

A REJECTION SAMPLING SCHEME FOR POSTERIOR PROBABILITY DISTRIBUTIONS VIA THE RATIO-OF-UNIFORMS METHOD

Luca Martino, Joaquín Míguez

Department of Signal Theory and Communications, Universidad Carlos III de Madrid.
Avenida de la Universidad 30, 28911 Leganés, Madrid, Spain.
E-mail: luca@tsc.uc3m.es, joaquin.miguez@uc3m.es

ABSTRACT

Accept/reject sampling is a well-known method to generate random samples from arbitrary target probability distributions. It demands the design of a suitable proposal probability density function (pdf) from which candidate samples can be drawn. The main limitation to the use of RS is the need to find an adequate upper bound for the ratio of the target pdf over the proposal pdf from which the samples are generated. There are no general methods to analytically find this bound, except when the target pdf is log-concave. In this paper we introduce a novel procedure using the ratio of uniforms method to efficiently perform rejection sampling for a large class of target densities. The candidate samples are generated using only two independent uniform random variables. In order to illustrate the application of the proposed technique, we provide a numerical example.

1. INTRODUCTION

Bayesian methods have become very popular in signal processing and, with them, there has been a surge of interest in the Monte Carlo techniques that are often necessary for the implementation of optimal a posteriori estimators [9]. Rejection sampling (RS) is a “universal” Monte Carlo technique that can be used to generate samples from any target probability density function (pdf) that we can evaluate up to a proportionality constant, by drawing from a possibly simpler proposal density. The sample is either accepted or rejected by an adequate test of the ratio of the two pdfs and it can be proved that accepted samples are actually distributed according to the target distribution. The performance of RS methods depends on the availability of tight bounds for the ratio of the target and proposal densities and, in general, the proportion of accepted candidate samples is related to the similarity of the two densities.

There are no general methods to construct adequate proposal pdf’s, though. One exception is the so-called adaptive rejection sampling (ARS) method [3] which, given a target density, provides a procedure to obtain a suitable proposal function (for which the bound is easy to compute). Unfortunately, this procedure is only valid when the target pdf is strictly log-concave, which is often not the case in practice. In [7], an alternative ARS-type algorithm is introduced to generate samples from posterior probability distributions. It relies on computing a piecewise-constant approximation of the likelihood and a proposal density that consists in a mixture of truncated pdf’s. The application of this technique has two potential shortcomings, though: it requires the ability to integrate the prior pdf of the random variable of interest in any finite interval and to (efficiently) draw from truncated densities (which is not necessarily straightforward).

In this paper, we introduce a novel method to generate samples from the posterior distribution of a random variable x given a collection of data \mathbf{y} . The procedure is based on the ratio-of-uniforms (RoU) method, that was developed in [4, 5, 6, 10] to draw from log-concave pdf’s. Here, we extend

its applicability to a larger class of densities, possibly multimodal. Similarly to the work in [7], the algorithm is based on obtaining a piecewise-constant bound for the likelihood, but the generation of candidate samples is much simpler (only two uniform variables are needed) and does not require the integrability of the prior of x .

The basic problem and background material is stated in Section 2. Some useful denitions and basic assumptions are introduced in Section 3. The calculation of bounds is presented in Section 4. The new algorithm is derived in Section 5 and in Section 6 we apply it to a simple numerical example. Finally, Section 7 is devoted to the conclusions.

2. PROBLEM STATEMENT AND BACKGROUND

2.1 Signal model

Many signal processing problems involve the estimation of an unobserved signal of interest (SoI), denoted $x \in \mathbb{R}$, from a sequence of related observations $\mathbf{y} \in \mathbb{R}^n$ (vectors are denoted as lower-case bold-face letters all through the paper). We consider n scalar observations, $\mathbf{y} = [y_1, \dots, y_n]^\top$, which are obtained through nonlinear transformations of the signal x contaminated with additive noise. Formally, we write

$$y_1 = g_1(x) + \vartheta_1, \dots, y_n = g_n(x) + \vartheta_n, \quad (1)$$

where $g_i : \mathbb{R} \rightarrow \mathbb{R}$, $i = 1, \dots, n$, are nonlinearities and $\boldsymbol{\vartheta} = [\vartheta_1, \dots, \vartheta_n]^\top$ is a vector of independent noise variables. The joint noise pdf¹ has the form

$$p(\vartheta_1, \dots, \vartheta_n) = k \exp \left\{ - \sum_{i=1}^n \bar{V}_i(\vartheta_i) \right\}, \quad (2)$$

where $k > 0$ is a real constant and $\bar{V}_i : \mathbb{R} \rightarrow [0, +\infty)$ are real functions, subsequently referred to as *marginal potentials*.

We also assume that every $\bar{V}_i(\vartheta_i)$ is a convex function with a minimum at $\vartheta = \mu_i$. Without loss of generality, we set $\mu_i = 0$, $i = 1, \dots, n$, in the sequel. Note that, we can always consider $\bar{V}_i(\vartheta_i + \mu_i)$ and modify the corresponding observation as $\bar{y}_i = y_i - \mu_i$ to obtain a system equivalent to Eq. (1).

The scalar observations are conditionally independent given the SoI x , hence the *likelihood function* $p(\mathbf{y}|x)$ can be written in terms of the marginal potentials as

$$p(\mathbf{y}|x) \propto \exp \left\{ - \sum_{i=1}^n \bar{V}_i(y_i - g_i(x)) \right\}, \quad (3)$$

¹We use $p(\cdot)$ to denote the pdf of a random variate, i.e., $p(x)$ denotes the pdf of x and $p(y)$ is the pdf of y , possibly different. The conditional pdf of x given the observation of y is written as $p(x|y)$.

We also assume a prior pdf for the SoI x of the form

$$p(x) \propto \exp\{-\bar{V}_{n+1}(g_{n+1}(x))\}, \quad (4)$$

where $\bar{V}_{n+1}(\vartheta_{n+1})$ is a convex function with a minimum at $\mu_{n+1} = 0$ and $g_{n+1}(x)$ is a nonlinear function. Taking Eqs. (3) and (4) together, the posterior pdf of the SoI takes the form

$$\begin{aligned} p(x|\mathbf{y}) &\propto p(\mathbf{y}|x)p(x) = \\ &= k \exp\left\{-\sum_{i=1}^n \bar{V}_i(y_i - g_i(x)) - \bar{V}_{n+1}(g_{n+1}(x))\right\}. \end{aligned} \quad (5)$$

The aim of this paper is to design an efficient RS method to generate independent and identically distributed (i.i.d.) samples from pdf's of the form of Eq. (5).

In order, to rewrite the posterior density $p(x|\mathbf{y})$ in a more compact manner, we define the generalized observation vector $\tilde{\mathbf{y}} \triangleq [y_1, \dots, y_n, 0]^\top$ and the generalized nonlinearity vector $\tilde{\mathbf{g}}(x) \triangleq [g_1(x), \dots, g_n(x), -g_{n+1}(x)]^\top$. Then, Eq. (5) induces a *system potential* $V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) : \mathbb{R} \rightarrow \mathbb{R}$, defined as

$$V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) \triangleq \sum_{i=1}^{n+1} \bar{V}_i(y_i - g_i(x)). \quad (6)$$

Note that, in order to apply RS, one often needs to find an upper bound for the posterior $p(x|\mathbf{y})$, which is equivalent to find a lower bound for the system potential $V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$.

2.2 Ratio of uniforms

The RoU method [2, 4] is a sampling technique that relies on the following theorem.

Theorem: Let $p(x) \geq 0$ be a pdf known only up to a proportionality constant. If (u, v) is a sample drawn from the uniform distribution on the set

$$\mathcal{A} = \left\{ (v, u) : 0 \leq u \leq \sqrt{p(v/u)} \right\}, \quad (7)$$

then $x = \frac{v}{u}$ is a sample from $p(x)$.

Proof: See [2, Theorem 7.1].

Therefore, if we are able to draw uniformly from \mathcal{A} , then we can also draw from the pdf $p(x)$. The cases of practical interest are that in which the region \mathcal{A} is bounded and \mathcal{A} is bounded if and only if, both $\sqrt{p(x)}$ and $x\sqrt{p(x)}$ are bounded.

Figure 1 (left) depicts a bounded set \mathcal{A} . Note that, for every angle $\alpha \in (0, \pi)$ rad, we can draw a straight line that passes through the origin $(0, 0)$ and contains points $(v_i, u_i) \in \mathcal{A}$ such that $x = \frac{v_i}{u_i} = \tan(\alpha)$, i.e., every point (v_i, u_i) in the straight line with angle α yields the same value of x .

From the definition of \mathcal{A} , $u_i \leq p(x)$ and $v_i = u_i x \leq x\sqrt{p(x)}$, hence, if we choose the point (v_2, u_2) that lies on the boundary of \mathcal{A} , $u_2 = \sqrt{p(x)}$ and $v_2 = x\sqrt{p(x)}$, and we can embed the set \mathcal{A} in the rectangular region

$$\begin{aligned} \mathcal{R} = \left\{ (v', u') : 0 \leq u' \leq \sup_x \sqrt{p(x)}, \right. \\ \left. \inf_x x\sqrt{p(x)} \leq v' \leq \sup_x x\sqrt{p(x)} \right\}, \end{aligned} \quad (8)$$

as drawn in Fig. 1 (middle).

Once \mathcal{R} is constructed, it is straightforward to draw uniformly from \mathcal{A} by RS: simply draw uniformly from \mathcal{R} and then check whether the candidate point belongs to \mathcal{A} .

2.3 Goal

In [5, 6], the RoU technique has been applied to obtain an alternative implementation of the adaptive RS algorithm for log-concave pdf's.

We aim at using the RoU method to draw i.i.d. samples from the posterior pdf $p(x|\mathbf{y})$, not necessarily log-concave, by a RS procedure. If \mathcal{A} is bounded, the advantage w.r.t. conventional rejection samplers is that we only need to simulate two uniform random variables, which is straightforward.

We will show that it is possible to obtain the embedding rectangular region \mathcal{R} by constructing a piecewise-constant approximation of $\sqrt{p(x|\mathbf{y})}$ and $x\sqrt{p(x|\mathbf{y})}$.

3. ASSUMPTIONS AND DEFINITIONS

Let $\mathcal{D} \subseteq \mathbb{R}$ denote the support of the vector function $\tilde{\mathbf{g}}(x)$, i.e., $\tilde{\mathbf{g}} : \mathcal{D} \subseteq \mathbb{R} \rightarrow \mathbb{R}^{n+1}$. We assume that there exists a partition of \mathcal{D} consisting of intervals $\mathcal{B}_1 = (b_1^-, b_1^+), \dots, \mathcal{B}_q = (b_q^-, b_q^+)$ such that $\mathcal{D} = \cup_{j=1}^q [\mathcal{B}_j]$ (where $[\mathcal{B}_j] = [b_j^-, b_j^+]$ is the closure of \mathcal{B}_j), $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$, $\forall i \neq j$, and $\forall x \in \mathcal{B}_j$ every function $g_i(x)$, $i = 1, \dots, n+1$, is

- invertible and,
- either convex or concave.

Assumptions a) and b) together mean that $\forall i \in \{1, \dots, n+1\}$ and $\forall x \in \mathcal{B}_j$, the first derivative $\frac{dg_i}{dx}$ is either strictly positive or strictly negative and the second derivative $\frac{d^2g_i}{dx^2}$ is either non-negative or non-positive.

We also introduce the set of simple estimates

$$\mathcal{X} \triangleq \{x' \in \mathbb{R} : g_i(x') = y_i, \quad i = 1, \dots, n+1\}. \quad (9)$$

and the set of support points $\mathcal{S} \triangleq \{s_1, \dots, s_m\}$, where $s_k \in \mathcal{D}$ $\forall k$ and $s_1 < s_2 < \dots < s_m$. We assume that \mathcal{S} contains, at least:

- Every $b_j^- > -\infty$ and $b_j^+ < +\infty$, $j = 1, \dots, q$. As a consequence, given any finite interval $\mathcal{I}_k = [s_k, s_{k+1}]$, $k = 1, \dots, m-1$, every $g_i(x)$ is invertible and either convex or concave in \mathcal{I}_k .
- Every simple estimate, i.e., $\mathcal{X} \subseteq \mathcal{S}$.
- The point 0, i.e., $\exists k'$ such that $s_{k'} = 0 \in \mathcal{S}$.

Besides the intervals $\mathcal{I}_1 = [s_1, s_2], \dots, \mathcal{I}_{m-1} = [s_{m-1}, s_m]$ constructed with the points in \mathcal{S} , we also define the semi-open intervals $\mathcal{I}_0 = (-\infty, s_1]$ and $\mathcal{I}_m = [s_m, +\infty)$.

4. COMPUTATION OF BOUNDS

In order to obtain the rectangular region \mathcal{R} that enables us to draw uniformly from \mathcal{A} and, therefore, from the target density $p(x|\mathbf{y})$, we need to find bounds for the functions $\sqrt{p(x|\mathbf{y})}$ and $x\sqrt{p(x|\mathbf{y})}$.

Specifically, since $p(x|\mathbf{y}) \geq 0$, we only need the upper bounds $L^{(1)} \geq \sqrt{p(x|\mathbf{y})}$, $L^{(2)} \geq x\sqrt{p(x|\mathbf{y})}$ and $L^{(3)} \geq -x\sqrt{p(x|\mathbf{y})}$ (note that $-L^{(3)}$ is a lower bound for $x\sqrt{p(x|\mathbf{y})}$).

We associate a potential $V^{(i)}$ for each function of interest. Specifically, since $p(x|\mathbf{y}) \propto \exp\{-V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})\}$ we readily obtain that

$$\begin{aligned} \sqrt{p(x|\mathbf{y})} &\propto \exp\{-V^{(1)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})\}, \\ x\sqrt{p(x|\mathbf{y})} &\propto \exp\{-V^{(2)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})\}, \\ -x\sqrt{p(x|\mathbf{y})} &\propto \exp\{-V^{(3)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})\}, \end{aligned}$$

where $V^{(1)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) = \frac{1}{2}V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$ and $V^{(2)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) = \frac{1}{2}V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) - \log(x)$ and $V^{(3)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) = \frac{1}{2}V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) - \log(-x)$ ($x < 0$), respectively.

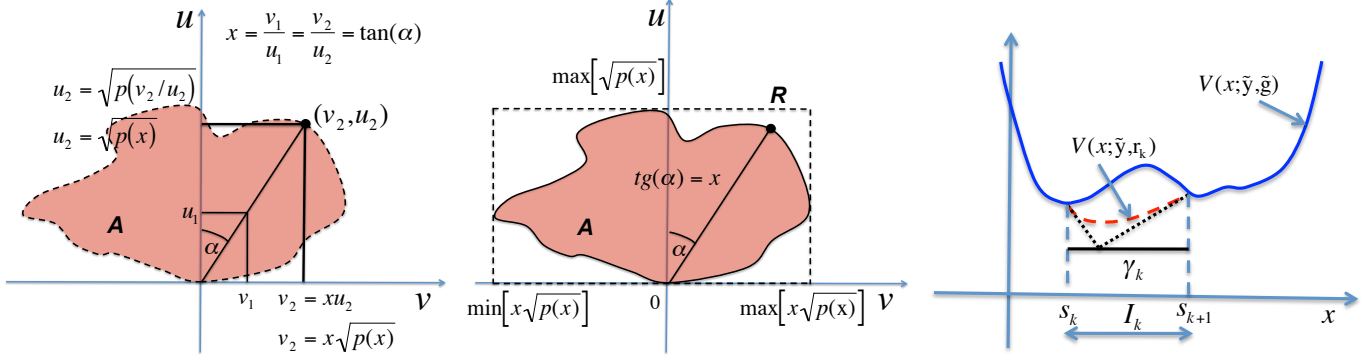


Figure 1: *Left:* A bounded region \mathcal{A} and the straight line $v = xu$ corresponding to the sample $x = \tan(\alpha)$. Every point in the intersection of the line $v = xu$ and the set \mathcal{A} yields the same sample x . The point on the boundary (v_2, u_2) has coordinates $v_2 = x\sqrt{p(x)}$ and $u_2 = \sqrt{p(x)}$. *Middle:* If the two functions $\sqrt{p(x)}$ and $x\sqrt{p(x)}$ are bounded, the set \mathcal{A} is bounded and embedded in the rectangle \mathcal{R} . *Right:* The solid line shows the system potential $V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$, the dashed line depicts the modified potential $V(x; \tilde{\mathbf{y}}, \mathbf{r}_k)$ and the dotted line illustrates the two linear functions tangent to the modified potential at s_k and s_{k+1} . The lower bound γ_k is obtained as the intersection of the two tangent lines.

Note that it is equivalent to maximize the functions $\sqrt{p(x|y)}$, $x\sqrt{p(x|y)}$, $-x\sqrt{p(x|y)}$ with respect to (w.r.t.) x and to minimize the corresponding potentials $V^{(j)}$, $j = 1, 2, 3$, also w.r.t. x . As a consequence, we may focus on the calculation of lower bounds $\gamma^{(j)} \leq V^{(j)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$, related to the upper bounds as $L^{(j)} = \exp\{-\gamma^{(j)}\}$, $j = 1, 2, 3$. This problem is far from trivial, though. Even for very simple marginal potentials, \bar{V}_i , $i = 1, \dots, n$, the potential functions, $V^{(j)}$, $j = 1, 2, 3$, can be highly multimodal w.r.t. x [8].

4.1 Lower bound for the system potential

We now briefly review a basic algorithm for the calculation of upper bounds for posterior pdf's [8]. Our goal is to obtain an analytical method for the computation of a scalar $\gamma_k \in \mathbb{R}$ such that $\gamma_k \leq \inf_{x \in \mathcal{I}_k} V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$ in an interval $x \in \mathcal{I}_k = [s_{k-1}, s_k]$, where $s_{k-1}, s_k \in \mathcal{S}$ (and $s_{k-1} < s_k$). The observations $\tilde{\mathbf{y}}$ are arbitrary but fixed and the nonlinearities $\tilde{\mathbf{g}}$ are assumed known.

In [8], it is described how to build, within each set \mathcal{I}_k , $k = 0, \dots, m$, adequate linear functions $\{r_{i,k}\}_{i=1}^{n+1}$ in order to replace the nonlinearities $\{g_i\}_{i=1}^{n+1}$. Given $\tilde{\mathbf{g}}(x) = [g_1(x), \dots, -g_{n+1}(x)]$, we construct a vector $\mathbf{r}_k(x) = [r_{1,k}(x), \dots, r_{n+1,k}(x)]^\top$ in a way that ensures

$$V(x; \tilde{\mathbf{y}}, \mathbf{r}_k) \leq V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}), \quad \forall x \in \mathcal{I}_k, \quad (10)$$

and, as a consequence,

$$\gamma_k = \inf_{x \in \mathcal{I}_k} V(x; \tilde{\mathbf{y}}, \mathbf{r}_k) \leq \inf_{x \in \mathcal{I}_k} V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}). \quad (11)$$

Therefore, it is possible to find a lower bound γ_k in \mathcal{I}_k for the system potential $V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$ by minimizing the modified potential $V(x; \tilde{\mathbf{y}}, \mathbf{r}_k)$ within \mathcal{I}_k .

It is often possible to obtain γ_k in closed-form. However, if the minimization of $V(x; \tilde{\mathbf{y}}, \mathbf{r}_k)$ remains intractable, we can exploit the fact that the modified system potential $V(x; \tilde{\mathbf{y}}, \mathbf{r}_k)$ is convex $\forall x \in \mathcal{I}_k$. Indeed, since the marginal potentials \bar{V}_i are convex and the functions $r_{i,k}(x)$ are linear, with $i = 1, \dots, n+1$, it turns out that $V(x; \tilde{\mathbf{y}}, \mathbf{r}_k) = \sum_{i=0}^{n+1} \bar{V}_i(y_i - r_{i,k}(x))$ is also convex in \mathcal{I}_k . In fact, each term

in the sum is convex, since

$$\begin{aligned} \frac{d^2 \bar{V}_i}{dx^2} &= \frac{d\bar{V}_i}{d\vartheta} \frac{d^2 r_{i,k}}{dx^2} + \left(\frac{dr_{i,k}}{dx} \right)^2 \frac{d^2 \bar{V}_i}{d\vartheta^2} \\ &= 0 + \left(\frac{dr_{i,k}}{dx} \right)^2 \frac{d^2 \bar{V}_i}{d\vartheta^2} \geq 0, \end{aligned} \quad (12)$$

where we have used that $\frac{d^2 r_{i,k}}{dx^2} = 0$ (since $r_{i,k}$ is linear) and the convexity of the marginal potentials $\bar{V}_i(\vartheta)$, $i = 1, \dots, n+1$.

The convexity of $V(x; \tilde{\mathbf{y}}, \mathbf{r}_k)$ implies that we can find a lower bound in \mathcal{I}_k for the system potential $V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$ at the intersection of the tangents to $V(x; \tilde{\mathbf{y}}, \mathbf{r}_k)$ at the limit points of \mathcal{I}_k , i.e. s_{k-1} and s_k . Figure 1 (right) depicts this procedure to obtain a lower bound in an interval \mathcal{I}_k for a system potential $V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$ (solid line) using the intersection of two tangent lines (dotted line) to the modified potential $V(x; \tilde{\mathbf{y}}, \mathbf{r}_k)$ (dashed line).

A global lower bound for the system potential can be found as

$$\gamma = \min_{k \in \{0, \dots, m\}} \gamma_k \quad (13)$$

Note that the quality of the bound depends of the number of support points. Hence, we can improve it by increasing the number of points in \mathcal{S} . Since in [8] we prove that all the stationary points (maxima and minima) of the posterior pdf are contained in the interval $\mathcal{J} = [\min \mathcal{X}, \max \mathcal{X}]$, though, the technique provides improved bounds only if we add support points inside the interval \mathcal{J} .

4.2 Lower Bounds for $V^{(1)}$, $V^{(2)}$ and $V^{(3)}$

It is straightforward to see that we can use the procedure described above to find a global lower bound $\gamma^{(1)}$ for the potential $V^{(1)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) = \frac{1}{2} V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$, associated to the function $\sqrt{p(x|y)}$, since $V^{(1)}$ is a scaled version of the system potential V .

The procedure of Section 4.1 can also be applied to find upper bounds for $x\sqrt{p(x|y)}$, with $x > 0$, and $-x\sqrt{p(x|y)}$ with $x < 0$. Indeed, recalling that the associated potentials are $V^{(2)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) = \frac{1}{2} V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) - \log(x)$, $x > 0$, and $V^{(3)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) = \frac{1}{2} V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}}) - \log(-x)$, $x < 0$, it is straight-

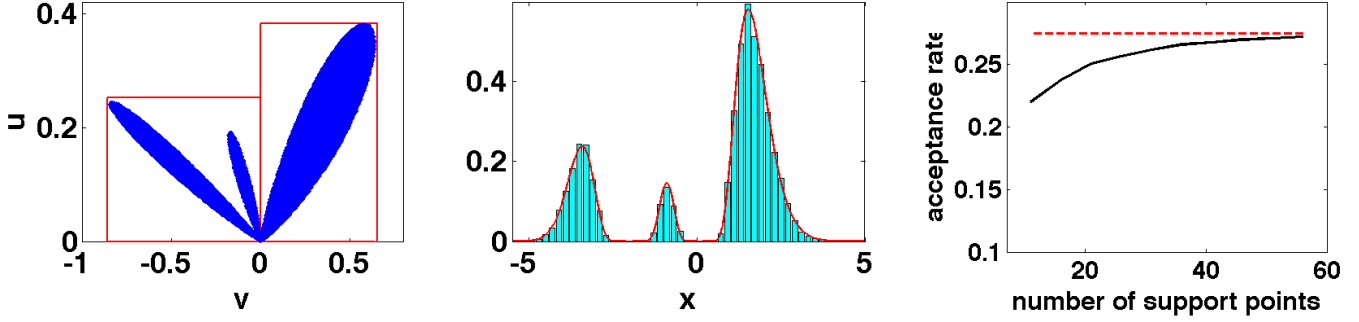


Figure 2: *Left:* The region \mathcal{A} corresponding to the target pdf $p(x|\mathbf{y})$ in the example, and the two rectangular regions that include \mathcal{A} , obtained using 20 support points in $\mathcal{J} = [\min \mathcal{X}, \max \mathcal{X}]$. *Middle:* The posterior density $p(x|\mathbf{y}) \propto \exp\{-V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})\}$ and the normalized histogram of $N = 10,000$ samples obtained using RS over the area \mathcal{A} . *Right:* The curve of acceptance rates (averaged over 10,000 simulations) as a function of the number m of support points chosen randomly in the interval $\mathcal{J} = [\min \mathcal{X} = -4.01, \max \mathcal{X} = 3.16]$.

forward to realize that the corresponding modified potentials

$$\begin{aligned} V^{(2)}(x; \tilde{\mathbf{y}}, \mathbf{r}_k) &\triangleq \frac{1}{2}V(x; \tilde{\mathbf{y}}, \mathbf{r}_k) - \log(x), \\ V^{(3)}(x; \tilde{\mathbf{y}}, \mathbf{r}_k) &\triangleq \frac{1}{2}V(x; \tilde{\mathbf{y}}, \mathbf{r}_k) - \log(-x), \end{aligned} \quad (14)$$

are convex in \mathcal{I}_k , since the functions $-\log(x)$ ($x > 0$) and $-\log(-x)$ ($x < 0$), are also convex. Therefore, it is always possible to compute lower bounds $\gamma^{(2)} \leq V^{(2)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$ and $\gamma^{(3)} \leq V^{(3)}(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$.

The corresponding upper bounds are $L^{(j)} = \exp\{-\gamma^{(j)}\}$, $j = 1, 2, 3$.

5. ALGORITHM

The basic RS algorithm to draw from $p(x|\mathbf{y})$ can be summarized in the following steps:

1. Compute the bounds $L^{(i)}$, $i = 1, 2, 3$ as shown in Sections 4.1 and 4.2.
2. Draw a point $[v', u']^\top \sim \mathcal{U}_{\mathcal{R}}$ uniformly from the rectangular region $\mathcal{R} = [-L^{(3)}, L^{(2)}] \times [0, L^{(1)}] \supseteq \mathcal{A}$. This is easily achieved by sampling $v' \sim \mathcal{U}(-L^{(3)}, L^{(2)})$ and $u' \sim \mathcal{U}(0, L^{(1)})$, where $\mathcal{U}(a, b)$ denotes the uniform distribution in the interval $[a, b]$.
3. If $u' \leq \sqrt{p(\frac{v'}{u'}|\mathbf{y})}$, then accept the sample $x' = \frac{v'}{u'}$.
4. If $u' > \sqrt{p(\frac{v'}{u'}|\mathbf{y})}$, then reject the sample $x' = \frac{v'}{u'}$.

A simple way to improve the acceptance rate of the rejection sampler is to compute two different bounds for $\sqrt{p(x|\mathbf{y})}$, namely $L_p^{(1)} \geq \sqrt{p(x|\mathbf{y})}$, $\forall x > 0$ (positive), and $L_p^{(1)} \geq \sqrt{p(x|\mathbf{y})}$, $\forall x < 0$ (negative). As a result, we define the rectangular regions

$$\begin{aligned} \mathcal{R}_1 &= [-L^{(3)}, 0] \times [0, L_n^{(1)}], \\ \mathcal{R}_2 &= [0, L^{(2)}] \times [0, L_p^{(1)}]. \end{aligned} \quad (15)$$

The union set contains the bounded region \mathcal{A} , i.e., $\mathcal{A} \subseteq \mathcal{R}_1 \cup \mathcal{R}_2$, hence we can perform RS by drawing candidate samples from $\mathcal{R}_1 \cup \mathcal{R}_2$. Note that, for a uniform distribution, the probability of \mathcal{R}_k is proportional to its measure $|\mathcal{R}_k|$, where $|\mathcal{R}_1| = |L^{(2)} - L^{(3)}| \cdot |L_n^{(1)}|$ and $|\mathcal{R}_2| = |L^{(2)} - L^{(3)}| \cdot |L_p^{(1)}|$. Specifically, the probability of \mathcal{R}_k is

$$w_k \triangleq \frac{|\mathcal{R}_k|}{|\mathcal{R}_1| + |\mathcal{R}_2|} \quad \text{with } k = 1, 2. \quad (16)$$

The modified RS algorithm can be summarized as follows:

1. Compute the bounds $L_n^{(1)}$, $L_p^{(1)}$, $L^{(2)}$ and $L^{(3)}$.
2. Draw an index $k \in \{1, 2\}$ from the probability distribution $P\{k = 1\} = w_1$, $P\{k = 2\} = w_2$.
3. Draw a point (v', u') uniformly from the region \mathcal{R}_k .
4. If $u' \leq \sqrt{p(\frac{v'}{u'}|\mathbf{y})}$, then accept the sample $x' = \frac{v'}{u'}$.
5. If $u' > \sqrt{p(\frac{v'}{u'}|\mathbf{y})}$, then reject the sample $x' = \frac{v'}{u'}$.

6. EXAMPLE

We illustrate the application of the proposed RS scheme by way of a simple numerical example. Consider the problem of drawing from the posterior distribution of $x \in \mathbb{R}$ given the observations

$$y_1 = \log[(x + 2)^2] + \vartheta_1, \quad y_2 = \log[(x - 0.1)^2] + \vartheta_2 \quad (17)$$

where $g_1(x) = \log[(x + 2)^2]$, $g_2(x) = \log[(x - 0.1)^2]$ and the noise variables ϑ_1 and ϑ_2 are independent and identically Gaussian distributed, $\vartheta_i \sim N(\vartheta_i; 0, 1/2)$, $i = 1, 2$, i.e., $p(\vartheta_i) \propto \exp\{-\bar{V}_i(\vartheta_i)\}$, where $\bar{V}_i(\vartheta_i) = \vartheta_i^2$.

We also assume a prior density of the form

$$p(x) \propto \exp\left\{-\frac{(x^2 - 10)^2}{100}\right\} \quad (18)$$

that we interpret as $\bar{V}(\vartheta_3) = \frac{\vartheta_3^2}{100}$ and $g_3(x) = x^2 - 10$. Therefore, the generalized observation vector and the generalized vector of nonlinearities are defined as $\tilde{\mathbf{y}} = [y_1, y_2, 0]^\top$ and $\tilde{\mathbf{g}}(x) = [g_1(x), g_2(x), -g_3(x)]^\top$, respectively, and the posterior pdf of x given \mathbf{y} can be written as $p(x|\mathbf{y}) \propto \exp\{-V(x; \tilde{\mathbf{y}}; \tilde{\mathbf{g}})\}$, where the generalized system potential

$$\begin{aligned} V(x; \tilde{\mathbf{y}}; \tilde{\mathbf{g}}) &= (y_1 - \log[(x + 2)^2])^2 + \\ &\quad (y_2 - \log[(x - 0.1)^2])^2 + \frac{1}{100}(x^2 - 10)^2, \end{aligned} \quad (19)$$

is, in general, non-convex.

The set of simple estimates is

$$\begin{aligned} \mathcal{X} &= \{\sqrt{\exp(y_1)} - 2, -\sqrt{\exp(y_1)} - 2, \\ &\quad \sqrt{\exp(y_2)} + 0.1, -\sqrt{\exp(y_2)} + 0.1, \\ &\quad \sqrt{10}, -\sqrt{10}\}, \end{aligned} \quad (20)$$

which setting, e.g., $y_1 = 1.4$ and $y_2 = 1$, yields $\mathcal{X} = \{0.01, -4.01, 1.74, -1.54, 3.16, -3.16\}$.

We can divide the domain \mathbb{R} in $q = 3$ intervals $\mathcal{B}_1 = (-\infty, -1]$, $\mathcal{B}_2 = [-1, 1]$ and $\mathcal{B}_3 = [1, +\infty]$ where the nonlinearities have first and second derivatives with constant sign. Hence, the set of support points \mathcal{S} have to include at least the values $-1, 0, 1$ and the set of the simple estimates \mathcal{X} , i.e., $\mathcal{S} = \mathcal{X} \cup \{-1, 0, 1\}$ sorted in ascending order. Therefore, we have at least $m = 9$ support points.

Figure 2 (left) depicts the bounded set \mathcal{A} corresponding to the target pdf $p(x|\mathbf{y})$ and the two rectangular regions \mathcal{R}_1 and \mathcal{R}_2 such that $\mathcal{A} \subseteq \mathcal{R}_1 \cup \mathcal{R}_2$, obtained using $m = 20$ support points in $\mathcal{J} = [\min \mathcal{X}, \max \mathcal{X}]$. With this setup, we apply the modified RS algorithm of Section 5 to generate 10,000 i.i.d. samples from $p(x|\mathbf{y})$. The resulting histogram, and the true function $p(x|\mathbf{y})$ are depicted in Figure 2 (middle).

Finally, Figure 2 (right) shows the acceptance rate² of the modified RS algorithm as a function of the number of support points, m , chosen randomly in the interval $\mathcal{J} = [\min \mathcal{X} = -4.01, \max \mathcal{X} = 3.16]$. With $m = 9$, the acceptance rate is $\approx 21\%$, while $m = 50$ yields an acceptance rate of $\approx 27\%$. The dashed line represents the highest achievable acceptance rate for this example (27.6%). It is clear that a trade off between computational complexity (that grows linearly with the number of support points) and performance can be sought.

7. CONCLUSIONS AND FUTURE WORK

We have proposed a novel rejection sampling method to draw from the (possibly multimodal) posterior pdf of a signal of interest x given a vector of nonlinear observations \mathbf{y} with arbitrary dimension. The proposed algorithm relies on the ratio-of-uniforms technique and, therefore, it only requires the generation of two independent uniform random variables from finite intervals in order to generate candidate sample of x (which may, in general, have an infinite domain).

The new technique can be used inside more elaborate Monte Carlo methods, such as MCMC algorithms and particle filters. Moreover, the proposed modified rejection sampler described in Section 5 can be readily extended to an adaptive version by covering the bounded set \mathcal{A} with k disjoint rectangular regions $\mathcal{R}_1, \dots, \mathcal{R}_k$ (in Section 5, we have restricted ourselves to the simplest case $k = 2$). In this way, an acceptance rate close to 100% is attainable.

Another way in which the proposed method can be extended involves the approximation of the target density by vertical bars [1].

These bars can be converted into triangles [4] $\mathcal{T}_1, \dots, \mathcal{T}_k$ (that cover the bounded set \mathcal{A}) in the $u - v$ reference axes and there are simple methods to draw uniformly from these triangular regions. With this approach, we can easily obtain acceptance rates close to 100% when the target density has a finite support, but there are difficulties, in general, to apply it when this support is unbounded (specifically to approximate the tails of the target pdf).

8. ACKNOWLEDGEMENTS

This work was partially supported by the Ministry of Science and Innovation of Spain (project TEC2009-14504-C02-01 DEIPRO and program Consolider-Ingenio 2010 CSD2008-00010 COMONSENS).

The final version of this paper has benefitted considerably from the insightful comments of one of the anonymous reviewers.

References

- [1] J. H. Ahrens. A one-table method for sampling from continuous and discrete distributions. *Computing*, 5(2):127–146, June 1995.
- [2] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, 1986.
- [3] W. R. Gilks and P. Wild. Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics*, 41(2):337–348, 1992.
- [4] A. J. Kinderman and J. F. Monahan. Computer generation of random variables using the ratio of uniform deviates. *ACM Transactions on Mathematical Software*, 3(3):257–260, September 1977.
- [5] Josef Leydold. Automatic sampling with the ratio-of-uniforms method. *ACM Trans. Math. Softw.*, 26(1):78–98, 2000.
- [6] Josef Leydold. Short universal generators via generalized ratio-of-uniforms method. *Mathematics of Computation*, 72:1453–1471, 2003.
- [7] L. Martino and J. Míguez. An adaptive accept/reject sampling algorithm for posterior probability distributions. *Proc. of the IEEE SSP*, August 2009.
- [8] L. Martino and J. Míguez. A novel rejection sampling scheme for posterior probability distributions. *Proc. of the 34th IEEE ICASSP*, April 2009.
- [9] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.
- [10] J. C. Wakefield, A. E. Gelfand, and A. F. M. Smith. Efficient generation of random variates via the ratio-of-uniforms method. *Statistics and Computing*, 1(2):129–133, August 1991.

²The acceptance rate is the proportion of generated candidate samples which are actually accepted.