Simulation from a target distribution based on discretization and weighting

Sonia Malefaki and George Iliopoulos[*]

Abstract

We present a simulation method which is based on discretization of the state space of the target distribution (or some of its components) followed by proper weighting of the simulated output. The method can be used in order to simplify certain Monte Carlo and Markov chain Monte Carlo algorithms. Its main advantage is that the autocorrelations of the weighted output almost vanish and therefore standard methods for iid samples can be used for estimating the Monte Carlo standard errors. We illustrate the method via toy examples as well as the well-known dugongs and Challenger datasets.

Key words and phrases: discretization; properly weighted samples; Markov chain Monte Carlo; autocorrelations.

1 Introduction

Let π be a (possibly multidimensional) distribution with support $\mathcal{X}$. Suppose that we wish to generate random variates from π and/or to calculate expectations of the form

$$\mathbf{E}_\pi(h) = \int_{\mathcal{X}} h(x)\pi(\mathrm{d}x)$$

for $h \in \mathcal{L}^1(\pi)$, that is, satisfying $\mathbf{E}_\pi(|h|) < \infty$. As the dimension of $\mathcal{X}$ increases, handling such problems becomes more difficult. For this purpose, many Monte Carlo (MC) and Markov chain Monte Carlo (MCMC) methods have been developed. A well-documented presentation of several MC and MCMC methods can be found in Robert and Casella (2004).

In this paper we discuss a simple approach which can be adopted in order to easily handle problems as calculating expectations as well as simulating approximate samples from a target distribution. Similarly to other standard simulation methods, it requires only the knowledge of the functional form of the target density (i.e., to be known up to a normalizing constant). This is usually the case in many situations as for example in Bayesian modeling with nonconjugate prior distributions. The approach is based on discretization of the state space $\mathcal{X}$ (or some of its components) combined with the proper weighting of the output.

---

[*]Corresponding author; Address: Department of Statistics and Insurance Science, University of Piraeus, 80 Karaoli & Dimitriou str., 18534 Piraeus, Greece, e–mail: `geh@unipi.gr`

Roughly speaking, "discretization" of a continuous random variable means partition of its support into a finite number of subsets followed by the introduction of another continuous random variable having constant probability density function (pdf) over each element of the partition. Then, under some appropriate conditions, the simulation of the original random variable may be replaced by simulation of its "discretized" version which is in general a much easier procedure. A discretization can be random or deterministic. For instance, Fu and Wang (2002) proposed a random discretization of a bounded support. More specifically, they discretized a bounded set of the form $(\alpha, \beta)^k$ into m contours by using nmk independent draws from the uniform distribution on (α, β). Every k successive observations were combined and were treated as an observation in $(\alpha, \beta)^k$. So, they produced nm independent identically and uniformly distributed observations on $(\alpha, \beta)^k$. They ordered the observations according to the height of the target density so that each contour contains n of the ordered observations and then simulated from the discretized distribution. Sainudiin and York (2006) proposed a deterministic discretization of the state space $\mathcal{X}$ in order to improve efficiency in Moore rejection sampler. Moreover, they proposed an adaptive partition of $\mathcal{X}$ by bisecting a properly chosen contour along the side with the maximal diameter, for further increasing the acceptance probability. There are several ways of choosing the proper contour, such as the contour with the largest volume or the contour with the largest diameter for its range enclosure. Liang, Liu and Carroll (2007) partitioned the sample space $\mathcal{X}$ into m contours according to the energy function $U(\mathbf{x}) = -\log \pi(\mathbf{x})$ in order to form the subsets $E_1 = \{\mathbf{x} : U(\mathbf{x}) \leqslant u_1\}, \ldots, E_i = \{\mathbf{x} : u_{i-1} < U(\mathbf{x}) \leqslant u_i\}, \ldots, E_m = \{\mathbf{x} : U(\mathbf{x}) > u_{m-1}\}$, where $u_1, \ldots, u_{m-1}$ are prespecified real numbers, and then used the Stochastic Approximation Monte Carlo Importance Sampling algorithm in order to sample from the target distribution π. Neil, Tailor and Marquez (2007) proposed a dynamic discretization of the state space $\mathcal{X}$ in order to approximate the true pdf π by the resulting discretized density g to an acceptable level of accuracy. They constructed a sequence of discretizations of $\mathcal{X}$ iteratively and at each step they tested if g has converged to π using a bound on their Kullback–Leibler divergence, called the entropy error. If the goal is not achieved, the procedure continues by splitting into two halves the highest entropy error interval and merging the consecutive intervals having the lowest entropy error or the ones that have zero mass.

In this paper, we propose a discretization in a deterministic way. First, we partition the support (or some of its components) into a predefined finite number of subsets. Then we simulate observations from the discretized version of the target distribution and we additionally weight them properly by assigning certain importance weights. Finally, we associate a jump process with the weighted sequence. Malefaki and Iliopoulos (2008) have already proven that this process converges weakly to the target distribution. We use the method in order to simplify certain MCMC algorithms. Interestingly, this approach seems to reduce considerably the

autocorrelations within the Markov sequence. This allows us to obtain quite accurate estimates of the Monte Carlo standard errors.

The rest of the paper is organized as follows. In Section 2, we define the notion of a simple distribution and we discuss a two-state procedure in order to sample from it. In Section 3, we give the definition of the jump process associated with properly weighted sequences, recall some already known results and describe our method explicitly. In Section 4, we illustrate the method using some toy examples as well as the well–known dugongs and Challenger datasets. We conclude the paper with a short summary. Finally, an appendix contains some technical details concerning the finiteness of the weights' variances in dugongs and Challenger datasets.

2   Discretization of the state space

Let π be a continuous distribution with support $\mathcal{X}$ which is known up to a normalizing constant. The first step of our approach is the replacement of π by a discretized version of it, that is, a "similar" simple distribution.

Definition 2.1. [SIMPLE DISTRIBUTION] Let $\mathcal{E} = (E_1, \ldots, E_m)$ be a measurable partition of the set $\mathcal{X}$. Any continuous distribution with constant pdf over each E_i will be called simple with respect to the partition $\mathcal{E}$.

In light of the above definition, let $\mathcal{E} = (E_1, \ldots, E_m)$ be a (Lebesgue) measurable partition of the state space $\mathcal{X}$ of π. To avoid trivialities, assume that $L(E_i) > 0$ for all i, where $L(E)$ denotes the Lebesgue measure of the set E. Choose points $\{x_1^*, \ldots, x_m^*\}$ such that for all $i = 1, \ldots, m$, $x_i^* \in \bar{E}_i$ (the boundary of E_i) with $\pi(x_i^*) > 0$. Then, for our target distribution π, the simple distribution

$$g(x) \propto \sum_{i=1}^{m} \frac{\pi(x_i^*)}{L(E_i)} I(x \in E_i), \tag{1}$$

is a discretization of π corresponding to the set $\{(E_i, x_i^*)\}_{i=1}^{m}$. For instance, if $\mathcal{X} = (\alpha, \beta)$ for some $-\infty < \alpha < \beta < \infty$ (i.e., a bounded subset of $\mathbb{R}$), a quite convenient discretization comes out by taking $E_i = \left(\alpha + \frac{i-1}{m}(\beta - \alpha), \alpha + \frac{i}{m}(\beta - \alpha)\right]$ and $x_i^* = \alpha + \frac{i-1/2}{m}(\beta - \alpha)$, that is, the midpoint of E_i, $i = 1, \ldots, m$. Then, the corresponding discretization of π becomes

$$g(x) \propto \sum_{i=1}^{m} \pi(x_i^*) I(x \in E_i), \tag{2}$$

since $L(E_i) \equiv (\beta - \alpha)/m$ for all i.

Note that Kozlov and Koller (1997) prove that the optimal —in terms of the Kullback–Leibler divergence— value of the simple distribution g in the subregion E_i is $\int_{E_i} \pi(x) \mathrm{d}x / L(E_i)$. However, in many cases this integral is quite difficult or even infeasible to be calculated exactly. For instance, this is the case when π is known up to a normalizing constant. So, in practice it

3

is more convenient to construct the simple distribution using suitable representative points x_i^*, such as the midpoints of the intervals or some local extrema of π.

In general, simulation from g in (1) can be carried out via the following simple two-stage procedure. First, the set E_i is chosen with probability proportional to $\pi(x_i^*)/L(E_i)$ and then x is drawn uniformly in E_i. Fu and Wang (2002) use this two-stage procedure in order to obtain a sample from g and estimate an expectation of interest with respect to π. More specifically, let $x_1, x_2, \ldots, x_n$ be an iid sequence from g. They prove that for any function $h \in \mathcal{L}^1(\pi)$ it holds

$$\lim_{n,m\uparrow\infty} \frac{1}{n} \sum_{i=1}^{n} h(x_i) \xrightarrow{p} \mathbf{E}_\pi(h) = \int_{\mathcal{X}} h(x)\pi(x)\mathrm{d}x. \tag{3}$$

As m increases, π is approximated better by g and the resulting estimators become more accurate. Fu and Wang (2002) also state that a value of m between 200 and 500 for a density of less than five dimensions and a value between 1000 and 100000 for a density of higher dimensions provides satisfactory results. On the contrary, our method needs significantly less number of subsets. Moreover, since the sample which is drawn from g is properly weighted with respect to π as it is presented in the next section, it converges in a sense to the target distribution π. This is actually true for the jump process associated with the weighted sample. Consequently, this method gives satisfactory results even if m is fairly small.

3 Jump processes associated with weighted sequences

The main principle of our approach is that of properly weighted samples. At first, properly weighted samples were introduced by Liu and Chen (1998) as a generalization of the standard importance sampling (IS) method. An equivalent and more convenient definition which is given also by Liu (2001), says that a set of weighted random samples $(X_i, \xi_i)_{1\leqslant i\leqslant n}$ is called proper with respect to π if

$$\mathbf{E}\{\xi_i | X_i = x\} = \kappa\pi(x)/g(x), \quad \text{for } i = 1, \ldots, n,$$

for some positive constant κ, where $X_i \sim g$. Malefaki and Iliopoulos (2008) defined the jump process which is associated with a weighted random sequence as follows:

Definition 3.1. For a weighted sequence $(X_n, \xi_n)_{n\in\mathbf{Z}_+} := ((X_0, \xi_0), (X_1, \xi_1), \ldots)$, where the ξ's are strictly positive weights, define $S_0 = 0$, $S_n = \sum_{i=0}^{n-1} \xi_i$, $n \geqslant 1$, and let

$$N_t := \sup\{n : S_n \leqslant t\}, \quad t \geqslant 0.$$

Then, the stochastic process $Y = (Y_t)_{t\geqslant 0}$ defined by $Y_t := X_{N_t}$, $t \geqslant 0$, is called the jump process associated with the weighted sequence $(X_n, \xi_n)_{n\in\mathbf{Z}_+}$.

Assume now that the sequence $X = (X_n)_{n\in\mathbf{Z}_+}$ is a homogeneous Harris ergodic Markov chain with state space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ having invariant probability distribution g and the distribution of

4

ξ_n depends solely on X_n with $\mathbf{E}\{\xi_n|X_n = x\} = \kappa w(x) = \kappa\pi(x)/g(x)$ for some $\kappa > 0$. Then, for the jump process $(Y_t)_{t \geqslant 0}$ associated with the weighted sequence $(X_n, \xi_n)_{n \in \mathbf{Z}_+}$ it holds that

$$\lim_{t \uparrow \infty} \mathbf{P}\{Y_t \in A\} = \pi(A), \quad \forall\, A \in \mathcal{B}(\mathcal{X}).$$

This fact follows from the standard theory of semi–Markov processes (cf. Limnios and Oprişan, 2001), since under the above assumptions, Y is a semi–Markov process with limit distribution π.

In the sequel we will concentrate on the special case $\xi_n \equiv w(X_n)$, that is, when the output of the Markov chain is deterministically weighted by the importance ratios. As it is shown in Malefaki and Iliopoulos (2008), in cases where the function $w(x)$ is known (up to a constant) there is no reason to consider random weights at all since the deterministic weights minimize the variances of the Monte Carlo estimators.

Working along the lines of the above approach, we are allowed to run more convenient MC or MCMC algorithms with a different target distribution g and then properly weight the outputs. For instance, in the standard Gibbs sampling setup, g could be a modified target distribution of which all full conditional distributions are easily handled. In any case, after properly weighted, the generated sample can be associated with the realization of a converging jump process.

Summarizing, our method consists of the following steps:

1. Discretize the state space (or some of its components) deterministically into a predefined finite number of subsets. In case the components to be discretized have unbounded support, apply suitable transformations to them in order to achieve this requirement (see for example Subsection 4.4).

2. Simulate an ergodic Markov chain with limit distribution the discretized version g of π.

3. After finishing the simulations, weight the simulated output $x_1, \ldots, x_n$ using the corresponding importance weights, that is, weight x_i by $w(x_i) \propto \pi(x_i)/g(x_i)$, $i = 1, \ldots, n$.

4. Treat the weighted sample $(x_1, w(x_1)), \ldots, (x_n, w(x_n))$ as a sample from the original target distribution π (possibly after discarding an initial part as a burn-in period).

The validity of Step 4 follows from the fact that the jump process $(y_t)_{t \geqslant 0}$ associated with the weighted sample $(x_i, w(x_i))_{1 \leqslant i \leqslant n}$ converges weakly to π. For example, setting $T = \sum_{i=1}^n w(x_i)$, the expectation $\mathbf{E}_\pi(h)$ can be estimated by

$$\frac{1}{T}\int_0^T h(y_t)\mathrm{d}t \equiv \frac{\sum_{i=1}^n w(x_i)h(x_i)}{\sum_{i=1}^n w(x_i)}.$$

Moreover, with $x_{(1)} \leqslant \cdots \leqslant x_{(n)}$ denoting the ordered observations, the upper α quantile of π can be estimated by $x_{(k)}$ where k satisfies $\sum_{i=k+1}^n w(x_{(i)})/T < \alpha \leqslant \sum_{i=k}^n w(x_{(i)})/T$. Note also

that Step 2 covers the case of iid sampling from g as well, since any iid sequence forms trivially a Markov chain. In fact, in this case the method is the original importance sampling with all its pros and cons; the only difference is that, by considering the jump process associated with the importance weighted sample, importance sampling is broadened to a more general simulation algorithm (see also Malefaki and Iliopoulos, 2008).

In the sequel, we will refer to the original simulated output $x_1, \ldots, x_n$ as unweighted sample and to the same values after weighted by their importance weights as weighted sample. Once more, we stress out that the target distribution of the unweighted sample is the simple distribution g whilst the target distribution of the jump process associated with the weighted sample (and in this sense of the weighted sample itself) is π.

In the next section we illustrate our approach by first presenting a couple of toy examples and then applying it to two real datasets which serve often as benchmark in Computational Bayesian Statistics.

4   Illustrative examples

4.1   Beta distribution

Consider the $\mathcal{B}eta(2,2)$ pdf,
$$\pi(x) \propto x(1-x), \quad 0 \leqslant x \leqslant 1.$$

We discretize the interval $[0,1]$ into $m = 10$ equal length bins and choose as target distribution the discretized version of π

$$g(x) \propto x_{[m]}(1 - x_{[m]}), \quad 0 \leqslant x \leqslant 1,$$

where $x_{[m]} = (2 \lfloor mx \rfloor + 1)/2m$. In order to sample from g, we first draw a bin from the discrete distribution
$$p(i) \propto \left(\frac{2i-1}{2m} \right) \left(1 - \frac{2i-1}{2m} \right), \quad i = 1, \ldots, m,$$

and then simulate $u \sim \mathcal{U}(0,1)$ and set $x = (u + i - 1)/m$. Let $w(x) = \pi(x)/g(x)$. The weighted sample $(x_n, w(x_n))_{n \geqslant 1}$ is proper with respect to π, so the jump process associated with the weighted output converges to π. At this point we should note that, in this particular example, the proposed method does not require a large m in order to achieve convergence. In Figure 1 we can notice the distributional convergence to the target distribution (the histogram of the weighted sample fits almost perfectly to its pdf) as well as the fast convergence of the weighted mean $\hat{x}_n^{IS} = \sum w(x_i)x_i / \sum w(x_i)$ to the mean of the distribution. For comparison purposes we have also simulated a random sample from the target distribution. In Figure 1 we can see the convergence of the corresponding sequence of the sample means as well. Observe that $\hat{x}_n^{IS}$ does not behave worse than the mean of a random sample from the target distribution. Moreover,
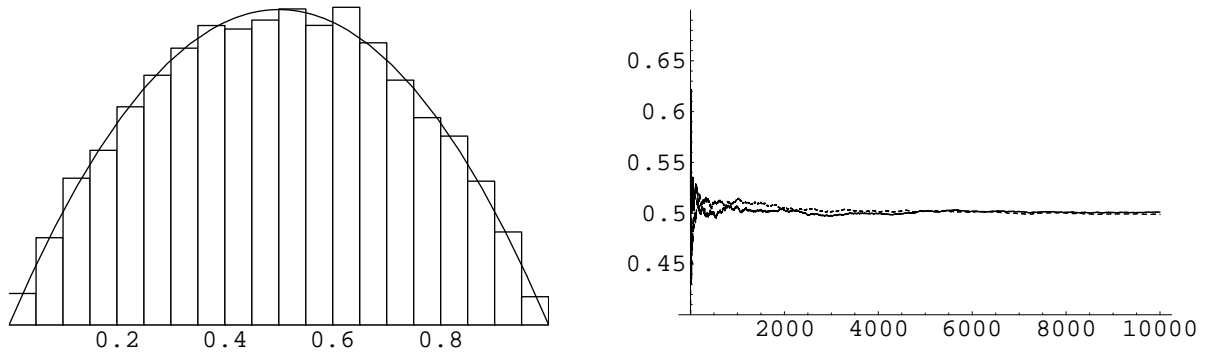
Figure 1: Histogram of the weighted sample $(x_i, w(x_i))_{1 \leqslant i \leqslant 10000}$ and convergence of the weighted mean $\hat{x}_n^{IS} = \sum w(x_i) x_i / \sum w(x_i)$ for $m = 10$ (solid line) and of the sample mean of an iid sample from the target distribution (dashed line).

Percentile	.05	.10	.25	.50	.75	.90	.95
True value	.1354	.1958	.3264	.5000	.6736	.8042	.8646
Estimated	.1317	.1932	.3283	.5022	.6786	.8031	.8641

Table 1: True and estimated percentiles of $\mathcal{B}eta(2,2)$ distribution.

in Table 1 we present the estimates of selected percentiles of the target distribution along with their corresponding true values. We can observe that the estimated values using the proposed method are very close to the true ones.

4.2 Dirichlet distribution

Let the target distribution be the two dimensional Dirichlet distribution $\mathcal{D}(\alpha, \beta; \gamma)$ with pdf

$$\pi(x) \equiv \pi(x_1, x_2) \propto x_1^{\alpha-1} x_2^{\beta-1} (1 - x_1 - x_2)^{\gamma-1}, \quad 0 \leqslant x_1, x_2, x_1 + x_2 \leqslant 1.$$

Here, the state space is the orthogonal triangle with vertices the origin and the points $(0, 1)$ and $(1, 0)$.

Two particular discretizations of the state space could be the following. Split first the original orthogonal triangle into $M = m^2$ smaller equal orthogonal triangles with area $1/(2m^2)$ as follows. Divide each of the axes into m equal length intervals and draw from their endpoints all lines parallel to the axes. Finally, draw the lines connecting the points $(i/m, 0)$ and $(0, i/m)$, $i = 1, \ldots, m - 1$. For this partition of the state space we choose as representative points

$$x_{[m1],i} = \frac{2i - 1}{2m}, \quad x_{[m2],jk} = \frac{j + 3(k - 1)}{3m},$$

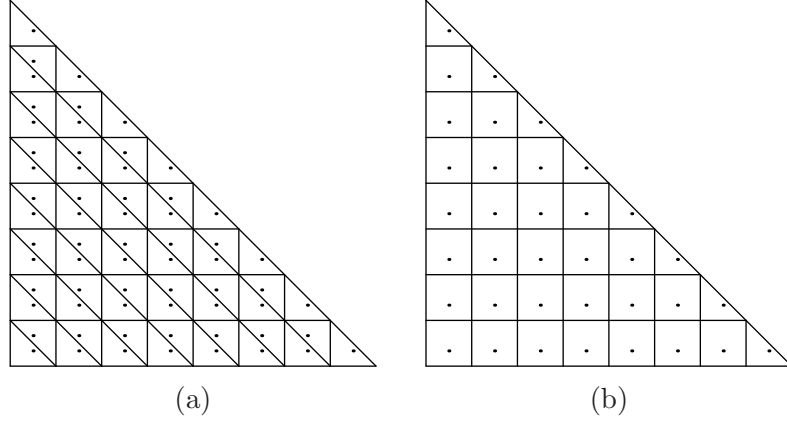for $i = 1, \ldots, m$, $j = 1, 2$ and $k = 1, \ldots, m - i - j + 2$ (see Figure 2(a)). Then, the corresponding

7

Figure 2: The two alternative discretizations of the state space of Dirichlet distribution.

simple distribution takes the form

$$g(x) \propto \sum_{i,j,k} x_{[m1],i}^{\alpha-1} x_{[m2],jk}^{\beta-1} (1 - x_{[m1],i} - x_{[m2],jk})^{\gamma-1} I(x \in E_{ijk}),$$

where E_{ijk} is the triangle containing the point $(x_{[m1],i} x_{[m2],jk})$. In order to simulate from g, a triangle is selected with probability proportional to

$$p(i,j,k) \propto x_{[m1],i}^{\alpha-1} x_{[m2],jk}^{\beta-1} (1 - x_{[m1],i} - x_{[m2],jk})^{\gamma-1}$$

and then (x_1, x_2) is drawn uniformly on E_{ijk}.

Alternatively, we can divide the original triangle into m triangles as before and $m(m+1)/2$ squares having double area of the triangles. In this case, we select as representative points

$$x_{[m1],i} = \frac{2i-1}{2m}, \quad x_{[m2],j} = \frac{3(j-1)+1}{3m},$$

for $i = 1, \ldots, m$, $j = 1, \ldots, m - i + 1$ (see Figure 2(b)). Since now the areas of triangles and squares differ, the corresponding simple distribution g takes the form

$$g(x) \propto \sum_{i,j} L(E_{ij})^{-1} x_{[m1],i}^{\alpha-1} x_{[m2],j}^{\beta-1} (1 - x_{[m1],i} - x_{[m2],j})^{\gamma-1} I(x \in E_{ij})$$

where $L(E_{ij}) = \{I(i+j \leqslant m) + I(i+j = m+1)/2\} m^{-2}$ is the area of E_{ij}.

The results obtained using both partitioning methods were similar. In Figure 3, the corresponding histograms of the weighted samples $(x_{1,i}, w(x_{1,i}))_{1 \leqslant i \leqslant n}$ and $(x_{2,i}, w(x_{2,i}))_{1 \leqslant i \leqslant n}$ clearly illustrate the distributional convergence of the jump process to the target distribution of (X_1, X_2), where $(X_1, X_2) \sim \mathcal{D}(2,3,4)$, $n = 10000$, $m = 15$. Furthermore, the right graphs show the fast convergence of the weighted means $\hat{x}_{1n}^{IS} = \sum w(x_{1,i}) x_{1,i} / \sum w(x_{1,i})$ and $\hat{x}_{2n}^{IS} = \sum w(x_{2,i}) x_{2,i} / \sum w(x_{2,i})$. Similarly to the previous example, we have included in Figure 3 the sequences of the
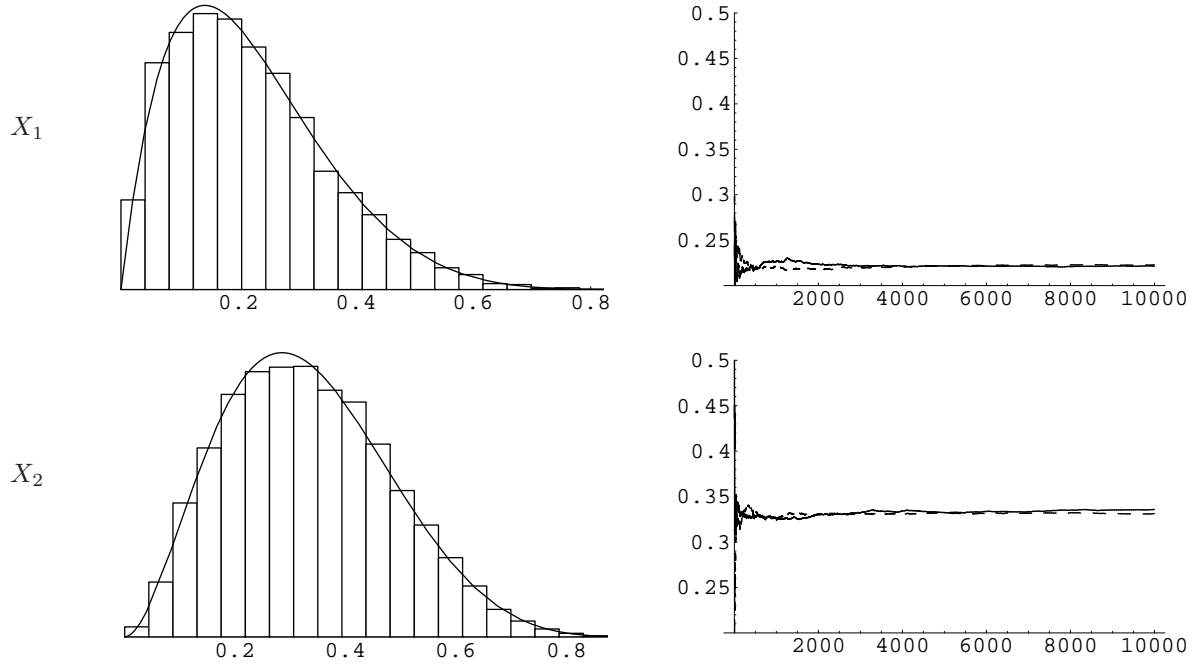
Figure 3: Histogram of the weighted sample $(x_{1,i}, w(x_{1,i}))_{1 \leqslant i \leqslant 10000}$ and $(x_{2,i}, w(x_{2,i}))_{1 \leqslant i \leqslant 10000}$ with $m = 15$ and the convergence of $\hat{x}_{1n}^{IS} = \sum w(x_{1,i}) x_{1,i} / \sum w(x_{1,i})$ and $\hat{x}_{2n}^{IS} = \sum w(x_{2,i}) x_{2,i} / \sum w(x_{2,i})$ and of the sample means of an iid sample from the target distributions (dashed line).

sample means obtained by an iid sample from the target distribution. Upon inspection, we may conclude that the rates of convergence of the iid-case sample mean and the weighted mean are comparable.

We can work in a similar way in higher dimensions. For example, consider the state space of the three dimensional Dirichlet distribution $(X_1, X_2, X_3) \sim \mathcal{D}(\alpha_1, \alpha_2, \alpha_3; \alpha_4)$ which is a pyramid with vertices the origin of the axes and the points $(1, 0, 0), (0, 1, 0), (0, 0, 1)$. Working among the lines of the above approach, the state space can be partioned in $m(m + 1)/2$ pyramids, $m(m - 1)/2$ polyhedra which are cubes without the grooved pyramid and $m(m - 1)(m - 2)/6$ cubes. Choosing appropriate points in the interior of these three-dimensional shapes we can easily simulate from the corresponding discretization of the target distribution. Note that in this case, $L(E_i)$ denotes the volume of the set E_i, $i = 1, \ldots, M = m(m + 1)(m + 2)/6$.

4.3 Dugongs dataset

The dugongs dataset is among the standard `WinBUGS` examples and has been used by many authors in order to illustrate and compare several sampling techniques. Initially, it was analyzed by Ratkowsky (1983). The data consist of length (y) and age (x) measurements for $n = 27$

x_i	1.0	1.5	1.5	1.5	2.5	4.0	5.0	5.0	7.0	8.0	8.5	9.0	9.5	9.5
y_i	1.80	1.85	1.87	1.77	2.02	2.27	2.15	2.26	2.35	2.47	2.19	2.26	2.40	2.39
x_i	10.0	12.0	12.0	13.0	13.0	14.5	15.5	15.5	16.5	17.0	22.5	29.0	31.5	
y_i	2.41	2.50	2.32	2.43	2.47	2.56	2.65	2.47	2.64	2.56	2.70	2.72	2.57	

Table 2: Dugongs dataset.

dugongs (sea cows) captured near Townsville, Queensland. The complete dataset is shown in Table 2.

Carlin and Gelfand (1991) modeled the data using a nonlinear growth curve with no inflection point and an asymptote as x tends to infinity. More specifically, they assumed that

$$y_i \sim \mathcal{N}(\alpha - \beta\gamma^{x_i}, \tau^{-1}), \quad i = 1, \ldots, n,$$

where $\alpha, \beta, \tau > 0$ and $0 < \gamma < 1$. They then proceed using a noninformative prior for the parameters, namely, $\pi(\alpha, \beta, \gamma, \tau) \propto \tau^{1/2}\alpha^{-1}$.

The data have been also analyzed by Sahu and Zhigljavsky (2003) who adopted vague priors similar to the default ones used in `WinBUGS`. However, these particular priors do not take into account the restrictions $\alpha, \beta > 0$. Malefaki and Iliopoulos (2008) analyzed the data as well but assuming the correct parameter space. They chose as prior distribution,

$$\alpha \sim \mathcal{N}(0, \tau_\alpha^{-1})I(\alpha > 0), \ \beta \sim \mathcal{N}(0, \tau_\beta^{-1})I(\beta > 0), \ \gamma \sim \mathcal{U}(0, 1), \ \tau \sim \mathcal{G}(k, k),$$

with $\tau_\alpha = \tau_\beta = 10^{-4}$ and $k = 10^{-3}$. Note that apart from the parameter constraints these are the same priors considered by Sahu and Zhigljavsky (2003). Then, the posterior distribution of $\theta = (\alpha, \beta, \gamma, \tau)$,

$$\begin{aligned}
\pi(\theta|\text{data}) &\propto f(\text{data}|\theta)f(\alpha)f(\beta)f(\gamma)f(\tau) \\
&\propto \tau^{n/2+k-1}\exp\left\{-\frac{\tau}{2}\sum_{i=1}^n(y_i - \alpha + \beta\gamma^{x_i})^2\right\} \times \\
&\quad \exp\left(-\tau k - \frac{\tau_\alpha\alpha^2}{2} - \frac{\tau_\beta\beta^2}{2}\right)I(\alpha > 0, \beta > 0, \tau > 0, 0 < \gamma < 1),
\end{aligned}$$

can be evaluated using the output of a suitable Gibbs sampler, so the full conditional distributions of the parameters are needed. Sampling from the full conditional (posterior) distributions of α, β (truncated normals) and τ (gamma) is a straightforward task but this is not the case for γ. Instead of using a Metropolis step, we can adopt the proposed strategy. We choose a different target distribution, namely, $g(\theta|\text{data})$ by discretizing the sample space of γ, i.e., the interval $(0, 1)$ into m equal length bins and take γ to be uniformly distributed within each bin.

10

The form of the new target distribution is

$$
\begin{aligned}
g(\theta|\text{data}) \quad &\propto \quad f(\text{data}|\alpha, \beta, \gamma_{[m]}, \tau) f(\alpha) f(\beta) f(\gamma) f(\tau) \\
&\propto \quad \tau^{n/2+k-1} \exp\left\{ -\frac{\tau}{2}\sum_{i=1}^{n}(y_i - \alpha + \beta\gamma_{[m]}^{x_i})^2 \right\} \times \\
&\quad \exp\left(-\tau k - \frac{\tau_\alpha \alpha^2}{2} - \frac{\tau_\beta \beta^2}{2} \right) I(\alpha > 0, \beta > 0, \tau > 0, 0 < \gamma < 1),
\end{aligned}
$$

where $\gamma_{[m]}$ is the point that the maximum of $\pi(\theta|\text{data})$ with respect of γ is achieved in each bin. These particular choice guarantees that the importance weights have finite variance (for the proof see in the Appendix). Note that Malefaki and Iliopoulos (2008) had chosen as $\gamma_{[m]}$ the middle point of each internal but this choice does not ensure the finiteness of the weights' variance.

With g as target distribution, sampling from the full conditional distribution of γ is now an easy task. One can first draw a bin from the discrete distribution

$$
p(j|\alpha, \beta, \tau, \text{data}) \propto g(\alpha, \beta, \gamma_{[j]}, \tau|\text{data}), \quad j = 1, \ldots, m,
$$

and then simulate $u \sim \mathcal{U}(0,1)$ and set $\gamma = (j+u-1)/m$. The values of the other parameters can be simulated as usual. At each iteration, the importance weight assigned to the whole parameter vector $\theta = (\alpha, \beta, \gamma, \tau)$ is

$$
w(\theta) \propto \frac{\pi(\theta|\text{data})}{g(\theta|\text{data})} = \frac{f(\text{data}|\alpha, \beta, \gamma, \tau)}{f(\text{data}|\alpha, \beta, \gamma_{[m]}, \tau)}.
$$

Then, the jump process associated with the weighted output converges to π.

We took $m = 5$ and ran 20000 iterations of the corresponding Gibbs sampler after an additional burn-in period of 5000 iterations. Figure 4 shows the histogram and the autocorrelations of the unweighted and weighted samples of the parameters as well as the convergence of the weighted means $\hat{\alpha}_n^{IS} = \sum_{i=1}^{n} w(\theta_i)\alpha_i / \sum_{i=1}^{n} w(\theta_i)$, $\hat{\beta}_n^{IS} = \sum_{i=1}^{n} w(\theta_i)\beta_i / \sum_{i=1}^{n} w(\theta_i)$ and $\hat{\gamma}_n^{IS} = \sum_{i=1}^{n} w(\theta_i)\gamma_i / \sum_{i=1}^{n} w(\theta_i)$ to the posterior means $\mathbf{E}\{\alpha|\text{data}\}$, $\mathbf{E}\{\beta|\text{data}\}$ and $\mathbf{E}\{\gamma|\text{data}\}$, respectively.

The most interesting feature arising from the above weighted scheme is that the autocorrelations almost vanish (see Figure 4). This indicates that the Monte Carlo standard errors could be possibly estimated using methods based on iid samples. In order to confirm that, we ran 50 independent chains with total length 25000 each. After a burn-in period of 5000 iterations, we estimated for each chain the standard errors for $\hat{\alpha}_n^{IS}$, $\hat{\beta}_n^{IS}$ and $\hat{\gamma}_n^{IS}$ using the formula for the variance of the ratio estimator in the case of independent samples. The averages of these estimates were found to be 0.0010 for $\hat{\alpha}_n^{IS}$, 0.00135 for $\hat{\beta}_n^{IS}$ and 0.0004 for $\hat{\gamma}_n^{IS}$. We also estimated the same standard errors using the 50 (independent) estimates of the posterior means, $(\hat{\alpha}_{n,i}^{IS})_{(1 \leqslant i \leqslant 50)}$, $(\hat{\beta}_{n,i}^{IS})_{(1 \leqslant i \leqslant 50)}$ and $(\hat{\gamma}_{n,i}^{IS})_{(1 \leqslant i \leqslant 50)}$. The corresponding estimates were 0.0017, 0.00133 and 0.0007
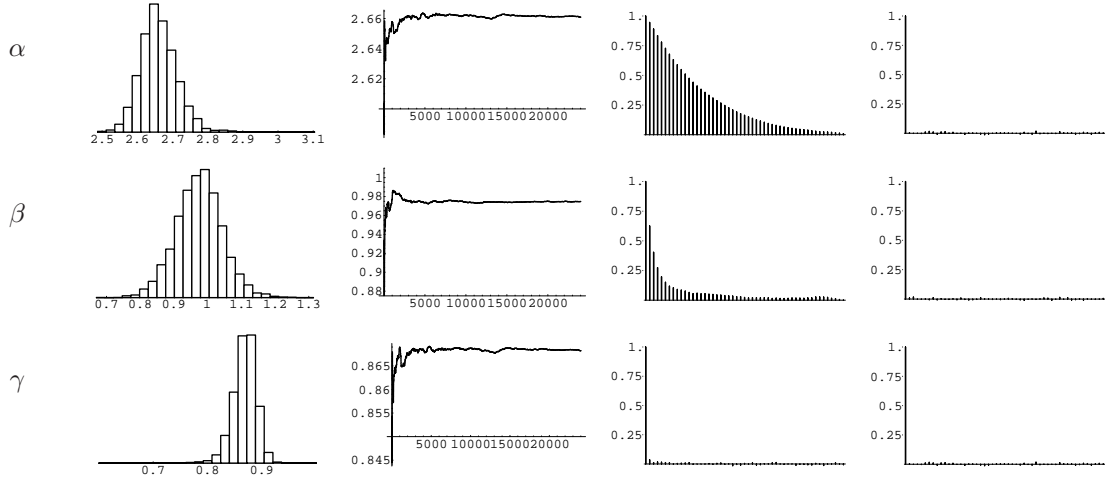
11

Figure 4: Histograms, convergence of the weighted means $\hat{\alpha}_n^{IS}$, $\hat{\beta}_n^{IS}$ and $\hat{\gamma}_n^{IS}$ and autocorrelations of the unweighted samples $(\alpha_i)_{1\leqslant i\leqslant 20000}$, $(\beta_i)_{1\leqslant i\leqslant 20000}$ and $(\gamma_i)_{1\leqslant i\leqslant 20000}$ and of the weighted samples $(\alpha_i, w(\theta_i))_{1\leqslant i\leqslant 20000}$, $(\beta_i, w(\theta_i))_{1\leqslant i\leqslant 20000}$ and $(\gamma_i, w(\theta_i))_{1\leqslant i\leqslant 20000}$ with $m = 5$ after a 5000 burn-in period for the dugongs dataset.

respectively, which are quite close to the previous ones. Hence, we conclude that it is acceptable to use the standard techniques for independent samples. Note that the autocorrelations are well defined since all parameters have finite posterior variances (for the proof see in the Appendix.)

In order to compare the precision of the above estimators with those obtained by the default method of `WinBUGS`, we also ran in that package 50 independent chains of length 25000 starting from the same values as before and took as burn-in period 5000 iterations as well. The estimated standard errors of the posterior means' estimates of α, β and γ across the 50 chains were found to be 0.0026, 0.00163 and 0.00104 respectively, that is, from 20 to 50% larger than our method's. Hence, we can safely conclude that the proposed method provides better Monte Carlo approximations to the posterior means of the above model's parameters.

4.4 Logistic regression - Challenger dataset

In 1986, the space shuttle Challenger exploded during the take-off. The explosion was the result of an O-ring failure. The accident was thought to be caused by the unusually cold weather at the time of launch, as there is reason to believe that the O-ring failure probabilities increase as temperature decreases (Dalal *et al.*, 1989). This has been concluded by data from previous space shuttle launches and O-ring failures. The dataset consist of $n = 23$ pairs (y_i, x_i) and is shown in Table 3.

A reasonable model for these data is the logistic regression model

$$\mathbf{P}(Y = 1|x) = p(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)},$$

12

Temperature (x_i)	53	57	58	63	66	67	67	67	68	69	70	70	70	70	72	73	75	75	76	76	78	79	81
Failure (y_i)	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0

Table 3: The Challenger dataset: $y_i = 1$ indicates failure of the ith O-ring.

that is, $p(x)$ is the probability of an O–ring failure at temperature x. Here, we consider the same prior as Robert and Casella (2004) for the parameters of the model. More specifically, α and β are considered a priori independent; an exponential prior with mean b is assigned to e^α and a flat prior to β, that is,

$$\pi_\alpha(\alpha)\pi_\beta(\beta) = e^{\alpha - e^\alpha/b}, \quad \alpha, \beta \in \mathbb{R}.$$

Let $\hat{\alpha}$ be the maximum likelihood estimate of α. Following Robert and Casella (2004), we set $b = \exp(\hat{\alpha} + \gamma)$, where $\gamma = 0.577216$ is the Euler's constant. Note that b is the value that makes $\mathbf{E}(\alpha) = \hat{\alpha}$. Then, the posterior distribution of the parameters is

$$
\begin{aligned}
\pi(\alpha, \beta|\text{data}) \quad \propto \quad & f(\text{data}|\alpha, \beta)\pi_\alpha(\alpha)\pi_\beta(\beta) \\
\propto \quad & \prod_{i=1}^n \left(\frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right)^{y_i} \left(\frac{1}{1 + \exp(\alpha + \beta x_i)} \right)^{1-y_i} e^{\alpha - e^\alpha/b}.
\end{aligned}
$$

For implementing the Gibbs sampler, the full conditional distributions of α and β are needed. However, they do not have a standard form. Therefore we will use the proposed approach, i.e., the discretization of the state space followed by proper weighing of the output. Since the state space of the parameters is unbounded, some transformations are needed. Setting

$$\theta = \frac{\exp\{(\alpha - \hat{\alpha})/\lambda_1\}}{1 + \exp\{(\alpha - \hat{\alpha})/\lambda_1\}} \quad \text{and} \quad \eta = \frac{\exp\{(\beta - \hat{\beta})/\lambda_2\}}{1 + \exp\{(\beta - \hat{\beta})/\lambda_2\}},$$

where λ_1 and λ_2 are some positive constants and $\hat{\beta}$ is the maximum likelihood estimator of β will serve our purpose because now the state space of both θ and η is the open interval $(0, 1)$.

It is easy to see that the posterior distribution of (θ, η) is

$$\pi(\theta, \eta|\text{data}) \propto \frac{f(\text{data}|\theta, \eta)\pi_\alpha\big(\hat{\alpha} + \lambda_1\text{logit}(\theta)\big)\pi_\beta\big(\hat{\beta} + \lambda_2\text{logit}(\eta)\big)}{\theta(1 - \theta)\eta(1 - \eta)}.$$

We now choose as target distribution

$$g(\theta, \eta|\text{data}) \propto \frac{f(\text{data}|\theta_{[m_1]}, \eta_{[m_2]})\pi_\alpha\big(\hat{\alpha} + \lambda_1\text{logit}(\theta_{[m_1]})\big)\pi_\beta\big(\hat{\beta} + \lambda_2\text{logit}(\eta_{[m_2]})\big)}{\theta_{[m_1]}(1 - \theta_{[m_1]})\ \eta_{[m_2]}(1 - \eta_{[m_2]})},$$

where $\theta_{[m_1]} = (2[m_1\theta] + 1)/2m_1$, $\eta_{[m_2]} = (2[m_2\eta] + 1)/2m_2$ and m_1 and m_2 the number of bins of the interval $(0, 1)$ corresponding to θ and η respectively.

The implementation of the Gibbs sampler with g as target distribution is an easy task. Then, in each iteration, (θ, η) is weighted by $w(\theta, \eta) = \pi(\theta, \eta|\text{data})/g(\theta, \eta|\text{data})$. Figure 5
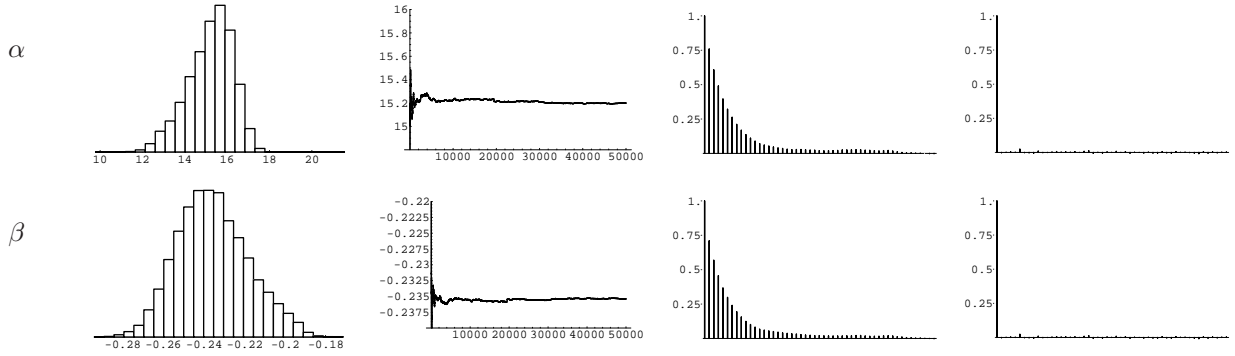
Figure 5: Histograms, convergence of the weighted means $\hat{\alpha}_n^{IS}$ and $\hat{\beta}_n^{IS}$ and autocorrelations of the unweighted samples $(\alpha_i)_{1\leqslant i\leqslant 50000}$ and $(\beta_i)_{1\leqslant i\leqslant 50000}$ and of the weighted samples $(\alpha_i, w(\theta_i, \eta_i))_{1\leqslant i\leqslant 50000}$ and $(\beta_i, w(\theta_i, \eta_i))_{1\leqslant i\leqslant 50000}$ with $m_\alpha = 20$, $m_\beta = 20$, $\lambda_1 = 0.5$ and $\lambda_2 = 0.1$ for the Challenger dataset.

shows the histogram, the autocorrelations and the convergence of the weighted means $\hat{\alpha}_n^{IS} = \sum_{i=1}^n w(\theta_i, \eta_i)\alpha_i / \sum_{i=1}^n w(\theta_i, \eta_i)$ and $\hat{\beta}_n^{IS} = \sum_{i=1}^n w(\theta_i, \eta_i)\beta_i / \sum_{i=1}^n w(\theta_i, \eta_i)$ to the posterior means $\mathbf{E}\{\alpha|\text{data}\}$ and $\mathbf{E}\{\beta|\text{data}\}$ respectively, computed from the output of 50000 updates after a burn-in period of 5000 iterations, with $m_\alpha = m_\beta = 20$, $\lambda_1 = 0.5$ and $\lambda_2 = 0.1$. We can immediately notice that the autocorrelations in the weighted samples almost vanish. (For the finiteness of the posterior variances and thus the existence of the autocorrelations see in the Appendix.) So, we can estimate the Monte Carlo standard errors using standard methods for iid samples. In order to confirm that, we repeated the procedure 50 times in order to obtain 50 independent chains of total length 50000, after corresponding burn-in periods of 5000. For each chain, we estimated the standard errors for $\hat{\alpha}_n^{IS}$ and $\hat{\beta}_n^{IS}$ using the formula for the estimation of standard error of the ratio estimator for independent samples. The averages of these estimates were 0.06381 for $\hat{\alpha}_n^{IS}$ and 9.4×10^{-4} for $\hat{\beta}_n^{IS}$. Using now the random samples consisting of the 50 independent estimates of $\mathbf{E}\{\alpha|\text{data}\}$ and $\mathbf{E}\{\beta|\text{data}\}$, i.e., $(\hat{\alpha}_{n,i}^{IS})_{(1\leqslant i\leqslant 50)}$ and $(\hat{\beta}_{n,i}^{IS})_{(1\leqslant i\leqslant 50)}$, we estimated the standard errors by 0.06562 and 9.7×10^{-4}. These values are quite close to the previous estimates and this justifies the use of methods for iid samples for estimating standard errors under our approach.

5 Summary

The aim of this paper is to present some applications of an easy-to-implement weighted scheme which is based on the discretization of the state space $\mathcal{X}$ and the proper weighting of the simulated output. As already mentioned, the method can significantly simplify MC and MCMC algorithms. We think that the most substantial benefit of the proposed weighted scheme is that, when employed in the context of MCMC algorithms, the autocorrelations almost vanish. Hence,

the standard errors of posterior expectations' estimates can be easily estimated using standard techniques for independent samples, a feature which is very useful in practical applications.

Appendix

Here, we will show that in the examples on dugongs and Challenger datasets, the parameters have finite posterior second moments. In fact, we will prove that this is the case for the corresponding importance weights; the desired result follows quite easily among the same lines.

Dugongs dataset

In the dugongs dataset example we have

$$
\int_\theta w^2(\theta) g(\theta) d(\theta) = \int_\theta w(\theta) \pi(\theta) \mathrm{d}\theta
$$

$$
= \int_{\alpha,\beta,\gamma,\tau} \exp\left\{ -\frac{\tau}{2} \left[\sum_i^n (y_i - \alpha + \beta\gamma^{x_i})^2 - \sum_i^n (y_i - \alpha + \beta\gamma_{[m]}^{x_i})^2 \right] \right\} \times
$$

$$
\tau^{n/2+k-1} \exp\left\{ \sum_i^n (y_i - \alpha + \beta\gamma^{x_i})^2 - \tau k - \frac{\tau_\alpha \alpha^2}{2} - \frac{\tau_\beta \beta^2}{2} \right\} \mathrm{d}\alpha\, \mathrm{d}\beta\, \mathrm{d}\gamma\, \mathrm{d}\tau
$$

$$
= \int_{\alpha,\beta,\gamma,\tau} \tau^{n/2+k-1} \exp\left\{ -\tau \left[\sum_i^n (y_i - \alpha + \beta\gamma^{x_i})^2 - \frac{1}{2}\sum_i^n (y_i - \alpha + \beta\gamma_{[m]}^{x_i})^2 + k \right] \right\} \times
$$

$$
\exp\left(-\frac{\tau_\alpha \alpha^2}{2} - \frac{\tau_\beta \beta^2}{2} \right) \mathrm{d}\alpha\, \mathrm{d}\beta\, \mathrm{d}\gamma\, \mathrm{d}\tau.
$$

By choosing now as $\gamma_{[m]}$ the value that maximizes the full conditional distribution of γ within each interval, we ensure that the quantity $\sum(y_i - \alpha + \beta\gamma^{x_i})^2 - \frac{1}{2}\sum(y_i - \alpha + \beta\gamma_{[m]}^{x_i})^2 + k$ is strictly positive. Thus, integrating with respect to τ we get

$$
\int_\theta w^2(\theta) g(\theta) \mathrm{d}\theta = \int_{\alpha,\beta,\gamma} \frac{\Gamma(n/2+k) \exp\left(-\frac{\tau_\alpha \alpha^2}{2} - \frac{\tau_\beta \beta^2}{2} \right)}{\left\{ \sum_i^n (y_i - \alpha + \beta\gamma^{x_i})^2 - \frac{1}{2}\sum_i^n (y_i - \alpha + \beta\gamma_{[m]}^{x_i})^2 + k \right\}^{n/2+k}} \mathrm{d}\alpha\, \mathrm{d}\beta\, \mathrm{d}\gamma
$$

$$
\leqslant \Gamma(n/2+k) k^{-n/2-k} \int_{\alpha,\beta} \exp\left(-\frac{\tau_\alpha \alpha^2}{2} - \frac{\tau_\beta \beta^2}{2} \right) \mathrm{d}\alpha\, \mathrm{d}\beta \ < \ \infty.
$$

The finiteness of the parameters' posterior second moments follows easily among the same lines.

Challenger dataset

Here we prove that the variance of the weights in the example presented in Subsection 4.4 is finite independently of the choice of the points $\theta_{[m_1]}$ and $\eta_{[m_2]}$.

The second moment of the weights is

$$\iint w(\theta,\eta)^2 g(\theta,\eta)\mathrm{d}\theta\,\mathrm{d}\eta = \iint w(\theta,\eta)\pi(\theta,\eta)\mathrm{d}\theta\,\mathrm{d}\eta$$

$$= \sum_{j=1}^{m_1}\sum_{k=1}^{m_2}\int_{\theta=(j-1)/m_1}^{j/m_1}\int_{\eta=(k-1)/m_2}^{k/m_2} w(\theta,\eta)\pi(\theta,\eta)\mathrm{d}\theta\,\mathrm{d}\eta$$

$$= \sum_{j=1}^{m_1}\sum_{k=1}^{m_2}C_{jk}\int_{\theta=(j-1)/m_1}^{j/m_1}\int_{\eta=(k-1)/m_2}^{k/m_2}\frac{\exp\left\{2\hat{\alpha}+2\lambda_1\mathrm{logit}(\theta)-2\exp\left[\hat{\alpha}+\lambda_1\mathrm{logit}(\theta)\right]/b\right\}}{\theta^2(1-\theta)^2\eta^2(1-\eta)^2}\times$$

$$\prod_{i=1}^{n}\frac{\exp\left\{2y_i\left[\hat{\alpha}+\lambda_1\mathrm{logit}(\theta)+\hat{\beta}x_i+\lambda_2 x_i\mathrm{logit}(\eta)\right]\right\}}{\left\{1+\exp\left[\hat{\alpha}+\lambda_1\mathrm{logit}(\theta)+\hat{\beta}x_i+\lambda_2 x_i\mathrm{logit}(\eta)\right]\right\}^2}\mathrm{d}\theta\,\mathrm{d}\eta,$$

where C_{jk}^{-1} equals the integrand evaluated at $\theta_{[m_1]}$, $\eta_{[m_2]}$, i.e., the midpoints of the intervals $[(j-1)/m_1, j/m_1]$, $[(k-1)/m_2, k/m_2]$, respectively. Setting now $C^* = \max C_{jk} < \infty$ we have that the above expression is less than or equal to

$$C^*\int_{\theta=0}^{1}\int_{\eta=0}^{1}\frac{\exp\left\{2\hat{\alpha}+2\lambda_1\mathrm{logit}(\theta)-2\exp\left[\hat{\alpha}+\lambda_1\mathrm{logit}(\theta)\right]/b\right\}}{\theta^2(1-\theta)^2\eta^2(1-\eta)^2}\times$$

$$\prod_{i=1}^{n}\frac{\exp\left\{2y_i\left[\hat{\alpha}+\lambda_1\mathrm{logit}(\theta)+\hat{\beta}x_i+\lambda_2 x_i\mathrm{logit}(\eta)\right]\right\}}{\left\{1+\exp\left[\hat{\alpha}+\lambda_1\mathrm{logit}(\theta)+\hat{\beta}x_i+\lambda_2 x_i\mathrm{logit}(\eta)\right]\right\}^2}\mathrm{d}\theta\,\mathrm{d}\eta =$$

$$C^{**}\int_{\theta=0}^{1}\int_{\eta=0}^{1}\frac{\exp\left\{2\lambda_1\mathrm{logit}(\theta)-2\exp\left[\lambda_1\mathrm{logit}(\theta)\right]/b\right\}}{\theta^2(1-\theta)^2\eta^2(1-\eta)^2}\times$$

$$\frac{\exp\left\{2\lambda_1\mathrm{logit}(\theta)\sum y_i+2\lambda_2\mathrm{logit}(\eta)\sum x_iy_i\right\}}{\prod_{i=1}^{n}\left\{1+C_1\exp\left[\lambda_1\mathrm{logit}(\theta)+\hat{\beta}x_i+\lambda_2 x_i\mathrm{logit}(\eta)\right]\right\}^2}$$

where $C^{**} = C^*\exp\{2\hat{\alpha}(2\sum y_i+1)+2\hat{\beta}\sum x_iy_i-2\exp(\hat{\alpha})/b\}$ and $C_1 = \exp(\hat{\alpha})$. Set $s_y = \sum y_i$, $s_x = \sum x_i$, $s_{xy} = \sum x_iy_i$. Then, the above integral is less than or equal to

$$\int_{\theta=0}^{1}\int_{\eta=0}^{1}\frac{\left(\frac{\theta}{1-\theta}\right)^{2\lambda_1(s_y+1)}\left(\frac{\eta}{1-\eta}\right)^{2\lambda_2 s_{xy}}\exp\left\{-2b^{-1}\left(\frac{\theta}{1-\theta}\right)^{\lambda_1}\right\}}{\theta^2(1-\theta)^2\eta^2(1-\eta)^2\prod_{i=1}^{n}\left\{1+C_1\exp(\hat{\beta}x_i)\left(\frac{\theta}{1-\theta}\right)^{\lambda_1}\left(\frac{\eta}{1-\eta}\right)^{\lambda_2 x_i}\right\}^2}\mathrm{d}\theta\,\mathrm{d}\eta.$$

By making now the transformations $v = \frac{\theta}{1-\theta}$, $u = \frac{\eta}{1-\eta}$, the integral takes the form:

$$I = \int_0^\infty\int_0^\infty\frac{v^{2\lambda_1(s_y+1)}u^{2\lambda_2 s_{xy}}}{\prod_{i=1}^{n}\left\{1+C_1\exp(\hat{\beta}x_i)v^{\lambda_1}u^{\lambda_2 x_i}\right\}^2}\left(1+\frac{1}{u}\right)^2\left(1+\frac{1}{v}\right)^2 e^{-2b^{-1}v^{\lambda_1}}\mathrm{d}u\mathrm{d}v. \qquad (4)$$

Consider first the integral with respect to u,

$$I_v \equiv \int_0^\infty\frac{u^{2\lambda_2 s_{xy}}}{\prod_{i=1}^{n}\left\{1+C_1\exp(\hat{\beta}x_i)v^{\lambda_1}u^{\lambda_2 x_i}\right\}^2}\left(1+\frac{1}{u}\right)^2\mathrm{d}u.$$

It holds $\prod_{i=1}^{n}\left\{1+C_1\exp(\hat{\beta}x_i)v^{\lambda_1}u^{\lambda_2 x_i}\right\} > 1+C_1\exp(\hat{\beta}s_x)v^{n\lambda_1}u^{\lambda_2 s_x}$. Thus,

$$I_v \leqslant \int_0^\infty\frac{u^{2\lambda_2 s_{xy}}}{\left\{1+C_2 v^{n\lambda_1}u^{\lambda_2 s_x}\right\}^2}\left(1+\frac{1}{u}\right)^2\mathrm{d}u \equiv I_v^*,$$

16

say, where $C_2 = C_1 \exp(\hat{\beta} s_x)$. By making the change of variable $t = C_2 v^{n\lambda_1} u^{\lambda_2 s_x}$ and setting $C_3 = C_2^{1/\lambda_2 s_x}$, $C_4 = \lambda_2 s_x C_3 C_2^{2 s_{xy}/s_x}$, we get

$$
\begin{aligned}
I_v^* &= \frac{1}{C_4 v^{\lambda_1 (2 s_{xy} + \lambda_2^{-1})/\bar{x}}} \int_0^\infty \frac{t^{(2 s_{xy} + \lambda_2^{-1})/s_x - 1}}{(1+t)^2} \left(1 + \frac{C_3 v^{\lambda_1/\lambda_2 \bar{x}}}{t^{1/\lambda_2 s_x}}\right)^2 \mathrm{d}t \\
&= \frac{1}{C_4 v^{\lambda_1 (2 s_{xy} + \lambda_2^{-1})/\bar{x}}} \left\{ \int_0^\infty \frac{t^{(2 s_{xy} + 1/\lambda_2)/s_x - 1}}{(1+t)^2} \, \mathrm{d}t + \right. \\
&\qquad \left. 2 C_3 v^{\lambda_1/\lambda_2 \bar{x}} \int_0^\infty \frac{t^{2 s_{xy}/s_x - 1}}{(1+t)^2} \, \mathrm{d}t + C_3^2 v^{2\lambda_1/\lambda_2 \bar{x}} \int_0^\infty \frac{t^{(2 s_{xy} - 1/\lambda_2)/s_x - 1}}{(1+t)^2} \, \mathrm{d}t \right\}.
\end{aligned}
$$

For any λ_2 such that $(2 s_{xy} + 1/\lambda_2)/s_x < 2$, all of the above integrals are finite. More specifically for $\lambda_2 = 0.1$ (i.e., the value we used to run the example), it holds $(2 s_{xy} + 10)/s_x = 902/1600 < 2$, so the above three integrals are equal to $\Gamma\left(\frac{451}{800}\right)\Gamma\left(\frac{1149}{800}\right)$, $\Gamma\left(\frac{223}{400}\right)\Gamma\left(\frac{577}{400}\right)$, $\Gamma\left(\frac{441}{800}\right)\Gamma\left(\frac{1159}{800}\right)$, respectively. Now, I in (4) will be finite if

$$
I_j \equiv \int_0^\infty v^{2\lambda_1 (s_y + 1) - \{2\lambda_1 s_{xy} + (1-j)\lambda_1/\lambda_2\}/\bar{x} - 2} (1+v)^2 e^{-2v/b} \, \mathrm{d}v < \infty, \quad \text{for } j = 0, 1, 2.
$$

This is true as long as $2\lambda_1(s_y + 1) - (2\lambda_1 s_{xy} + \lambda_1/\lambda_2)/\bar{x} - 1 > 0$. In particular, for $\lambda_1 = 0.5$ and $\lambda_2 = 0.1$ it holds $2 s_y - (2 s_{xy} + 10)/\bar{x} = 1654/1600 > 0$.

The finiteness of the parameters' posterior second moments can be easily proven among the same lines.

References

Carlin, B.P. and Gelfand, A.E. (1991). "An iterative Monte Carlo method for nonconjugate Bayesian analysis." *Statistics & Computing*, **1**, 119–128.

Dalal, S., Fowlkes, E. and Hoadley, B. (1989). Risk analysis of the space shuttle: pre–Challenger prediction of failure. *Journal of the American Statistical Association*, **84**: 945–957.

Fu, C.J. and Wang, L. (2002). A Random–Discretization based Monte Carlo sampling method with its applications. *Methodology and Computing in Applied Probability*, **4**, 5–25.

Kozlov, A.V. and Koller, D. (1997). Nonuniform dynamic discretization in hybrid networks. In *Proceedings of the Thirteenth Conference on uncertainty in Artificial Intelligence* (UAI–97), 314–325.

Liang, F., Liu, C. and Carroll, R.J. (2007). Stochastic approximation in Monte Carlo computation. *Journal of the American Statistical Association*, **102**, 305–320.

Liu, J.S. (2001). *Monte Carlo Strategies in Scientific Computing*. New York: Springer–Verlag.

Liu, J.S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, **93**, 1032–1044.

Limnios, N. and Oprişan, G. (2001). *Semi–Markov processes and reliability.* Statistics for industry and technology, Birkhäuser.

Malefaki, S. and Iliopoulos, G. (2008). On convergence of properly weighted samples to the target distribution. *Journal of Statistical Planning and Inference*, **138**, 1210–1225.

Neil, M., Tailor, M., and Marquez, D. (2007). Inference in hybrid Bayesian networks using dynamic discretization. *Statistics and Computing*, **17**, 219–233.

Ratkowsky, D. (1983). *Nonlinear regression modeling.* Marcel Dekker, New York.

Robert, C.P. and Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.), New York: Springer–Verlag.

Sahu, S.K., and Zhigljavsky, A.A. (2003). Self–regenerative Markov Chain Monte Carlo with adaptation. *Bernoulli*, **9**, 395–422.

Sainudiin, R. and York, T. (2006). An auto-validating rejection sampler. `http://arxiv.org/abs/math/0611479`.