) and is summarized in Algorithm 1.

Algorithm 1 SARIS algorithm

Input: $(\gamma_k)_{k \geq 0}$, r_0 , stopping criterion
 Until stopping criterion:
 Draw Z_{k+1} from π_{r_k}
 Update $r_{k+1} = r_k + \gamma_k \frac{f_0(Z_{k+1}) - r_k f_1(Z_{k+1})}{\pi_{r_k}(Z_{k+1})}$
 $k = k + 1$
 Return r_k

Remark 2. *The iterative structure of our procedure enables to introduce a stopping criterion. This is not the case in many other methods, in particular for ratio importance sampling and Bridge sampling. Indeed in those two cases it is not possible to compute r_{k+1} given r_k , therefore one should fix the sampling size at the beginning of the procedure.*

Remark 3. *In most real-life applications, it is difficult to independently and exactly simulate from complex distributions. Therefore, the simulation step in (14) might be intractable. It is however possible to use the transition kernel of an ergodic Markov Chain having π_r as invariant distribution. One common practical choice for such Markov Chain Monte Carlo (MCMC) sampling scheme is the Metropolis-Hastings or the Metropolis-within-Gibbs algorithm (Robert and Casella, 1999). The recursive scheme (14) can therefore be generalized as follows for every positive integer k :*

$$r_{k+1} = r_k + \gamma_{k+1} \frac{f_0(Z_{k+1}) - r_k f_1(Z_{k+1})}{\pi_{r_k}(Z_{k+1})} \quad Z_{k+1} \sim \Pi_{r_k}(\cdot; Z_k) \quad (15)$$

where $\Pi_r(\cdot, \cdot)$ is a transition kernel of an ergodic Markov chain with invariant distribution π_r .

Remark 4. *We emphasize that the SARIS algorithm can also be used to compute a single marginal likelihood. If we know a density p only up to a normalizing constant c , $p = f/c$, then by introducing a known normalized density g , the SARIS algorithm can be used to compute the ratio $r^* = c$.*

Remark 5. *Finally, the proposed procedure enables to directly estimate any strictly monotonous and invertible transformation g of the ratio. Suppose that the objective*

is to compute $g(r^*)$ (for example $g = -2\log$ to obtain a likelihood ratio statistic). The recursive scheme (14) can be easily modified to estimate $g(r^*)$ with the sequence $(g_k)_{k \geq 0}$ defined as follows:

$$g_{k+1} = g_k + \gamma_{k+1} \frac{f_0(Z_{k+1}) - g^{-1}(g_k) f_1(Z_{k+1})}{\pi_{g^{-1}(g_k)}(Z_{k+1})} \quad Z_{k+1} \sim \pi_{g^{-1}(g_k)} \quad (16)$$

Algorithm 1 can be easily adapted using this recursion, only changing the updating rule.

3.2 Theoretical property of the SARIS algorithm

In this section we study the theoretical convergence property of the sequence $(r_k)_{k \geq 0}$ generated by the SARIS procedure described in Algorithm 1. We emphasize that such a setting is less general than the recursion defined in (15), however its theoretical study corresponds to the one of Robbins and Monro (1951). Moreover it allows a more fair comparison with the Monte Carlo methods presented in section 2.2, for which the theory was established for independent and identically distributed sampling.

We first state some regularity assumptions on the functions f_0, f_1 and on the densities $\{\pi_r, r \in \mathbb{R}\}$:

Assumption 1. *The functions f_0 and f_1 are positive integrable and for every $z \in \mathcal{Z}$, $r \mapsto \pi_r(z)$ is continuous.*

Assumption 2.

$$\mathbb{E}_0 \left(\sup_{r \in \mathbb{R}} \left| \frac{f_0(Z) - r f_1(Z)}{\pi_r(Z)} \right| \right) + \mathbb{E}_1 \left(\sup_{r \in \mathbb{R}} \left| \frac{f_0(Z) - r f_1(Z)}{\pi_r(Z)} \right| \right) < +\infty.$$

This assumption ensures the integrability of the main quantities involved in the algorithm. We also state a common assumption on the sequence of step sizes $(\gamma_k)_{k \geq 0}$.

Assumption 3. *The sequence $(\gamma_k)_{k \geq 0}$ is positive, decreasing and verifies $\sum_{k=0}^{+\infty} \gamma_k = +\infty$ and $\sum_{k=0}^{+\infty} \gamma_k^2 < +\infty$.*

We can now state the almost sure (a.s.) convergence of the sequence (r_k) .

Proposition 1. *Considering the sequence $(r_k)_{k \geq 0}$ generated by Algorithm 1, under Assumptions 1, 2 and 3, we get:*

$$\lim_{k \rightarrow +\infty} r_k = r^* \quad a.s.$$

The proof is postponed to the appendix. We now require a stronger regularity assumption on the functions f_0, f_1, π_r to derive the asymptotic distribution of the sequence $(r_k)_{k \geq 0}$.

Assumption 4. *The functions f_0 and f_1 are not proportional to each other.*

Assumption 5. *There exists $\delta > 0$ such that*

$$\sup_{k \geq 0} \mathbb{E}_0 \left(\left| \frac{f_0(Z) - r_k f_1(Z)}{\pi_{r_k}(Z)} \right|^{1+\delta} \right) + \sup_{k \geq 0} \mathbb{E}_1 \left(\left| \frac{f_0(Z) - r_k f_1(Z)}{\pi_{r_k}(Z)} \right|^{1+\delta} \right) < +\infty.$$

Assumption 6. *There exists $\frac{1}{2} < \epsilon < 1$, $a > 0$, $b > 0$ such that the sequence of step sizes (γ_k) is of the form $\gamma_k = \frac{a}{b+k^\epsilon}$.*

These integrability and step sizes assumptions are classical ones to obtain asymptotic normality results for martingales. The next result states the asymptotic normality of the averaged sequence defined as $(r_k^{AV})_{k \geq 0} = \left(\frac{1}{k} \sum_{j=0}^k r_j \right)_{k \geq 0}$.

Proposition 2. *Considering the sequence $(r_k)_{k \geq 0}$ generated by Algorithm 1 and its averaged version $(r_k^{AV})_{k \geq 0}$, under Assumptions 1, 4, 5 and 6, we get:*

$$\sqrt{k} (r_k^{AV} - r^*) \xrightarrow[k \rightarrow +\infty]{d} \mathcal{N}(0, V_{saris}(\pi_{r^*}))$$

with

$$V_{saris}(\pi_{r^*}) = \frac{1}{c_1^2} \mathbb{E}_{r^*} \left[\left(\frac{f_0(Z) - r^* f_1(Z)}{\pi_{r^*}(Z)} \right)^2 \right]$$

where the expectation \mathbb{E}_{r^*} is taken with respect to the density π_{r^*} . Furthermore, the optimal proposal π_{saris}^{opt} defined as the one which minimizes the asymptotic variance is given as

$$\pi_{r^*}^{opt}(z) \propto |p_1(z) - p_0(z)|,$$

corresponding to the optimal variance:

$$V_{saris}^{opt} = r^{*2} \left(\int_{\mathcal{Z}} |p_1(z) - p_0(z)| \mu(dz) \right)^2.$$

The proof is postponed to the appendix and relies on similar arguments as the derivation of optimal importance function in importance sampling (see [Robert and Casella \(1999\)](#) for more details). We achieve the same optimal variance and retrieve the optimal proposition density of ratio importance sampling. Moreover in [Chen and Shao \(1997b\)](#) the authors show that this variance is smaller than the variance of the optimal Bridge sampling estimator.

Remark 6. *If sampling is done through the use of a Markov transition kernel, similar theoretical results can be obtained assuming additional regularity conditions on the Markov kernel. For further details we refer to [Allasonniere and Kuhn \(2015\)](#) and [Fort \(2015\)](#).*

It is important to notice that the algorithm presented in this section is not always applicable. In particular, the analytical expression of the optimal proposal density $\pi_{r^*}^{opt}$ is unknown, and therefore the update rule of the sequence r_k defined in Algorithm 1 is not computable. The next section solves this issue.

3.3 A practical extension of the SARIS algorithm

Suppose that for all z , $\pi_r(z) = \tilde{\pi}_r(z)/c(r)$ where the analytical expression of $\tilde{\pi}_r(z)$ is known, then equation (11) is equivalent to:

$$\mathbb{E}_{\pi_{r^*}} \left[\frac{f_0(Z) - r^* f_1(Z)}{\tilde{\pi}_{r^*}(Z)} \right] = 0. \quad (17)$$

Based on this identity, one can consider the following extension of Algorithm 1 called SARIS-EXT, that uses the known unnormalised density $\tilde{\pi}_r$:

Algorithm 2 SARIS-EXT algorithm

Input: $(\gamma_k)_{k \geq 0}$, r_0 , stopping criterion
Until stopping criterion:
Draw Z_{k+1} from $\pi_{r_k} \propto \tilde{\pi}_{r_k}$
Update $r_{k+1} = r_k + \gamma_k \frac{f_0(Z_{k+1}) - r_k f_1(Z_{k+1})}{\tilde{\pi}_{r_k}(Z_{k+1})}$
 $k = k + 1$
Return r_k

We state mild additional conditions on $\tilde{\pi}_r$ similar to Assumptions 2 and 5, to prove that the sequence $(r_k)_{k \geq 0}$ generated by Algorithm 2 verifies the same properties as the one generated by Algorithm 1.

Assumption 7.

$$\mathbb{E}_0 \left(\sup_{r \in \mathbb{R}} \left| \frac{f_0(Z) - r f_1(Z)}{\tilde{\pi}_r(Z)} \right| \right) + \mathbb{E}_1 \left(\sup_{r \in \mathbb{R}} \left| \frac{f_0(Z) - r f_1(Z)}{\tilde{\pi}_r(Z)} \right| \right) < +\infty.$$

Proposition 3. *Considering the sequence $(r_k)_{k \geq 0}$ generated by Algorithm 2, under Assumptions 1, 3 and 7, we get:*

$$\lim_{k \rightarrow +\infty} r_k = r^* \quad a.s.$$

Assumption 8. *There exists $\delta > 0$ such that*

$$\sup_{k \geq 0} \mathbb{E}_0 \left(\left| \frac{f_0(Z) - r_k f_1(Z)}{\tilde{\pi}_{r_k}(Z)} \right|^{1+\delta} \right) + \sup_{k \geq 0} \mathbb{E}_1 \left(\left| \frac{f_0(Z) - r_k f_1(Z)}{\tilde{\pi}_{r_k}(Z)} \right|^{1+\delta} \right) < +\infty.$$

Assumption 9. *There exists a neighborhood U of r^* such that $r \mapsto h(r) = \frac{c_0 - r c_1}{c(r)}$ is continuously differentiable and $h'(r^*) < 0$.*

Assumption 10. *For all $r \in \mathbb{R}$ the set $\mathcal{A}_r = \{z \in \mathcal{Z}, f_0(z) = r f_1(z)\}$ satisfies $\mu(\mathcal{A}_r) = 0$.*

Remark 7. *i) Assumption 9 is required to apply a central limit theorem on the sequence $(r_k)_{k \geq 0}$. For the SARIS algorithm, $h(r) = c_0 - r c_1$ which does not require regularity conditions, as it is affine. ii) Assumption 10 is required to ensure that $c^{opt}(r)$ defined as the normalizing constant of $z \mapsto \tilde{\pi}_r^{opt}(z) = |f_0(z) - r f_1(z)| \propto \pi_r^{opt}(z)$ verifies Assumption 9. The following proposition extends the theoretical results of the SARIS algorithm to the SARIS-EXT algorithm.*

Proposition 4. *Considering the sequence $(r_k)_{k \geq 0}$ generated by Algorithm 2 and its averaged version $(r_k^{AV})_{k \geq 0}$, under Assumptions 1, 4, 6, 8 and 9, we get:*

$$\sqrt{k} (r_k^{AV} - r^*) \xrightarrow[k \rightarrow +\infty]{d} \mathcal{N}(0, V_{ext}(\tilde{\pi}_{r^*}))$$

Moreover we have :

$$V_{ext}(\tilde{\pi}_{r^*}) = V_{saris}(\pi_{r^*}).$$

Furthermore, under Assumption 10, an optimal unnormalised proposal $\tilde{\pi}_r^{opt}$ defined as one of which minimizes the asymptotic variance is given as

$$\tilde{\pi}_{r^*}^{opt}(z) = |f_0(z) - r^* f_1(z)| \propto \pi_{r^*}^{opt}(z)$$

corresponding to the optimal variance:

$$V_{saris}^{opt} = r^{*2} \left(\int_{\mathcal{Z}} |p_1(z) - p_0(z)| \mu(dz) \right)^2.$$

This proposition shows that theoretically, the extended algorithm has the same asymptotic performances as the initial SARIS algorithm.

4 Non optimal methods using draws from p_0 and p_1

4.1 Estimating ratios of normalizing constants using only distributions p_0 and p_1

It is interesting to compare methods that use draws solely from p_0 and p_1 . As explained in Section 2.1 with the two examples, in latent variable models, these represent the posterior distributions of latent variables given data under two different hypotheses. In Bayesian inference, they denote the posterior distributions of parameters under two distinct models. In both contexts, draws from these distributions are essential for inference, making it convenient to consider proposal distributions based on them. Numerically, this allows for the reuse of already simulated samples. Practically, it simplifies the method by eliminating the need to build new samplers.

A natural choice is a mixture of p_0 and p_1 , similar to the approach proposed by [Chen and Shao \(1997b\)](#) in Section 5, and closely related to bridge sampling. We define the proposal density based on this mixture as follows:

$$z \mapsto \pi_{r^*}^{mixt}(z) = \frac{1}{2} \{p_0(z) + p_1(z)\} \propto f_0(z) + r^* f_1(z) \quad (18)$$

As long as we know how to draw samples from p_1 and p_0 , it is easy to sample from this mixture, by sampling uniformly randomly from one or the other distribution. Of course the analytical expression of π_r^{mixt} is unknown, but the extended algorithm is applicable considering the unnormalised density $\tilde{\pi}_r^{mixt}(z) = f_0(z) + r f_1(z)$. The

simulating step in Algorithm 2 requires to sample from $\pi_{r_k}^{mixt}$, which verifies for every $z \in \mathcal{Z}$:

$$\begin{aligned} \pi_{r_k}^{mixt}(z) &\propto f_0(z) + r_k f_1(z) \\ &\propto c_0 p_0(z) + r_k c_1 p_1(z) \\ &\propto r^* p_0(z) + r_k p_1(z) \\ \iff \pi_{r_k}^{mixt} &= \left(1 - \frac{r_k}{r_k + r^*}\right) p_0(z) + \frac{r_k}{r_k + r^*} p_1(z) \end{aligned} \quad (19)$$

where the weights of the mixture depend on r^* . Therefore it is not possible to sample from $\pi_{r_k}^{mixt}$ using simply draws from p_0 and p_1 .

Remark 8. *We emphasize that even if it is not possible to sample from $\pi_{r_k}^{mixt}$ using draws from p_0 and p_1 , it is still possible to use a MCMC procedure to sample from $\pi_{r_k}^{mixt} \propto f_0 + r_k f_1$, but this loses the practical benefits of using separately the two distributions. This scheme is still illustrated in the simulation study.*

However, as simulating from the mixture $\pi_{r^*}^{mixt}$ is possible, one can consider the following alternative recursive scheme :

$$r_{k+1} = r_k + \gamma_{k+1} \frac{f_0(Z_{k+1}) - r_k f_1(Z_{k+1})}{\tilde{\pi}_{r_k}^{mixt}(Z_{k+1})}, \quad Z_{k+1} \sim \pi_{r^*}^{mixt} \quad (20)$$

which defines a new estimation procedure that is summarized in Algorithm 3, and called SARIS-MIXT.

Algorithm 3 SARIS-MIXT algorithm

Input: $(\gamma_k)_{k \geq 0}$, r_0 , stopping criterion

Until stopping criterion:

Draw Z_{k+1} from $\pi_{r^*}^{mixt} = \frac{1}{2}(p_0 + p_1)$

Update $r_{k+1} = r_k + \gamma_k \frac{f_0(Z_{k+1}) - r_k f_1(Z_{k+1})}{f_0(Z_{k+1}) + r_k f_1(Z_{k+1})}$

$k = k + 1$

Return r_k

Remark 9. *Contrary to the SARIS-EXT algorithm, in this procedure the distribution used for the sampling step does not depend on the current ratio r_k .*

It can be shown that under mild conditions, the sequence generated by Algorithm 3 converges almost surely towards r^* and is asymptotically Gaussian. The proof follows the same lines as for the two other SARIS algorithms, the main steps are given in the proof of the next result.

Unfortunately, the estimator obtained from the algorithm SARIS-MIXT presents a higher asymptotic variance than the one obtained using SARIS-EXT with $\tilde{\pi}_r^{mixt}$ as proposal, $V_{ext}(\tilde{\pi}_{r^*}^{mixt})$. The following proposition formalizes this statement.

Proposition 5. *Let V_{saris}^{mixt} be the asymptotic variance of the averaged sequence generated by the SARIS-MIXT Algorithm 3. Let $\tilde{\pi}_r^{mixt}(z) = f_0(z) + r f_1(z)$ be the*

unnormalised mixture between p_0 and p_1 . Let Ψ be the quantity defined as:

$$\Psi = \int_{\mathcal{Z}} \frac{p_1(z)p_0(z)}{\frac{1}{2}\{p_1(z) + p_0(z)\}} dz$$

Under Assumptions 1, 4, 6, we get :

$$V_{ext}(\tilde{\pi}_{r^*}^{mixt}) = 4r^{*2} (1 - \Psi) = \Psi^2 V_{saris}^{mixt}.$$

The quantity Ψ can be seen as an overlap index between p_0 and p_1 . It is easy to show that $0 \leq \Psi \leq 1$. If the two distributions have disjoint supports, then $\Psi = 0$. If $p_0 = p_1$ then $\Psi = 1$.

This index Ψ is very convenient to compare the asymptotic variances derived in Proposition 5 with those of the optimal Bridge sampling estimator and ratio importance sampling estimator using $\pi_{r^*}^{mixt}$ as proposal. For a detailed description of the later called Bridge-like ratio importance sampling method (RIS-MIXT), we refer to [Chen and Shao \(1997b\)](#), section 5. Let V_{ris}^{mixt} denote the asymptotic variance of the Bridge-like ratio importance sampling estimator. The following relationship exists between the various variances discussed in this paragraph:

$$V_{ext}(\tilde{\pi}_{r^*}^{mixt}) = \Psi \times V_{bridge}^{opt} = \Psi^2 \times V_{ris}^{mixt} = \Psi^2 \times V_{saris}^{mixt} \quad (21)$$

Note that the SARIS-MIXT estimator reaches the same asymptotic variance as the Bridge-like ratio importance sampling estimator, which is not surprising as they are both based on the same identity (6).

It is noticeable that $V_{ext}(\tilde{\pi}_{r^*}^{mixt})$ is bounded, unlike the other variances that diverge as Ψ approaches 0. In fact, when $\Psi = 0$, π_r^{mixt} aligns with π_r^{opt} . Therefore, finding a way to sample from the distribution defined by (19) using only draws from p_0 and p_1 would be very beneficial. This would allow for variance reduction, as described in equation (21), while still maintaining the simplicity of simulating only from p_0 and p_1 . This result indicates that Bridge sampling is still the best method when using only draws from p_0 and p_1 . However, when Ψ is closed to one (which is the case with a lot of overlap between p_0 and p_1), these methods remain comparable.

The next section explores the extension of the SARIS-MIXT algorithm to a joint procedure with parameter inference in latent variables models.

4.2 A joint procedure for model parameter estimation and LRT statistic computation in latent variables models

In this section, we extend the use of the SARIS estimator based on mixtures between p_0 and p_1 to the context of likelihood ratio test (LRT) in latent variables model and introduce a joint procedure for model parameter inference and LRT statistic computation.

Consider two random variables Y on \mathcal{Y} and Z on \mathcal{Z} . Assume that the joint density of (Y, Z) belongs to a parametric family $\{f_\theta, \theta \in \Theta\}$, with $\Theta \subset \mathbb{R}^p$ with p a positive

integer. We only observe a realization y of Y , the random variable Z being unobserved. The maximum likelihood estimator $\hat{\theta}$ is defined as:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta; y)$$

where the marginal likelihood $L(\theta; y)$ is equal to the complete likelihood integrated over the latent variable:

$$L(\theta; y) = \int_{\mathcal{Z}} f_{\theta}(y, z) \mu(dz)$$

The above integral is often untractable, which makes the optimization process difficult. To solve this issue, stochastic methods can be used. Two popular ones are the stochastic approximation expectation maximization (SAEM) algorithm (Delyon et al., 1999; Kuhn and Lavielle, 2004) and the stochastic gradient descent (SGD) algorithm (Baey et al., 2023). Both methods require draws from the posterior distribution of the latent variables given the data whose density is denoted by p_{θ} in the sequel.

Both SAEM and SGD are iterative algorithms that can be summarized as follows, at each step $k > 0$:

1. Draw Z_k from p_{θ_k} .
2. Update θ_k with a gradient step when using SGD or a maximization step when using SAEM.

Consider now the context of Example 1 where the objective is to test the hypotheses:

$$H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \Theta_1$$

where $\Theta_0 \subset \Theta_1 \subset \Theta$. The LRT statistic equals:

$$LR = -2 \log \left(\frac{L(\hat{\theta}_0; y)}{L(\hat{\theta}_1; y)} \right).$$

We propose to combine the estimation procedures for $\hat{\theta}_0$ and $\hat{\theta}_1$ with the computation of the marginal likelihood ratio $r^* = \exp(-LR/2)$, taking advantage of the computational effort of the inference task. The procedure is detailed in Algorithm 4.

Algorithm 4 Joint parameter and LRT statistic estimation in latent variables models

Input: $z_{0,0}, z_{1,0}, \theta_{0,0}, \theta_{1,0}, r_0, (\gamma_k)_{k \geq 0}$ $k = 0$

Until convergence criterion:

Draw $z_{0,k+1}$ from $p_{\theta_{0,k}}$ and $z_{1,k+1}$ from $p_{\theta_{1,k}}$ Update $\theta_{0,k+1}$ and $\theta_{1,k+1}$ using a SGD or SAEM stepDraw \tilde{z}_{k+1} from a uniform distribution on $\{z_{0,k+1}, z_{1,k+1}\}$ Update r_{k+1} as:

$$r_{k+1} = r_k + \gamma_k \frac{f_{\theta_{0,k+1}}(y, \tilde{z}_{k+1}) - r_k f_{\theta_{1,k+1}}(y, \tilde{z}_{k+1})}{f_{\theta_{0,k+1}}(y, \tilde{z}_{k+1}) + r_k f_{\theta_{1,k+1}}(y, \tilde{z}_{k+1})}$$

 $k = k + 1$ Return $r_k, \theta_{0,k}, \theta_{1,k}$

Remark 10. When the two estimation processes can not be applied jointly, this procedure can be carried out post-estimation, provided the sequences $(\theta_{i,k}, z_{i,k})_{i=0,1, k \geq 0}$ are kept in memory.

Remark 11. If only one marginal likelihood is to be estimated, the procedure applies by introducing a proposal density q (or a sequence of proposal densities $(q_k)_{k \geq 0}$) from which it is possible to sample from and to proceed as follows:

1. Draw z_{k+1} from p_{θ_k}
2. Update θ_{k+1} with SGD or SAEM step
3. With probability 0.5 define $\tilde{z}_{k+1} = z_{k+1}$, otherwise draw \tilde{z}_{k+1} from $q_{k+1}(\cdot)$
4. Update r_{k+1} as:

$$r_{k+1} = r_k + \gamma_k \frac{f_{\theta_{k+1}}(y, \tilde{z}_{k+1}) - r_k q_{k+1}(\tilde{z}_{k+1})}{f_{\theta_{k+1}}(y, \tilde{z}_{k+1}) + r_k q_{k+1}(\tilde{z}_{k+1})}$$

The R package `bridgesampling` (Gronau et al., 2020) proposes to use a Gaussian approximation as a second distribution when considering the computation of a single marginal likelihood. However, in our procedure prior knowledge on the distribution of interest (mean and variance for example) is not available as inference has not been performed yet. A possible approach to overcome this issue could be to use a Gaussian proposal with adaptive mean m_k and variance σ_k^2 , where m_k and σ_k^2 are defined at each step k as follows:

$$\begin{aligned} m_{k+1} &= m_k + \gamma_k (z_{k+1} - m_k) \\ v_{k+1} &= v_k + \gamma_k (z_{k+1}^2 - v_k) \\ \sigma_{k+1}^2 &= v_{k+1} - m_{k+1}^2 \end{aligned}$$

5 Numerical experiments and practical considerations

This section is devoted to numerical experiments. We first illustrate the performances of the three SARIS estimators compared to the RIS estimator of Chen and Shao

(1997a) and the optimal Bridge sampling estimator of Meng and Wong (1996). We then provide an example of the joint procedure introduced in Section 4.2.

5.1 Simulation study in a one dimensional Gaussian setting

We first illustrate our method with a one dimensional Gaussian setting. We consider two Gaussian distributions, $f_0 = \phi$ and $f_1 = \phi(\cdot - \mu)$ where ϕ is the standardized Gaussian density and $\mu \in \mathbb{R}$. Since these densities are already normalized, we have $c_0 = c_1 = 1$ and therefore $r^* = 1$.

In the simulation study, we compare the performances of three SARIS estimators presented in this paper. These methods are compared with the optimal bridge sampling estimator (BRIDGE OPT) and the RIS estimator based on $\pi_{r^*}^{mixt}$ (RIS-MIXT). For the entire simulation study, the sampling steps are performed using one step of an adaptive Metropolis Hastings (MH) algorithm (see for example Roberts and Rosenthal (2009, section 3)) in order to stick to most real life applications, implemented manually.

The three SARIS estimators presented in Figure 1 are the following:

1) *Optimal SARIS extended using $\tilde{\pi}_r^{opt}$ (SARIS-EXT opt)*

This is the estimator generated by Algorithm 2 using $\tilde{\pi}_r^{opt}(z) = |f_0(z) - rf_1(z)|$. In this procedure, the increment $\frac{f_0(Z_{k+1}) - rf_1(Z_{k+1})}{|f_0(Z_{k+1}) - rf_1(Z_{k+1})|} \in \{0; 1\}$ only takes two values. The drawback is that the increment has no intensity, i.e. it gives no indication on the order of magnitude of the descent step. However, it can still solve computational issues, in particular when the evaluation of the likelihood can be complicated. In order to use this method, one only need to know how to evaluate the unnormalised densities up to a non decreasing transformation, as only comparison of them is required, and not their evaluation.

2) *SARIS using mixture between p_0 and p_1 (SARIS-MIXT)*

This distribution has already been discussed in Section 4. As long as we know how to draw samples from p_1 and p_0 , it is easy to sample from the mixture by randomly sampling from one or the other distribution with probability 1/2:1/2. This estimator corresponds to the one generated by Algorithm 3.

3) *SARIS extended using $\tilde{\pi}_r^{mixt} = f_0(z) + rf_1(z)$ (SARIS-EXT-mixt)*

This proposal is not of practical use, as it neither uses draws from p_0 and p_1 nor is an approximation of the optimal scheme. However, it is interesting to distinguish it from the previous one, as they are very similar. This distribution is also a mixture between p_0 and p_1 as explained in Section 4. Even if it is supposed to approximate $\pi_{r^*}^{mixt}$, it benefits from a significant gain in performance, as Proposition 5 theoretically justifies it, and Figure 1 illustrates it.

Note that the three different proposal considered above lead to a bounded increment. This property is crucial both from a practical point of view, as it enhances

the numerical stability of the procedure, and a theoretical one, as it guarantees Assumptions 2 and 5 to be verified.

Remark 12. *The optimality criterion considered in this article is given by the asymptotic variance of the exact sampling scheme, which is most of the time intractable in practice. Indeed, in most real life applications sampling is performed through the use of transition kernels in MCMC algorithms. In this setting, even if central limit theorems may apply under regularity conditions, the asymptotic variance is in general not explicit and there is no guarantee that the proposal which minimizes this variance is the same as in the exact sampling case. Therefore it might still be of practical interest to consider other proposals. Some proposal distributions are worth mentioning, e.g. the geometric mean between p_0 and p_1 discussed in Meng and Wong (1996) and Chen and Shao (1997a) for example, or an element of the q -path between p_0 and p_1 that generalizes the harmonic mean and geometric mean between the two distributions (see Brekelmans et al. (2020) for details).*

For each of the SARIS procedure the following step size is considered:

$$\begin{cases} \gamma_k = 0.1 & \text{for } k < K_{heat} \\ \gamma_k = \frac{0.1}{1+k^{2/3}} & \text{otherwise,} \end{cases} \quad (22)$$

which is standard in the stochastic gradient descent literature. It verifies Assumption 3 and 6 while presenting a heating phase that enables a wider exploration of the parameter space at the beginning of the algorithm.

As discussed in Section 2.2, the optimal Bridge sampling estimator is not available in a closed form, so we relied on the recursive algorithm described in equation (5), which was run until convergence. For the RIS estimator, we used equation (7) using $\pi_{r^*}^{mixt}$ as proposal distribution.

We adjusted the sample sizes of each method to make sure that they are all comparable in terms of number of calls to functions f_0 and f_1 . Four MH samplers were implemented. The first two are MH samplers whose invariant distributions are p_0 and p_1 . These are used for BRIDGE-OPT, RIS-MIXT and SARIS-MIXT. Then two samplers generating non homogenous Markov chains with, at each step, invariant distributions being $\pi_{r_k}^{opt}$ and $\pi_{r_k}^{mixt}$, were built to compute respectively the SARIS-EXT-opt and the SARIS-EXT-mixt estimators.

For each of the estimators, a budget of $2K+2K_{heat}$ draws were allocated, with $2K$ samples used to compute the estimators and $2K_{heat}$ for heating the MH samplers. For the experiments, we used $K = 5000$ and $K_{heat} = 300$.

We also considered the estimation of $\log(r^*)$ using recursion (16). From a theoretical point of view, the use of this transformation is equivalent to using a delta method. Therefore, results obtained for the estimation of $\log r^*$ or r^* are comparable in terms of performances. However, since the numerical stability of the procedure is greater in the former case, in the sequel we only present results associated with the estimation of $\log r^*$.

We first consider two cases: 1) a strong overlap between p_0 and p_1 with $\mu = 1$ and 2) a small overlap between p_0 and p_1 with $\mu = 5$. Results are presented in Figure 1.

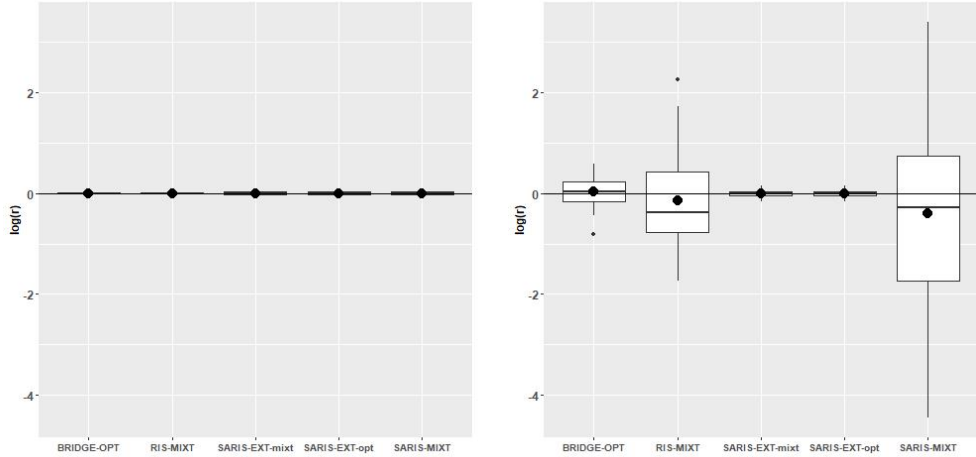


Fig. 1 Estimation of $\log r^*$ using the optimal bridge sampling estimator (BRIDGE-OPT), the bridge like ratio importance sampling (RIS-MIXT), SARIS-EXT-mixt, SARIS-EXT-opt and SARIS-MIXT. The black line represents the true value $\log(r^*) = 0$. The boxplots were computed on 50 repetitions of the experiments. The two graphs illustrate strong overlap between p_0 and p_1 (on the left, in the case $\mu = 1$) and little overlap (on the right, in the case $\mu = 5$)

As the theory suggests, the optimal bridge sampling estimator outperforms the two other methods based on samples from p_0 and p_1 . However, when considering the case $\mu = 1$, the five methods present performances of the same order. The fact that the bridge like ratio importance sampling performs better than the SARIS-MIXT estimator can be interpreted by the fact that the ratio importance sampling estimator imposes at each step the estimator to solve the empirical version of the SARIS identity (11), which might enhance the stability of the procedure. This difference should be diminished by considering adaptive step sizes, to mimic the differences between \hat{r}_K^{ris} and \hat{r}_{K+1}^{ris} .

Figure 1 illustrates the fact that the SARIS-EXT estimators presented are much more robust to little overlap than ratio importance sampling and Bridge sampling. However, it also illustrates the fact that in more simple cases such as the one illustrated in Figure 1, methods that only use draws from p_0 and p_1 , which are easier to apply might be sufficient.

To show intermediate results between the cases $\mu = 1$ and $\mu = 5$, and worse ones, we compare the optimal Bridge sampling estimator with the SARIS-EXT-opt estimator for varying values of μ between 1 and 10. For the simulations we use $K = 5000$ samples for each expectation in the bridge sampling estimation procedure, and therefore use $2K$ samples for the SARIS procedure. Results are displayed in Figure 2. For each value of μ , 50 repetitions were computed. The dots represent the empirical means and the errorbars the empirical standard deviations computed over the 50 repetitions. Figure 2 illustrates the robustness of the proposed procedure in comparison to the bridge sampling estimator that highly deteriorates when the overlap reduces.

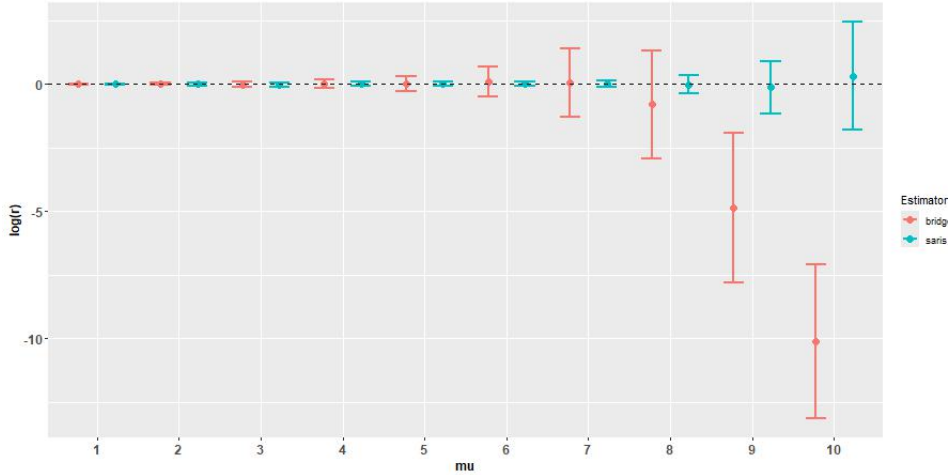


Fig. 2 Estimation of the log ratio of normalizing constants of the densities of a $\mathcal{N}(0, 1)$ and the one of a $\mathcal{N}(\mu, 1)$ for varying values of μ , using the optimal Bridge sampling estimator BRIDGE-OPT (red) and SARIS-EXT-opt to compute the ratio. The dots represent the empirical means and the error bars the empirical standard deviations computed over 50 repetitions.

This makes the proposed procedure appealing when little information is available on the distributions.

5.2 Joint estimation in latent variables models

To illustrate the joint procedure presented above, we consider the following model of linear regression with missing values. Let $i = 1, \dots, n$, we observe the response $y_i \in \mathbb{R}$ modeled as :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \quad (23)$$

where $(\varepsilon_i)_{i=1, \dots, n}$ is a sequence of independent and identically distributed Gaussian noise with known variance σ^2 , $\beta = (\beta_0, \beta_1, \beta_2)^T$ is an unknown vector of regression coefficients, and $(x_i)_{i=1, \dots, n} = (x_{i1}, x_{i2})_{i=1, \dots, n}^T$ is an independent and identically distributed sample of covariates from a $\mathcal{N}\left((\mu_1, \mu_2)^T, \begin{pmatrix} \gamma_1^2 & 0 \\ 0 & \gamma_2^2 \end{pmatrix}\right)$. We suppose that for $i = 1, \dots, r$ we only observe (y_i, x_{i1}) and for $i = r+1, \dots, n$ we observe (y_i, x_{i1}, x_{i2}) . This example is borrowed from the lecture notes of Julie Josse "Handling Missing values" available on this [website](https://juliejosse.com/wp-content/uploads/2018/07/LectureMissing-Weij_modifAude.html)¹. Here the parameter to estimate is $\theta = (\beta, \gamma_1^2, \gamma_2^2, \mu_1, \mu_2)$. In order to make the notations as simple as possible, we will confound the notations of the random variables and their observed realizations. Furthermore, we are going to write $f_\theta(z)$ the density of the random variable Z evaluated at z . For example $f_\theta(y_i|x_{i1})$ is the conditional density of y_i given x_{i1} .

With these notations, the complete likelihood $L_n(\theta)$ is given by :

¹https://juliejosse.com/wp-content/uploads/2018/07/LectureMissing-Weij_modifAude.html

$$\begin{aligned}
L_n(\theta) &= \prod_{i=1}^n f_\theta(y_i, x_i) \\
&= \prod_{i=1}^n f_\theta(y_i|x_i) f_\theta(x_i)
\end{aligned}$$

However we do not observe the first r $(x_{i2})_i$ that are handled as latent variables, therefore the observed likelihood is marginalized over their distribution :

$$\begin{aligned}
L_n(\theta) &= \left[\prod_{i=1}^r f_\theta(y_i, x_{i1}) \right] \times \prod_{i=r+1}^n f_\theta(y_i, x_i) \\
&= \left[\prod_{i=1}^r f_\theta(y_i|x_{i1}) f_\theta(x_{i1}) \right] \times \prod_{i=r+1}^n f_\theta(y_i|x_i) f_\theta(x_i) \\
&= \left[\prod_{i=1}^r \int_{x_{i2}} f_\theta(y_i|x_{i1}, x_{i2}) f_\theta(x_{i1}) f_\theta(x_{i2}) dx_{i2} \right] \times \prod_{i=r+1}^n f_\theta(y_i|x_i) f_\theta(x_i)
\end{aligned}$$

In fact we can compute exactly the marginal likelihood as the conditional distribution of y_i given x_{i1} is a $\mathcal{N}(\beta_1 x_{i1}, \beta_2^2 \gamma_2^2 + \sigma^2)$. Given a realization $\mathbf{x}_2 = (x_{i2})_{i=1, \dots, r}$ of the unobserved variables, we introduce the complete log likelihood defined as :

$$l_n(x_{12}, \dots, x_{r2}; \theta) = \sum_{i=1}^n \log(f_\theta(y_i, x_i))$$

We consider here the following test :

$$H_0 : \beta_0 = 0 \quad \text{against} \quad H_1 : \beta_0 \neq 0$$

To apply the joint procedure described in Section 4.2 we consider the unconstrained parameter space Θ_1 corresponding to the alternative hypothesis, and the constrained parameter Θ_0 that corresponds to the case $\beta_0 = 0$. For the two estimation procedures, we used stochastic gradient descent. At each step k the procedure computes an estimator g_k of the log likelihood ratio, jointly to the estimators $\hat{\theta}_{0,k}$ and $\hat{\theta}_{1,k}$ of respectively the restricted and unrestricted maximum likelihood estimators as follows :

1. Draw $\mathbf{x}_{02}^{(k+1)}$ from the posterior distribution of $(x_{i2})_{i=1, \dots, r}$ given $((y_i, x_i)_{i=1, \dots, r}, \hat{\theta}_{0k})$ and $\mathbf{x}_{12}^{(k+1)}$ from the posterior distribution of $(x_{i2})_{i=1, \dots, r}$ given $((y_i, x_i)_{i=1, \dots, r}, \hat{\theta}_{1k})$
2. Update $\hat{\theta}_{0,k+1}$ and $\hat{\theta}_{1,k+1}$ each with a gradient step :

$$\begin{aligned}
\hat{\theta}_{0,k+1} &= \hat{\theta}_{0,k} - \gamma_k \nabla_\theta l_n(\mathbf{x}_{02}^{(k+1)}; \hat{\theta}_{0,k}) \\
\hat{\theta}_{1,k+1} &= \hat{\theta}_{1,k} - \gamma_k \nabla_\theta l_n(\mathbf{x}_{12}^{(k+1)}; \hat{\theta}_{1,k})
\end{aligned}$$

3. Draw $\tilde{\mathbf{x}}_2^{(k+1)}$ from a uniform distribution on $\{\mathbf{x}_{02}^{(k+1)}, \mathbf{x}_{12}^{(k+1)}\}$
4. Update g_{k+1} as :

$$g_{k+1} = g_k + \gamma_k \frac{\exp \left\{ l_n(\tilde{\mathbf{x}}_2^{(k+1)}; \hat{\theta}_{0,k}) \right\} - \exp \left\{ l_n(\tilde{\mathbf{x}}_2^{(k+1)}; \hat{\theta}_{1,k}) + g_k \right\}}{\exp \left\{ l_n(\tilde{\mathbf{x}}_2^{(k+1)}; \hat{\theta}_{0,k}) \right\} + \exp \left\{ l_n(\tilde{\mathbf{x}}_2^{(k+1)}; \hat{\theta}_{1,k}) + g_k \right\}}$$

For the following experiments, we used a constant step size of $\gamma_k = 0.1$ for the SARIS procedure and the estimation processes to keep the speed of convergence of the different iterations as similar as possible. The parameters used to generate the data are $\mu = (1, 1)^T$, $(\gamma_1, \gamma_2) = (1, 1)$, $\sigma^2 = 2$ and $\beta = (0.1, 1, 1)$. Finally, we considered $n = 200$ individuals and two different numbers of missing values. We considered $r = 20$ which corresponds to 10% of missing values for the second covariate, and then $r = 50$ which corresponds to 25%. We used $K = 250$ iterations for the estimation process. We reproduced the experiment 20.

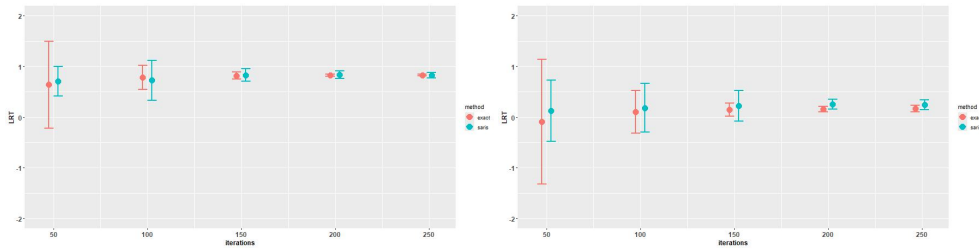


Fig. 3 Estimation of the log likelihood in the latent variable model of Section 5.2 in the case $r = 10\%n$ (on the left) and $r = 25\%n$ (on the right). Comparison over the iterations of the exact likelihood ratio $\left(-2 \log \frac{L_n(\hat{\theta}_{0,k})}{L_n(\hat{\theta}_{1,k})}\right)_{k \geq 0}$ (exact in red) and its approximation $(-2g_k)_{k \geq 0}$ (approx in blue) using the joint procedure described in Section 4.2. The dots represents the means taken over 20 repetitions, the errorbars correspond to the empirical standard deviation.

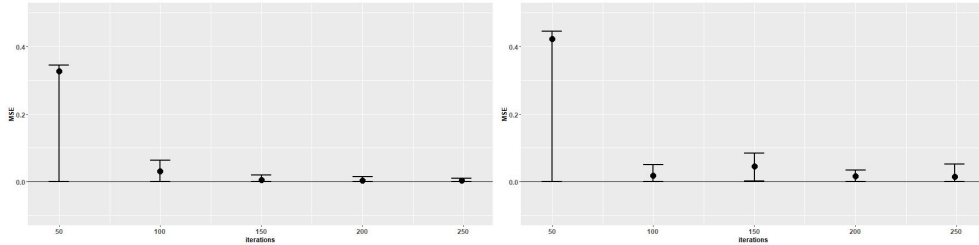


Fig. 4 Evolution of the empirical mean square error between the exact log likelihood ratio and its approximation using the joint procedure described in Section 4.2 in latent variable model of Section 5.2 in the case $r = 10\%n$ (on the left) and $r = 25\%n$ (on the right). The dots represents the means taken over 20 repetitions, the errorbars reach the 5% and 95% empirical quantiles.

Figure 3 displays the evolution of the exact likelihood ratio $-2 \log \frac{L_n(\hat{\theta}_{0,k})}{L_n(\hat{\theta}_{1,k})}$ over the iterations, compared to the one of the SARIS estimator obtained using the joint procedure described in Section 4.2. The dots represent the mean, and the errorbars

the standard deviations computed over the 20 repetitions of the experiment. Figure 4 plots the mean squared error between the exact log likelihood ratio along the iteration process and its approximation using the SARIS joint procedure. The dots are the mean over the 20 repetitions, and the errorbars reach the 5% and 95% corresponding empirical quantiles.

In the case where $r = 10\%n$ we observe that the joint procedure accurately tracks the exact value of the likelihood ratio along the estimation process which is very encouraging. We observe that in the more complex case where $r = 25\%n$ the joint procedure correctly tracks the exact likelihood ratio statistic but with a small bias. There might be several sources for this bias such as the difference in convergence speed of both estimation processes or the higher variance due to the higher number of missing data. However, the approximation still tracks the exact value along iterations, which makes the proposed joined procedure appealing as a first step of a marginal likelihood computation. Indeed it gives a first estimator at the end of the estimation process that might be used as a starting point for a SARIS procedure, or an estimator \hat{r} to use the approximated optimal scheme of [Chen and Shao \(1997b\)](#) to use as proposal distribution $\pi \propto |f_0 - \hat{r}f_1|$ in a ratio importance sampling procedure.

All the experiments were computed on R version 4.3.3 (2024-02-29 ucrt). For reproducibility the scripts are available at [this link](#)².

6 Conclusion

We proposed a new methodology to compute ratios of normalizing constants that relies on the principle of stochastic approximation. Our procedure presents good theoretical properties which makes it competitive with the best methods from the literature. More precisely, our estimator is consistent and asymptotically Gaussian as the number of iterations goes to infinity. Moreover, the practical implementation of the algorithm reaches an asymptotic variance which is smaller than the optimal variance of the Bridge sampling estimator. Another important advantage is that our estimator does not require to fix in advance the computational effort thanks to its iterative nature. Indeed our procedure can be stopped in practice once a given convergence criterion is reached. Furthermore, our estimator seems more robust to little overlap between the two unnormalised distributions considered and outperforms the Bridge sampling estimator in some of the numerical examples considered. The proposed methodology also allows for the computation of single marginal likelihoods. Moreover, in the context of likelihood ratio test statistics in latent variables models, our procedure can be integrated in the parameter estimation process to reduce the computational effort.

Besides these positive points, there are several interesting perspectives to investigate. Thanks to the rich literature on stochastic approximation and more specifically on stochastic gradient descent, many refinements can be explored, such as acceleration, variance reduction or adaptive step sizes. Similarly, refinements used in classical Monte Carlo such as the Warp Bridge sampling ([Meng and Schilling, 2002](#); [Wang et al.](#),

²<https://github.com/tguedon/saris>

2022) could be applied to reduce the asymptotic variance of the estimator. Furthermore, as a method to compute ratios of normalizing constants, the SARIS procedure can benefit from the use of intermediate distributions that enable to decompose the problem into several simpler sub-problems, in the principle of stepping stone sampling.

7 Declarations

7.1 Funding

This work was funded by the Stat4Plant project ANR-20-CE45-0012.

7.2 Conflict of interest

The authors declare that they have no conflict of interest.

References

- Annis, J., Evans, N.J., Miller, B.J., Palmeri, T.J.: Thermodynamic integration and steppingstone sampling methods for estimating bayes factors: A tutorial. *Journal of mathematical psychology* **89**, 67–86 (2019)
- Allasonniere, S., Kuhn, E.: Convergent stochastic expectation maximization algorithm with efficient sampling in high dimension. application to deformable template model estimation. *Computational Statistics & Data Analysis* **91**, 4–19 (2015)
- Baey, C., Delattre, M., Kuhn, E., Leger, J.-B., Lemler, S.: Efficient preconditioned stochastic gradient descent for estimation in latent variable models. *Proceedings of the 40th International Conference on Machine Learning* (2023)
- Bennett, C.H.: Efficient estimation of free energy differences from monte carlo data. *Journal of Computational Physics* **22**(2), 245–268 (1976)
- Brekelmans, R., Masrani, V., Bui, T., Wood, F., Galstyan, A., Steeg, G.V., Nielsen, F.: Annealed importance sampling with q-paths. *arXiv preprint arXiv:2012.07823* (2020)
- Benveniste, A., Métivier, M., Priouret, P.: *Adaptive Algorithms and Stochastic Approximations* vol. 22. Springer, & Business Media (2012)
- Chen, M.-H., Shao, Q.-M.: Estimating ratios of normalizing constants for densities with different dimensions. *Statistica Sinica*, 607–630 (1997)
- Chen, M.-H., Shao, Q.-M.: On monte carlo methods for estimating ratios of normalizing constants. *The Annals of Statistics* **25**(4), 1563–1594 (1997)
- Delyon, B., Lavielle, M., Moulines, E.: Convergence of a stochastic approximation version of the em algorithm. *Annals of statistics*, 94–128 (1999)

- Del Moral, P., Doucet, A., Jasra, A.: Sequential monte carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **68**(3), 411–436 (2006)
- Duflo, M.: *Algorithmes Stochastiques* vol. 23. Springer, (1996)
- Fort, G.: Central limit theorems for stochastic approximation with controlled markov chain dynamics. *ESAIM: Probability and Statistics* **19**, 60–80 (2015)
- Frühwirth-Schnatter, S.: Estimating marginal likelihoods for mixture and markov switching models using bridge sampling techniques. *The Econometrics Journal* **7**(1), 143–167 (2004)
- Friel, N., Wyse, J.: Estimating the evidence—a review. *Statistica Neerlandica* **66**(3), 288–308 (2012)
- Fan, Y., Wu, R., Chen, M.-H., Kuo, L., Lewis, P.O.: Choosing among partition models in bayesian phylogenetics. *Molecular biology and evolution* **28**(1), 523–532 (2011)
- Gelfand, A.E., Dey, D.K.: Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Methodological)* **56**(3), 501–514 (1994)
- Geyer, C.J.: Estimating normalizing constants and reweighting mixtures (1994)
- Gelman, A., Meng, X.-L.: Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, 163–185 (1998)
- Gronau, Q.F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D.S., Forster, J.J., Wagenmakers, E.-J., Steingroever, H.: A tutorial on bridge sampling. *Journal of mathematical psychology* **81**, 80–97 (2017)
- Gronau, Q.F., Singmann, H., Wagenmakers, E.-J.: bridgesampling: An r package for estimating normalizing constants. *Journal of Statistical Software* **92**(10), 1–29 (2020)
- Kuhn, E., Lavielle, M.: Coupling a stochastic approximation version of em with an mcmc procedure. *ESAIM: Probability and Statistics* **8**, 115–131 (2004)
- Llorente, F., Martino, L., Delgado, D., Lopez-Santiago, J.: Marginal likelihood computation for model selection and hypothesis testing: an extensive review. *SIAM Review* **65**(1), 3–58 (2023)
- Lartillot, N., Philippe, H.: Computing bayes factors using thermodynamic integration. *Systematic biology* **55**(2), 195–207 (2006)
- Meng, X.-L., Schilling, S.: Warp bridge sampling. *Journal of Computational and Graphical Statistics* **11**(3), 552–586 (2002)

- Meng, X.-L., Wong, W.H.: Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, 831–860 (1996)
- Neal, R.M.: Annealed importance sampling. *Statistics and computing* **11**, 125–139 (2001)
- Newton, M.A., Raftery, A.E.: Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **56**(1), 3–26 (1994)
- Robert, C.P., Casella, G.: *Monte Carlo Statistical Methods* vol. 2. Springer, New York, NY (1999)
- Robbins, H., Monro, S.: A stochastic approximation method. *The annals of mathematical statistics*, 400–407 (1951)
- Russel, P.M., Meyer, R., Veitch, J., Christensen, N.: The stepping-stone sampling algorithm for calculating the evidence of gravitational wave models. arXiv preprint arXiv:1810.04488 (2018)
- Roberts, G.O., Rosenthal, J.S.: Examples of adaptive mcmc. *Journal of computational and graphical statistics* **18**(2), 349–367 (2009)
- Shirts, M.R., Chodera, J.D.: Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of chemical physics* **129**(12) (2008)
- Torrie, G.M., Valleau, J.P.: Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**(2), 187–199 (1977)
- Vaart, A.W.: *Asymptotic Statistics* vol. 3. Cambridge university press, (2000)
- Wang, L., Jones, D.E., Meng, X.-L.: Warp bridge sampling: The next generation. *Journal of the American Statistical Association* **117**(538), 835–851 (2022)
- Xie, W., Lewis, P.O., Fan, Y., Kuo, L., Chen, M.-H.: Improving marginal likelihood estimation for bayesian phylogenetic model selection. *Systematic biology* **60**(2), 150–160 (2011)

Appendix A Proofs

A.1 Proof of Proposition 1

We first rewrite the iterative scheme (14) for every positive integer k :

$$r_{k+1} = r_k + \gamma_{k+1} H_{\pi_{r_k}}(Z_{k+1}), \quad Z_{k+1} \sim \pi_{r_k}$$

introducing the notation $H_{\pi_r}(z, r) = \frac{f_0(z) - rf_1(z)}{\pi_r(z)}$.

We apply the almost sure convergence theorem of Robbins-Monro algorithms (Robbins and Monro, 1951) stated in section 5.1 of Benveniste et al. (2012) to prove Proposition 1.

Under Assumption 1 we can define the function h for every $r \in \mathbb{R}$ by $h(r) = \mathbb{E}_r [H_{\pi_r}(Z, r)]$ where \mathbb{E}_r stands for \mathbb{E}_{π_r} for sake of simplicity. It follows directly that $h(r) = c_0 - rc_1$ for every $r \in \mathbb{R}$ and $h(r^*) = 0$. We define the filtration $(\mathcal{F}_k)_{k \geq 0}$ corresponding to the increasing family of σ -algebra generated by (r_0, Z_1, \dots, Z_k) . We now verify the following assumptions of the theorem stated in section 5.1 of Benveniste et al. (2012):

1. for any positive measurable function g defined on $\mathcal{Z} \times \mathbb{R}$ we have $\mathbb{E} [g(Z_{k+1}, r_k) | \mathcal{F}_k] = \mathbb{E}_{r_k} [g(Z, r_k)]$.
2. there exists $C > 0$, such that for every $r \in \mathbb{R}$ $\mathbb{E}_r [H_{\pi_r}(Z, r)^2] \leq C(1 + r^2)$
3. there exists $r^* > 0$ such that for every $r \in \mathbb{R} \setminus \{r^*\}$:

$$(r - r^*)h(r) < 0$$

4. $\sum \gamma_k = +\infty$ and $\sum \gamma_k^2 < +\infty$

The first point is straightforward considering the sampling scheme of equation (14). We now consider point 2):

$$\begin{aligned} \mathbb{E}_r [H_{\pi_r}(Z, r)^2] &= \int_{\mathcal{Z}} \left(\frac{f_0(z) - rf_1(z)}{\pi_r(z)} \right)^2 \pi_r(z) \mu(dz) \\ &= \int_{\mathcal{Z}} \frac{(f_0(z) - rf_1(z))^2}{\pi_r(z)} \mu(dz) \\ &= \int_{\mathcal{Z}} |H_{\pi_r}(z, r)| |f_0(z) - rf_1(z)| \mu(dz) \\ &\leq \int_{\mathcal{Z}} |H_{\pi_r}(z, r)| f_0(z) \mu(dz) + \int_{\mathcal{Z}} |H_{\pi_r}(z, r)| f_1(z) |r| \mu(dz) \\ &\leq c_0 \mathbb{E}_0 [|H_{\pi_r}(z, r)|] + |r| c_1 \mathbb{E}_1 [|H_{\pi_r}(z, r)|] \\ &\leq \tilde{C}(1 + r^2) \end{aligned}$$

thanks to Assumption 1.

Point 3) is straightforward as $h(r) = c_1(r^* - r)$. Finally point 4) is implied by Assumption 3. \square

A.2 Proof of Proposition 2

We apply the central theorem for Robbins-Monro algorithms (Duflo (1996) chapter 4) to prove Proposition 2. The additional assumptions to verify are:

1. There exists a neighborhood U of r^* such that the function h is continuously differentiable on U and for all $r \in U$, $h'(r) < 0$.

2. There exists $\Gamma > 0$, such that almost surely:

$$\lim_{k \rightarrow +\infty} \mathbb{E} \left[\left(H_{\pi_{r_k}}(Z_{k+1}, r_k) - h(r_k) \right)^2 \middle| \mathcal{F}_k \right] = \Gamma$$

3. There exists $\delta > 0$ such that

$$\sup_k \mathbb{E} \left[\left(H_{\pi_{r_k}}(Z_{k+1}, r_k) - h(r_k) \right)^{2+\delta} \middle| \mathcal{F}_k \right] < +\infty$$

4. There exist $\frac{1}{2} < \epsilon < 1$, $a > 0$, $b > 0$ such that the sequence of step sizes (γ_k) is of the form $\gamma_k = \frac{a}{b+k^\epsilon}$.

The first point is straightforward as h is linear in r and h' is constant equal to $-c_1 < 0$. To prove the second point, let us introduce $\xi_{k+1} = H_{\pi_{r_k}}(Z_{k+1}, r_k) - h(r_k)$ for all integer k . After some calculation, we get:

$$\mathbb{E}(\xi_{k+1}^2 | \mathcal{F}_k) = c_0 \mathbb{E}_0 \left[H_{\pi_{r_k}}(Z, r_k) \right] - r_k c_1 \mathbb{E}_1 \left[H_{\pi_{r_k}}(Z, r_k) \right] - h(r_k)^2$$

Under Assumptions 1 and 6, the sequence $(r_k)_k$ converges almost surely to r^* , $h(r^*) = 0$, and, applying the dominated convergence theorem, the following convergence holds almost surely:

$$\lim_{k \rightarrow +\infty} \mathbb{E}(\xi_{k+1}^2 | \mathcal{F}_k) = \mathbb{E}_{r^*} \left[H_{\pi_{r^*}}(Z, r^*)^2 \right]$$

The right hand side term is positive under Assumption 4. Finally we show that Assumption 5 implies condition 3) using similar calculation and arguments. Assumption 6 corresponds to condition 4).

Applying the central theorem for Robbins-Monro algorithms (Duflo (1996) chapter 4) we get that:

$$V_{saris}(\pi_{r^*}) = \mathbb{E}_{r^*} (H_{\pi_{r^*}}(Z, r^*)^2) / c_1^2$$

We then apply Jensen inequality to get the minorization:

$$\begin{aligned} V_{saris}(\pi_{r^*}) &\geq (\mathbb{E}_{r^*} [|H_{\pi_{r^*}}(Z, r^*)|])^2 / c_1^2 \\ &= \left(\int_{\mathcal{Z}} |f_0(z) - r^* f_1(z)| \mu(dz) \right)^2 / c_1^2 \\ &= r^{*2} \left(\int_{\mathcal{Z}} |p_0(z) - p_1(z)| \mu(dz) \right)^2 \end{aligned}$$

Moreover the equality case holds in Jensen inequality for $\pi_{saris}^{opt}(z) \propto |f_0(z) - r^* f_1(z)|$ which leads to the result. \square

A.3 Proof of Proposition 3

This proof follows the lines of the proof of Proposition 1. The only difference is that the objective function h to nullify is :

$$h : r \mapsto \frac{c_0 - rc_1}{c(r)}$$

that verifies for every $r \in \mathbb{R}$, $(r - r^*)h(r) = \frac{-c_1(r - r^*)^2}{c(r)} < 0$. The remaining hypothesis are verified thanks to Assumption 7.

Remark 13. *The fact that $c(r)$ is strictly positive for all r is implied by Assumption 7. $c(r) = 0$ would mean that $\tilde{\pi}_r(Z) = 0$ μ -almost surely.* □

A.4 Proof of Proposition 4

In order to prove this proposition we proceed in two steps. The first step establishes the central limit theorem, the second step shows that the use of the unnormalized density $\tilde{\pi}_r^{opt}(z) = |f_0(z) - rf_1(z)|$ enables to reach the optimal asymptotic variance V_{saris}^{opt} .

The first step follows the same lines as the proof of Proposition 2, replacing Assumptions 2 and 5 by Assumptions 7 and 8. Assumption 9 ensures the differentiability of h in a neighborhood U of r^* , and the fact that $h'(r^*) \neq 0$. Furthermore, $h'(r^*) = \frac{-c_0c(r^*) - c'(r^*)(c_0 - r^*c_1)}{c(r^*)^2} = \frac{-c_1}{c(r^*)}$. The continuity of c and $r \mapsto \tilde{\pi}_r(z)$ for every z are guaranteed by Assumption 1 and the differentiability of h . The central limit theorem of [Duflo \(1996\)](#) chapter 4 concludes this step, and :

$$\begin{aligned} V_{ext}(\tilde{\pi}_{r^*}) &= \frac{c(r^*)^2}{c_1^2} \mathbf{E}_{r^*} \left[\left(\frac{f_0(Z) - r^*f_1(Z)}{\tilde{\pi}_{r^*}(Z)} \right)^2 \right] \\ &= \frac{1}{c_1^2} \mathbf{E}_{r^*} \left[\left(\frac{f_0(Z) - r^*f_1(Z)}{\pi_{r^*}(Z)} \right)^2 \right] \\ &= V_{saris}(\pi_{r^*}) \end{aligned}$$

Let $r \in \mathbb{R}$, $\tilde{\pi}_r^{opt} : z \mapsto |f_0(z) - rf_1(z)|$ and $c^{opt} : r \mapsto \int_{\mathcal{Z}} \tilde{\pi}_r^{opt}(z) dz$. The second step proves that $\tilde{\pi}_r^{opt}$ verifies Assumptions 7, 8 and 9.

First for every $r \in \mathbb{R}$ and $z \in \mathcal{Z}$, the quantity $\frac{f_0(z) - rf_1(z)}{\tilde{\pi}_r^{opt}(z)}$ is bounded since its absolute value equals 1, and therefore Assumptions 7 and 8 are verified. The main point to verify is Assumption 9. This is not straightforward because of the absolute value in the definition of the c^{opt} function. Let $r \in \mathbb{R}$, $\mathcal{A}_r^+ = \{z \in \mathcal{Z} : f_0(z) > rf_1(z)\}$ and $\mathcal{A}_r^- = \{z \in \mathcal{Z} : f_0(z) < rf_1(z)\}$. We get

$$\begin{aligned} c(r) &= \int_{\mathcal{Z}} |f_0(z) - rf_1(z)| \mu(dz) \\ &= - \int_{\mathcal{A}_r^-} (f_0(z) - rf_1(z)) \mu(dz) + \int_{\mathcal{A}_r^+} (f_0(z) - rf_1(z)) \mu(dz) \\ &= -c_0 + rc_1 + 2 \int_{\mathcal{A}_r^+} (f_0(z) - rf_1(z)) \mu(dz) \end{aligned}$$

using Assumption 10, and that for $i = 0, 1$, $c_i = \int_{\mathcal{A}_r^+} f_i(z)\mu(dz) + \int_{\mathcal{A}_r^-} f_i(z)\mu(dz)$. To study the differentiability of $m : r \mapsto \int_{\mathcal{A}_r^+} f_0(z) - rf_1(z)\mu(dz)$, we consider $r_0 \in \mathbb{R}$ and $\epsilon > 0$:

$$m(r_0 + \epsilon) - m(r_0) = \int_{\mathcal{A}_r^+ \setminus \mathcal{A}_{r+\epsilon}^+} (rf_1(z) - f_0(z))\mu(dz) - \epsilon \int_{\mathcal{A}_{r+\epsilon}^+} f_1(z)\mu(dz)$$

Let study the first term.

$$\begin{aligned} \mathcal{A}_r^+ \setminus \mathcal{A}_{r+\epsilon}^+ &= \{z \in \mathcal{Z} : rf_1(z) < f_0(z) \leq (r + \epsilon)f_1(z)\} \\ &= \{z \in \mathcal{Z} : 0 < f_0(z) - rf_1(z) \leq \epsilon f_1(z)\} \end{aligned}$$

f_1 is μ -almost surely upper-bounded as it is a (unnormalized) density, let \bar{f}_1 its upperbound.

$$\begin{aligned} \left| \int_{\mathcal{A}_r^+ \setminus \mathcal{A}_{r+\epsilon}^+} rf_1(z) - f_0(z)\mu(dz) \right| &\leq \int_{\mathcal{A}_r^+ \setminus \mathcal{A}_{r+\epsilon}^+} |rf_1(z) - f_0(z)|\mu(dz) \\ &\leq \epsilon \bar{f}_1 \mu(\mathcal{A}_r^+ \setminus \mathcal{A}_{r+\epsilon}^+) \end{aligned}$$

and $\mu(\mathcal{A}_r^+ \setminus \mathcal{A}_{r+\epsilon}^+) \xrightarrow{\epsilon \rightarrow 0} 0$ with Assumption 10. The same calculations hold with $\epsilon < 0$ and we get that m is differentiable on \mathbb{R} and for every $r \in \mathbb{R}$ $m'(r) = - \int_{\mathcal{A}_r^+} f_1(z)\mu(dz)$ and finally c is differentiable on \mathbb{R} and specifically :

$$c'(r^*) = c_1 - 2 \int_{\mathcal{A}_{r^*}^+} f_1(z)\mu(dz)$$

therefore h is differentiable on \mathbb{R} which concludes the proof, as the first part of the proposition shows that $V_{ext}(\tilde{\pi}_{r^*}) = V_{saris}(\pi_{r^*})$.

A.5 Proof of Proposition 5

We first derive the expression of $V_{ext}(\tilde{\pi}_{r^*}^{mixt})$.

We recall the notations: $\pi_{r^*}^{mixt}(z) = \frac{1}{2}(p_0(z) + p_1(z)) \propto f_0(z) + r^* f_1(z) = \tilde{\pi}_{r^*}^{mixt}(z)$.

We start from the result of Proposition 2:

$$\begin{aligned} V_{ext}(\tilde{\pi}_{r^*}^{mixt}) &= \int_{\mathcal{Z}} \frac{(f_0(z) - r^* f_1(z))^2}{\pi_{r^*}^{mixt}(z)} \mu(dz) / c_1^2 \\ &= 2r^{*2} \int_{\mathcal{Z}} \frac{(p_0(z) - p_1(z))^2}{p_0(z) + p_1(z)} \mu(dz) \\ &= 2r^{*2} \left(\int_{\mathcal{Z}} \frac{(p_0(z) + p_1(z))^2 - 4p_0(z)p_1(z)}{p_0(z) + p_1(z)} \mu(dz) \right) \end{aligned}$$

$$\begin{aligned}
&= 2r^{*2} \left(2 - \int_{\mathcal{Z}} \frac{4p_0(z)p_1(z)}{p_0(z) + p_1(z)} \mu(dz) \right) \\
&= 4r^{*2} (1 - \Psi)
\end{aligned}$$

using $(a - b)^2 = (a + b)^2 - 4ab$ for all a and b reals. The result of Proposition 4 also holds as $h(r) = \frac{c_0 - rc_1}{c_0 + rc_1}$ verifies Assumption 9.

We then present the main steps to prove the almost sure convergence and the asymptotic normality of the sequence obtained using Algorithm 3. The analysis of this procedure follows the same lines as the one of the SARIS and SARIS-ext procedure. We introduce the notation

$$\tilde{H}(z, r) = \frac{f_0(z) - rf_1(z)}{f_0(z) + rf_1(z)}$$

and we define under Assumption 1, with $\pi = \pi^{mixt}$, the function

$$\tilde{h}(r) = \mathbb{E}_{\pi_{r^*}^{mixt}} \left(\frac{f_0(Z) - rf_1(Z)}{f_0(Z) + rf_1(Z)} \right).$$

After some calculation, we get

$$\tilde{h}'(r^*) = -\frac{\Psi}{2r^*}$$

and

$$\tilde{\Gamma} = \mathbb{E}_{\pi_{r^*}^{mixt}} \left[\left(\frac{f_0(Z) - r^* f_1(Z)}{f_0(Z) + r^* f_1(Z)} \right)^2 \right] = (1 - \Psi),$$

which is positive under Assumption 4. Following the same lines as in the proofs of Propositions 1 and 2, it is straightforward to check that all the assumptions required for the consistency and asymptotic normality hold. Applying these results to the sequence generated by Algorithm 3 leads to the following asymptotic variance V_{saris}^{mixt} :

$$V_{saris}^{mixt} = 4r^{*2}(1 - \Psi)/\Psi^2$$

The optimal bridge sampling asymptotic variance is given in equation (4) by $V_{bridge}^{opt} = 4r^{*2}(\Psi^{-1} - 1)$ and $V_{ris}^{mixt} = 4r^{*2}(1 - \Psi)/\Psi^2$ (see for example Chen and Shao (1997a, Theorem 5.2.)), which concludes the proof. \square