

PERFECTLY RANDOM SAMPLING OF TRUNCATED MULTINORMAL DISTRIBUTIONS

PEDRO J. FERNÁNDEZ,* *Fundação Getulio Vargas*

PABLO A. FERRARI,** *Universidade de São Paulo*

SEBASTIAN P. GRYNBERG,*** *Universidad de Buenos Aires*

Abstract

The target measure μ is the distribution of a random vector in a box \mathcal{B} , a Cartesian product of bounded intervals. The Gibbs sampler is a Markov chain with invariant measure μ . A ‘coupling from the past’ construction of the Gibbs sampler is used to show ergodicity of the dynamics and to perfectly simulate μ . An algorithm to sample vectors with multinormal distribution truncated to \mathcal{B} is then implemented.

Keywords: Perfect sampling; simulation; normal distribution

2000 Mathematics Subject Classification: Primary 65C35

Secondary 65C05

1. Introduction

The use of latent variables has several interesting applications in statistics, econometrics, and related fields like quantitative marketing. Models like tobit, probit, ordered probit, and multinomial probit are good examples. References and examples can be found in Geweke *et al.* (1997) and Allenby and Rossi (1999), especially for the multinomial probit model.

When the estimation process uses some simulation technique, in particular in Bayesian analysis, the need for drawing a sample from the distribution of the latent variable naturally arises. This procedure *augments* the observed data Y with a new variable X , which will be referred to as a *latent* variable. It is usually referred to as *data augmentation* (Tanner and Wong (1987), Tanner (1991, Chapter 5), Gelfand and Smith (1990)) and in other contexts as *imputation* (Rubin (1987, Chapter 5)).

The data augmentation algorithm is used when the likelihood function or posterior distribution of the parameter, given the latent data X , is simpler than the posterior, given the original observed data Y . If the distribution of the X variables is a multivariate normal in d dimensions and $Y = \mathbf{1}\{X \in \mathcal{B}\}$ is the indicator function of the set \mathcal{B} , the drawing is made from the normal distribution truncated to \mathcal{B} (and/or its complement \mathcal{B}^c).

The goal of this paper is to produce exact (or perfect) samples from random variables with distributions supported on a d -dimensional box \mathcal{B} ; the term *box* denotes the Cartesian product of d bounded intervals. We construct a discrete-time stationary Markov process $(\zeta_t, t \in \mathbb{Z})$ in the state space \mathcal{B} whose time-marginal distribution at any time t (that is, the law of ζ_t) has

Received 5 January 2007; revision received 25 September 2007.

* Postal address: Fundação Getulio Vargas, Av. Canal Marapendi 2915, Bloco. 1, Apto. 1502, Barra da Tijuca, Rio de Janeiro, 22631-050, Brazil.

** Postal address: Instituto de Matemática e Estatística, Universidade de São Paulo, Caixa Postal 66281, São Paulo, 05311-970, Brazil. Email address: pablo@ime.usp.br

*** Postal address: Departamento de Matemática, Facultad de Ingeniería, Universidad de Buenos Aires, Paseo Colón 850, Buenos Aires, 1063, Argentina.

a given density distribution g with support in \mathcal{B} . The construction is then implemented to perfectly simulate normal vectors of reasonable large dimension truncated to bounded boxes. The approach is also useful to show uniqueness of the invariant measure for a family of processes in a infinite-dimensional space $[a, b]^{\mathbb{Z}^d}$ with truncated Gaussian distributions and nearest-neighbor interactions (in preparation, see also Section 6).

The process is the Gibbs sampler popularized by Geman and Geman (1984) and used for the truncated normal case by Geweke (1991) and Robert (1995) in a Markov chain Monte Carlo algorithm to obtain samples with distribution approximating g . To describe the process in our setting, let $d \geq 2$, and take a box $\mathcal{B} = B(1) \times \dots \times B(d)$, where $B(k)$ are bounded intervals: $B(k) = [a_k, b_k] \subset \mathbb{R}$, $k = 1, \dots, d$. If $\mathbf{x} = (x_1, \dots, x_d)$ is the configuration at time $t - 1$ then at time t a site $\kappa(t)$ is chosen uniformly in $\{1, \dots, d\}$, say $\kappa(t) = k$, and x_k is substituted by a value y chosen with the density g conditioned to the values of the other coordinates; that is, with the density $g_k(\cdot | \mathbf{x}) : B(k) \rightarrow \mathbb{R}^+$ given by

$$g_k(y | \mathbf{x}) := \frac{g(\mathbf{y})}{\int_{B(k)} g(\mathbf{z}) \, d\mathbf{z}}, \quad y \in B(k), \tag{1}$$

where $k = \kappa(t)$, $\mathbf{y} = (x_1, \dots, x_{k-1}, y, x_{k+1}, \dots, x_d)$ and $\mathbf{z} = (x_1, \dots, x_{k-1}, z, x_{k+1}, \dots, x_d)$. Note that $g_k(y | \mathbf{x})$ does not depend on x_k . The distribution in \mathcal{B} with density g is a reversible measure for this dynamics.

We construct a stationary Gibbs sampler $(\zeta_t, t \in \mathbb{Z})$ as a function of a sequence $\underline{U} = (U_t, t \in \mathbb{Z})$ of independent, identically distributed (i.i.d.), and uniform-in-[0, 1] random variables and the updating schedule $\underline{\kappa} = (\kappa(t), t \in \mathbb{Z})$. We define (backwards) stopping times $\tau(t) \in (-\infty, t]$ such that $\{\tau(t) = s\}$ is measurable with respect to the field generated by the uniform random variables and the schedule in $[s + 1, t]$. The configuration at time t depends on the uniform random variables and the schedule in $[\tau(t) + 1, t]$.

In fact we simultaneously construct processes $(\zeta_{[s,t]}^\zeta, -\infty < s \leq t < \infty, \zeta \in \mathcal{B})$. For each fixed s and ζ , $(\zeta_{[s,t]}^\zeta, s \leq t < \infty)$ is the Gibbs sampler starting at time s with configuration ζ . For each fixed t , $(\zeta_{[s,t]}^\zeta, s \leq t, \zeta \in \mathcal{B})$ is a *maximal coupling*; see Thorisson (2000, Chapter 1) and Section 2. For each fixed time interval $[s, t]$, the process $(\zeta_{[s,t']}, s \leq t' \leq t)$ is a function of the uniform random variables, the schedule in $[s + 1, t]$, and the initial ζ . The crucial property of the coupling is that $\zeta_{\tau(t),t}^\zeta$ does not depend on ζ .

The construction is a particular implementation of the coupling-from-the-past algorithm of Propp and Wilson (1996) to obtain samples of a law ν as a function of a finite (but random) number of uniform random variables. For an annotated bibliography on the subject, see the web page of Wilson (1998). A Markov process having ν as a unique invariant measure is constructed as a function of the uniform random variables. The processes starting at all possible initial states run from negative time s to time 0 using the uniform random variables U_s, \dots, U_0 and a fixed updating schedule. If at time 0 all the realizations coincide then this common state has distribution ν . If they do not coincide then start the algorithm at time $s - 1$ using the random variables U_{s-1}, \dots, U_0 and continue this way up to the moment all realizations coincide at time 0; $\tau(0)$ is the maximal $s < 0$ such that this holds. The difficulty is that unless the dynamic is monotone, it is necessary to effectively couple all (noncountable) initial states. Murdoch and Green (1998) proposed various procedures to transform the infinite set of initial states into a more tractable finite subset.

In the normal case, for some intervals and covariance matrices, the maximal coupling works without further treatment. Møller (1999) studied the continuous state spaces when the process undergoes some kind of monotonicity; this is not the case here unless all correlations are

nonnegative. Philippe and Robert (2003) proposed an algorithm to perfectly simulate normal vectors conditioned to be positive using coupling from the past and slice sampling; this method is efficient in low dimensions.

Consider a d -dimensional random vector $Y \in \mathbb{R}^d$ with normal distribution of density

$$f(\mathbf{x}; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-Q(\mathbf{x}; \mu, \Sigma)/2}, \quad \mathbf{x} \in \mathbb{R}^d,$$

where the covariance matrix Σ is a positive-definite matrix, μ is the mean vector, and

$$Q(\mathbf{x}; \mu, \Sigma) = (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu),$$

where y' denotes the transpose of (a vector or matrix) y . In this case we say that $Y \sim \mathcal{N}(\mu, \Sigma)$. We denote by X the truncation of Y to the box \mathcal{B} ; its density is given by

$$g(\mathbf{x}; \mathcal{B}) = \frac{1}{Z(\mathcal{B})} f(\mathbf{x}) \mathbf{1}\{\mathbf{x} \in \mathcal{B}\},$$

where $Z(\mathcal{B}) = P(X \in \mathcal{B}) = \int_{\mathcal{B}} f(\mathbf{x}) \, d\mathbf{x}$ is the normalizing constant.

We have tested the algorithm in some examples. In particular, for the matrix $\Sigma^{-1} = \frac{1}{2}I + \frac{1}{2}\mathbf{1}\mathbf{1}'$, where I is the identity matrix and $\mathbf{1}' = (1, \dots, 1)$. For the box $[0, \frac{1}{2}]^d$ and for $d \geq 14$, our method is faster than the rejection method. The simulations indicate that the computer time for the rejection method grows exponentially with d while it grows as d^4 for our method. Using a simple program in MATLAB[®] on a desktop microcomputer, the method permits us to simulate up to dimension $d = 30$ for the interval $[0, \frac{1}{2}]^d$. For bigger boxes or boxes apart from the mean, for instance $\mathcal{B} = [\frac{1}{2}, 1]^d$, our method is less efficient, but in any case it is much faster than the rejection method. Our method makes use of the same one-dimensional uniform random numbers several times. Hence, for the sake of comparison, we have counted the number of times each number has been used. It is expected that if the box \mathcal{B} is sufficiently small and the covariance matrix has a ‘neighbor structure’, the number of steps may grow as $(d \ln d)^2$. In this case we should be able to simulate the state of an infinite-dimensional Gaussian field in a finite box.

Correlated normal vectors can be mapped into independent, standard normal vectors using the Cholesky transformation; see Devroye (1986, Chapter XI). In the case of truncated normals to a box, the transformation maps the box into a set. We may simulate standard normals and reject them if they do not belong to the transformed box. The difficulties are similar to those of the rejection method, i.e. as the dimension grows, the transformed set has small probability and the expected number of iterations grows exponentially with the dimension.

When the truncating set is not a box, our method generally fails as the coupling event for each coordinate has probability 0 in general. We illustrate this with an example in Section 6.

Another possibility, also discussed by Devroye (1986, p. 555), is to compute the marginal of the first coordinate, then the second marginal conditioned to the first one, and so on. The problem here is that the computation of the marginals conditioned to all the other coordinates in \mathcal{B} may be as complicated as the computation of the whole truncated vector.

In Section 2 we define coupling and maximal coupling. In Section 3 we describe the stationary theoretical construction of the Gibbs sampler and the properties of the coupling. Section 4 is devoted to the pseudocode of the perfect simulation algorithm. In Section 5 we compute the functions entering the algorithm for the truncated normal case. In Section 6 we give some examples and compare our perfect simulation algorithm with the rejection algorithm based on the uniform distribution.

2. Coupling

A *coupling* of a collection of random variables X_λ , $\lambda \in \Lambda$, with Λ a label set, is a random vector $(\hat{X}_\lambda : \lambda \in \Lambda)$ with the same marginals; that is, such that

$$\hat{X}_\lambda \stackrel{D}{=} X_\lambda, \quad \lambda \in \Lambda,$$

where ‘ $\stackrel{D}{=}$ ’ denotes equality in distribution. An event C is called a *coupling event* if the \hat{X}_λ s are identical in C , that is, if

$$C \subseteq \{\hat{X}_\lambda = \hat{X}_{\lambda'} \text{ for all } \lambda, \lambda' \in \Lambda\}.$$

We consider continuous random variables X_λ in \mathbb{R} with densities f_λ satisfying

$$P(C) \leq \int_{\mathbb{R}} \inf_{\lambda \in \Lambda} f_\lambda(y) dy \tag{2}$$

for any coupling event C . This is always true if Λ is countable and in the normal case treated here. When there is a coupling event C such that equality in (2) holds, the coupling is called *maximal*. See Thorisson (2000) for a complete treatment of coupling, including historical quotations.

A natural maximal coupling of these variables is the following. Let $G_\lambda(x) = \int_{-\infty}^x g_\lambda(y) dy$ be the corresponding cumulative distribution functions, and define

$$R(x) := \int_{-\infty}^x \inf_{\lambda \in \Lambda} g_\lambda(y) dy, \quad R := R(\infty).$$

Let $\hat{G}_\lambda(x) = G_\lambda(x) - R(x) + R$. Let U be a random variable uniformly distributed in $[0, 1]$, and let

$$\hat{X}_\lambda = R^{-1}(U) \mathbf{1}\{U \leq R\} + \hat{G}_\lambda^{-1}(U) \mathbf{1}\{U > R\},$$

where G^{-1} , the *generalized inverse* of G , is defined by

$$G(x) \geq y \iff x \geq G^{-1}(y).$$

The marginal distribution of \hat{X}_λ is G_λ :

$$P(\hat{X}_\lambda < x) = P(U < R(x)) + P(R < U < \hat{G}_\lambda(x)) = R(x) + \hat{G}_\lambda(x) - R = G_\lambda(x).$$

The process $(\hat{X}_\lambda, \lambda \in \Lambda)$ is a maximal coupling for the family $(X_\lambda, \lambda \in \Lambda)$. We use this coupling to construct the dynamics starting from different initial conditions. Then we show how to compute $R(x)$ and R in the normal case.

In the sequel we use a family of couplings. Let $(\Lambda_\ell, \ell \geq 1)$ be a decreasing sequence of parameter sets: $\Lambda_{\ell+1} \subset \Lambda_\ell$ for all $\ell \geq 1$. Let $R_{[0]}(x) \equiv 0$ and, for $\ell \geq 1$,

$$R_{[\ell]}(x) := \int_{-\infty}^x \inf_{\lambda \in \Lambda_\ell} g_\lambda(y) dy, \quad R_{[\ell]} := R_{[\ell]}(\infty).$$

Assume that $\lim_{\ell \rightarrow \infty} R_{[\ell]} = 1$. Define $D_{[\ell]}(x) = R_{[\ell]}(x) - R_{[\ell-1]}(x) + R_{[\ell-1]}$ and, for $\lambda \in \Lambda_\ell \setminus \Lambda_{\ell+1}$, define $\hat{G}_{\lambda, [\ell]}(x) = G_\lambda(x) - R_{[\ell]}(x) + R_{[\ell]}$ and

$$\hat{X}_\lambda = \sum_{i=1}^{\ell} D_{[i]}^{-1}(U) \mathbf{1}\{U \in [R_{i-1}, R_i]\} + \hat{G}_{\lambda, [\ell]}^{-1}(U) \mathbf{1}\{U > R_{[\ell]}\}.$$

Then $(\hat{X}_\lambda, \lambda \in \Lambda_\ell)$ is a maximal coupling of $(X_\lambda, \lambda \in \Lambda_\ell)$ for each ℓ .

3. A stationary construction of the Gibbs sampler

The goal is to define a stationary process $(\zeta_t, t \in \mathbb{Z})$ in \mathcal{B} with the Gibbs sampler dynamics associated to a density g in \mathcal{B} . The time marginal of the process at any time will be the distribution with density g . The Gibbs sampler is defined as follows. Assume that $\zeta_{t-1} \in \mathcal{B}$ is known, and then at time t choose a random site $\kappa(t)$ uniformly in $\{1, \dots, d\}$ and independently of ‘everything else’. Then set $\zeta_t(k) = \zeta_{t-1}(k)$ if $k \neq \kappa(t)$ and, for $k = \kappa(t)$, choose $\zeta_t(k)$ with the law g conditioned to the values of ζ_{t-1} at the other coordinates, i.e.

$$P(\zeta_t(k) \in A \mid \zeta_s, s < t) = \int_A g_k(y \mid \zeta_{t-1}) \, dy \quad \text{if } k = \kappa(t),$$

where $g_k(y \mid \mathbf{x})$, $y \in B(k) := [a_k, b_k]$, is the conditional density of the k th component of g given the other coordinates in (1). Reversibility follows from the identity

$$\int_{\mathcal{B}} \int_{B(k)} H_1(\mathbf{y}) H_2(\mathbf{x}) g_k(y \mid \mathbf{x}) \, dy \, g(\mathbf{x}) \, d\mathbf{x} = \int_{\mathcal{B}} \int_{B(k)} H_1(\mathbf{x}) H_2(\mathbf{y}) g_k(y \mid \mathbf{x}) \, dy \, g(\mathbf{x}) \, d\mathbf{x}$$

for any continuous functions H_1 and H_2 from \mathcal{B} to \mathbb{R} , where $\mathbf{x} = (x_1, \dots, x_d)$ and $\mathbf{y} = (x_1, \dots, x_{k-1}, y, x_{k+1}, \dots, x_d)$. In particular, the law with density g is invariant for the Gibbs sampler. It is indeed the unique invariant measure for the dynamics, as we show in Theorem 1.

To define a stationary version of the process we construct a family of couplings for each time interval $[s, t]$. The coupling starts with all possible configurations at time s and uses random variables associated to each time to decide the updating. All couplings use the same updating variables. To reflect the evolution of all possible configurations, we introduce a process on the set of boxes contained in \mathcal{B} . The initial configuration of this process is \mathcal{B} at time s , but at times $s' > s$ it can be (and will be as s decreases) reduced to a point. Let ξ be a d -tuple of closed bounded intervals of \mathbb{R} , $\xi(i) \subset B(i)$, $i = 1, \dots, d$. We abuse the notation by also calling ξ the box

$$\xi = \xi(1) \times \dots \times \xi(d) \subset \mathcal{B}.$$

Note that ξ has dimension less than d when $\xi(i)$ is just a point for some i . Define

$$R_k(x \mid \xi) = \int_{a_k}^x \min_{\mathbf{x} \in \xi} g_k(y \mid \mathbf{x}) \, dy, \quad x \in B(k), \tag{3}$$

$$R_k(\xi) = R_k(b_k \mid \xi). \tag{4}$$

These functions depend on ξ only through $(\xi_i, i \neq k)$. Here $R_k(\xi)$ is the probability that if the k th coordinate is updated when the other coordinates belong to the set ξ , then a coupled event is attained for the k th coordinate. This means that, for any configuration $\mathbf{x} \in \xi$, the updated value of the k th coordinate will be the same, say x ; its law is given by $R_k(x \mid \xi)$. In this case the updated interval $\xi(k)$ is also reduced to the point x . If the coupled event is not attained, $\xi(k)$ is kept as the interval $B(k)$.

The updating random variables consist of two families: $\underline{U} = (U_t : t \in \mathbb{Z})$, a family of independent variables with uniform distribution in $[0, 1]$; and $(\kappa(t) : t \in \mathbb{Z})$, a family of independent variables with uniform distribution in $\{1, \dots, d\}$ and independent of \underline{U} .

Now, for each $s \in \mathbb{Z}$, we define a process $(\eta_{[s,t]}, t \geq s)$, taking values on the boxes contained in \mathcal{B} as a function of $((U_t, \kappa(t)), t > s)$. The initial state is $\eta_{[s,s]} = \mathcal{B}$, and later each coordinate k is either a (random) point or the full interval $B(k)$. More precisely, for $s \in \mathbb{Z}$ and $x \in \mathbb{R}$, set

$$R_{[s,s]} = 0, \quad R_{[s,s]}(x) = 0, \quad D_{[s,s]}(x) = 0, \quad \text{and} \quad \eta_{[s,s]} = \mathcal{B}.$$

Fix $n \geq 1$, and assume that $R_{[s,t]}(x)$, $R_{[s,t]}$, and $\eta_{[s,t]}$ are defined if $0 \leq t - s \leq n - 1$. Then, for $t - s = n$ and $x \in B(\kappa(t))$, define

$$\begin{aligned}
 R_{[s,t]} &= R_{\kappa(t)}(\eta_{[s,t-1]}), \\
 R_{[s,t]}(x) &= R_{\kappa(t)}(x \mid \eta_{[s,t-1]}), \\
 D_{[s,t]}(x) &= R_{[s+1,t]} + R_{[s,t]}(x) - R_{[s+1,t]}(x), \\
 \eta_{[s,t]}(k) &= \begin{cases} \eta_{[s,t-1]}(k) & \text{if } k \neq \kappa(t), \\ \eta_{[s+1,t]}(k) \mathbf{1}\{U_t \leq R_{[s+1,t]}\} \\ + D_{[s,t]}^{-1}(U_t) \mathbf{1}\{R_{[s+1,t]} < U_t \leq R_{[s,t]}\} + B(k) \mathbf{1}\{U_t > R_{[s,t]}\} & \text{if } k = \kappa(t); \end{cases} \tag{5}
 \end{aligned}$$

in words, $k = \kappa(t)$ defines the coordinate to be updated at time t . Here $R_{[s,t]}$ is the probability that the coupling event is attained at coordinate $\kappa(t)$ for all the processes starting at times smaller than or equal to s . The coupling event is attained for the process starting at s when $U_t < R_{[s,t]}$. In case the coupling event is attained, the value of the k th coordinate is given by $D_{[s',t]}^{-1}(U_t)$ for s' given by the biggest $s \leq t$ such that $U_t < R_{[s,t]}$ (the second term in (5)). This value is the same for each $s \leq s'$ (the first term in (5)). When the coupling event is not attained (that is, for $s > s'$), the k th coordinate is kept equal to the full interval $B(k)$ (the third term in (5)). From this construction it follows that $(R_{[t-1,t]}(\cdot), R_{[t-1,t]}, D_{[t-1,t]})$ does not depend on \underline{U} and, for $s < t - 1$,

$$\begin{aligned}
 (R_{[s,t]}(\cdot), R_{[s,t]}, D_{[s,t]}) &\text{ is a function of } ((U_n, \kappa(n)), n = s + 1, \dots, t - 1), \tag{6} \\
 \eta_{[s,t]} &\text{ is a function of } (R_{[s,t]}(\cdot), R_{[s,t]}, D_{[s,t]}) \text{ and } (U_t, \kappa(t)).
 \end{aligned}$$

The k th coordinate of $\eta_{[s,t]}$ is either the interval $B(k)$ or a point. The process is monotone in the following sense:

$$\eta_{[s-1,t]} \subset \eta_{[s,t]} \quad \text{for all } s < t.$$

In particular, if some coordinate is a point at time t for the process starting at s , then it will be the *same* point for the processes starting at previous times, i.e.

$$\text{if } \eta_{[s,t]}(k) \text{ is a point then } \eta_{[s-1,t]}(k) = \eta_{[s,t]}(k) \quad \text{for all } s < t.$$

For each time t , define

$$\tau(t) = \max\{s < t : \eta_{[s,t]}(k) \text{ is a point for all } k\}.$$

Using (6), we conclude that $\tau(t) + 1$ is a stopping time for the filtration generated by $((U_{t-n}, \kappa(t-n)), n \geq 0)$; the event $\{\tau(t) = s\}$ is a function of $((U_n, \kappa(n)), n = s + 1, \dots, t)$.

Assume that $\tau(t) > -\infty$ almost surely for all t , and define the process $(\zeta_t, t \in \mathbb{Z})$ in \mathcal{B} by

$$\zeta_t := \eta_{[\tau(t),t]} \in \mathcal{B}.$$

Noting that $\tau(t) > -\infty$ is equivalent to $R_{[n,t]} \rightarrow 1$ as $n \rightarrow -\infty$, we obtain the following explicit expression for ζ_t :

$$\zeta_t(k) = \begin{cases} \zeta_{t-1}(k) & \text{if } k \neq \kappa(t), \\ \sum_{n \leq t-1} D_{[n,t]}^{-1}(U_t) \mathbf{1}\{R_{[n+1,t]} < U_t \leq R_{[n,t]}\} & \text{if } k = \kappa(t). \end{cases} \tag{7}$$

For each $s < t$ and $\zeta \in \mathcal{B}$, we now construct the process $(\zeta_{[s,t]}^\zeta, t \geq s)$, the Gibbs sampler starting with ζ at time s . Let $G_k(x | \mathbf{x}) = \int_{a_k}^x g_k(z | \mathbf{x}) dz$ and

$$\hat{G}_{[s,t]}(x | \zeta) = R_{[s,t]} + G_{\kappa(t)}(x | \zeta) - R_{[s,t]}(x).$$

For each $\zeta \in \mathcal{B}$, define $\zeta_{[s,s]}^\zeta = \zeta$ and, for $t > s$,

$$\zeta_{[s,t]}^\zeta(k) = \begin{cases} \zeta_{[s,t-1]}^\zeta(k) & \text{if } k \neq \kappa(t), \\ \sum_{n=s}^{t-1} \mathbf{1}\{U_t \in [R_{[n+1,t]}, R_{[n,t]}]\} D_{[n+1,t]}^{-1}(U_t) \\ \quad + \mathbf{1}\{U_t > R_{[s,t]}\} \hat{G}_{[s,t]}^{-1}(U_t | \zeta_{[s,t-1]}^\zeta) & \text{if } k = \kappa(t). \end{cases}$$

For each fixed t and k , $((\zeta_{[s,t]}^\zeta(k), \zeta \in \mathcal{B}), s \leq t)$ is a maximal coupling among the processes starting with all possible configurations ζ at all times $s \leq t$.

Theorem 1. Assume that $\tau(t) > -\infty$ almost surely for all $t \in \mathbb{Z}$. Then the process $(\zeta_t, t \in \mathbb{Z})$, defined in (7), is a stationary Gibbs sampler related to g . The distribution of ζ_t at each time t is the distribution with density g in \mathcal{B} . This distribution is the unique invariant measure for the process and

$$\sup_{\zeta} \mathbb{P}(\zeta_{[s,t]}^\zeta \neq \zeta_t) \leq \mathbb{P}(\tau(t) < s), \tag{8}$$

where $\zeta_{[s,t]}^\zeta$ is the process starting with configuration ζ at time s .

Proof. Stationarity follows from the construction. Indeed, for all $s \leq t$ and $\ell \in \mathbb{Z}$, $\eta_{[s,t]}$ as a function of $((U_n, \kappa(n)), n \in \mathbb{Z})$ is identical to $\eta_{[s+\ell,t+\ell]}$ as a function of $((U_{n+\ell}, \kappa(n+\ell)), n \in \mathbb{Z})$.

The function $\Psi_t: [0, 1] \times [0, 1]^{t-1} \rightarrow B(\kappa(t))$, defined in (7) by

$$\Psi_t(U; U_{t-1}, U_{t-2}, \dots) = \sum_{n \leq t-1} D_{[n,t]}^{-1}(U) \mathbf{1}\{R_{[n+1,t]} < U \leq R_{[n,t]}\},$$

satisfies, for $k = \kappa(t)$,

$$\mathbb{P}(\Psi_t(U; U_{t-1}, U_{t-2}, \dots) \in A) = \int_A g_k(y | \zeta_{t-1}) dy,$$

where we have used the fact that ζ_{t-1} is a function of U_{t-1}, U_{t-2}, \dots and $\kappa_{t-1}, \kappa_{t-2}, \dots$. This is sufficient to show that ζ_t evolves according to the Gibbs sampler.

To complete the proof we need to show that the distribution with density g is the marginal law of ζ_t . Since the updating is performed with the conditioned distribution, the distribution with density g is invariant for the Gibbs sampler.

Since $\zeta_{[s,t]}^\zeta = \zeta_{[s,t]}$, if $s \leq \tau(t)$ then (8) follows. This also implies that $\lim_{s \rightarrow -\infty} \zeta_{[s,t]}^\zeta = \zeta_t$ almost surely. In particular, $\zeta_{[s,t]}^\zeta$ converges weakly to the law of ζ_t as s tends to ∞ . The same is true for $\lim_{t \rightarrow \infty} \zeta_{[s,t]}^\zeta$. Taking ζ random with law g , this proves that ζ_t must have law g . Then taking ζ with law \tilde{g} invariant for the dynamics, we conclude that $\tilde{g} = g$ and the uniqueness of the invariant measure follows.

Lemma 1. *A sufficient condition for $\tau(t) > -\infty$ almost surely is that $R_k(\mathcal{B}) > 0$ for all k . In this case $P(\tau(t) < s)$ decays exponentially fast with $t - s$.*

Proof. We have $\tau(t) > \tau^0(t) := \max\{s < t - d : \kappa(s+n) = n \text{ and } U_{s+n} < R_{\kappa(s+n)}(\eta), n = 1, \dots, d\}$, the last time in the past the coordinates have been successively updated to a point independently of the other coordinates. At $\tau^0(t) + d$ and successive times $t' \geq \tau^0(t) + d$, each coordinate of $\eta_{[\tau^0(t), t']}$ has been reduced to a point. For all t , $\tau_0(t)$ is finite because the event $\{\kappa(s+n) = n \text{ and } U_{s+n} < R_{\kappa(s+n)}(\eta), n = 1, \dots, d\}$ occurs for infinitely many values of s with probability 1. The same argument shows the exponential decay of the tail of $\tau(t)$.

Remark. The velocity of convergence of the Gibbs sampler to equilibrium depends on the values $(R_k(\mathcal{B}), k = 1, \dots, d)$. In turn, these values depend on the size of the box \mathcal{B} and the correlations of the distribution g . Strongly correlated vectors produce small values $R_k(\mathcal{B})$ and, hence, slow convergence to equilibrium. The efficiency of the algorithms discussed in the next sections will also depend on these values.

4. The perfect simulation algorithm

The construction of Section 3 is implemented in a perfect simulation algorithm. Let ξ be a box contained in \mathcal{B} . For each $k = 1, \dots, d$, let $R_k(\cdot \mid \xi)$ and $R_k(\xi) = R_k(b_k \mid \xi)$ be as defined in (3) and (4), respectively. When $\xi = \mathcal{B}$, we denote by R_k the value of $R_k(\mathcal{B})$. Let $\kappa(t)$ be the updating schedule. It can be either a family of i.i.d. random variables chosen uniformly in $\{1, \dots, d\}$ or the periodic sequence $\kappa(t) = [t - 1]_{\text{mod } d} + 1$. For each pair η, ξ of boxes contained in \mathcal{B} and $1 \leq k \leq d$, let $D_{k,\eta,\xi} : \mathcal{B}(k) \rightarrow [0, 1]$ be the function defined by

$$D_{k,\eta,\xi}(x) = R_k(x \mid \xi) - R_k(x \mid \eta) + R_k(\eta), \quad x \in \mathcal{B}(k).$$

Let ξ_1 and ξ_2 be boxes contained in \mathcal{B} , let $u \in [0, 1]$, and let $1 \leq k \leq d$. The auxiliary coupler function ϕ is defined as follows.

```
function  $\phi(\xi_1, \xi_2, u, k)$ :
 $\xi \leftarrow \xi_1$ 
 $\xi(k) \leftarrow \xi_2(k)$ 
if  $\xi_2(k)$  is not a point in  $\mathbb{R}$  and  $u \leq R_k(\xi)$ ,
    update the  $k$ th coordinate of  $\xi$ :
     $\xi(k) \leftarrow D_{k,\xi_2,\xi}^{-1}(u)$ 
end
 $\xi_2 \leftarrow \xi$ 
Return ( $\xi_2$ )
```

For the perfect simulation algorithm, perform the following steps.

```
 $T \leftarrow 0$ 
start ( $\eta_0$ )
while  $\eta_0$  is not a point in  $\mathcal{B}$ 
     $T \leftarrow T - 1$ 
    start ( $\eta_T$ )
     $t \leftarrow T$ 
    while  $t < 0$  update  $\eta_{t+1}$  using the coupler function  $\phi$ :
         $\eta_{t+1} \leftarrow \phi(\eta_t, \eta_{t+1}, U_{t+1}, \kappa(t + 1))$ 
```

```

        t ← t + 1
    end
end
Return (η0)

For each T ≤ 0, ‘start(ηT)’ is the following subroutine.

start (ηT) :
ηT ← ℬ
Generate UT, a uniform random variable in [0, 1]
if UT ≤ Rκ(T), compute the κ(T)th coordinate of ηT :
    ηT(κ(T)) ← Rκ(T)-1(UT | ηT)
end
    
```

5. Truncated normal distributions

In this section we implement the construction in the normal case. We start with elementary facts of the one-dimensional normal distribution.

5.1. One dimension

Let $\phi(x) = (1/\sqrt{2\pi}) e^{-x^2/2}$ be the standard normal density in \mathbb{R} and $\Phi(x) = \int_{-\infty}^x \phi(y) dy$ be the corresponding distribution function. Let $a < b$ be real numbers. A random variable X with mean μ and variance σ^2 has a *truncated normal distribution* to the interval $[a, b]$ if its density is given by

$$g_{a,b}(x; \mu, \sigma^2) = \frac{1}{\sigma} \frac{\phi((x - \mu)/\sigma)}{\Phi((b - \mu)/\sigma) - \Phi((a - \mu)/\sigma)}, \quad a \leq x \leq b,$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. This distribution is called $\mathcal{T}_{a,b}\mathcal{N}(\mu, \sigma^2)$. The truncated normal is the law $\mathcal{N}(\mu, \sigma^2)$ conditioned to $[a, b]$. Let

$$A_{a,b}(\mu) = \Phi(b - \mu) - \Phi(a - \mu).$$

Let $X_\mu, \mu \in I = [\mu^-, \mu^+]$, be a family of normal distributions truncated to the interval $[a, b]$, i.e. $X_\mu \sim \mathcal{T}_{a,b}\mathcal{N}(\mu, \sigma^2)$. Let $A^\pm = A_{a,b}^\pm$ and $x(I, \sigma) = x_{a,b}(I, \sigma)$ be defined by

$$A^- = \Phi\left(\frac{b - \mu^-}{\sigma}\right) - \Phi\left(\frac{a - \mu^-}{\sigma}\right), \quad A^+ = \Phi\left(\frac{b - \mu^+}{\sigma}\right) - \Phi\left(\frac{a - \mu^+}{\sigma}\right),$$

and

$$x(I, \sigma) = \frac{\mu^- + \mu^+}{2} - \frac{\sigma^2}{\mu^+ - \mu^-} \ln\left(\frac{A^-}{A^+}\right).$$

Let

$$R(x | I) = \int_a^x \inf_{\mu \in I} g_{a,b}(y; \mu, \sigma^2) dy, \quad x \in [a, b], \tag{9}$$

$$R(I) = R(b | I).$$

Proposition 1, below, proven later, gives an explicit way of computing the functions participating in the multicoupling. It states that the infimum of normal densities with the same variances σ and means in the interval $I = [\mu^-, \mu^+]$ coincides with the normal with mean μ^+ up to $x(I, \sigma)$ and with the normal with mean μ^- from this point on.

Proposition 1. *The integrand in (9) has the following explicit expression:*

$$\inf_{\mu \in I} g_{a,b}(x; \mu, \sigma^2) = g_{a,b}(x; \mu^+, \sigma^2) \mathbf{1}\{x < x(I, \sigma)\} + g_{a,b}(x; \mu^-, \sigma^2) \mathbf{1}\{x \geq x(I, \sigma)\}. \tag{10}$$

As a consequence,

$$\begin{aligned} R(x | I) &= \frac{1}{A^+} \left(\Phi\left(\frac{x - \mu^+}{\sigma}\right) - \Phi\left(\frac{a - \mu^+}{\sigma}\right) \right) \mathbf{1}\{x < x(I, \sigma)\} \\ &+ \frac{1}{A^+} \left(\Phi\left(\frac{x(I, \sigma) - \mu^+}{\sigma}\right) - \Phi\left(\frac{a - \mu^+}{\sigma}\right) \right) \mathbf{1}\{x \geq x(I, \sigma)\} \\ &+ \frac{1}{A^-} \left(\Phi\left(\frac{x - \mu^-}{\sigma}\right) - \Phi\left(\frac{x(I, \sigma) - \mu^-}{\sigma}\right) \right) \mathbf{1}\{x \geq x(I, \sigma)\}. \end{aligned}$$

In particular,

$$\begin{aligned} R(I) &= \frac{1}{A^+} \left(\Phi\left(\frac{x(I, \sigma) - \mu^+}{\sigma}\right) - \Phi\left(\frac{a - \mu^+}{\sigma}\right) \right) \\ &+ \frac{1}{A^-} \left(\Phi\left(\frac{b - \mu^-}{\sigma}\right) - \Phi\left(\frac{x(I, \sigma) - \mu^-}{\sigma}\right) \right), \end{aligned}$$

and, for each $u \in [0, R(I)]$, there exists a unique $x_u \in [a, b]$ such that $R(x_u | I) = u$.

Remark. We have $0 < R(I) < 1$. Indeed, $R(I) > 0$ because $\inf_{\mu \in I} g_{a,b}(x; \mu, \sigma^2) > 0$. If $R(I) = 1$ then all densities coincide unless $\mu^- = \mu^+$; by hypothesis this trivial case is excluded.

5.2. The multivariate normal

Let $Y \sim \mathcal{N}(\mu, \Sigma)$, let $X = Y|_{\mathcal{B}}$ be the vector conditioned to \mathcal{B} , and let $\mathbf{x} \in \mathcal{B}$. Let $\Sigma^{-1} = (a_{ij})$, then the law of X_k conditioned to $(X_i, i \neq k) = (x_i, i \neq k)$ is the truncated normal $\mathcal{T}_{a_k, b_k} \mathcal{N}(\hat{\mu}_k(\mathbf{x}), \hat{\sigma}_k^2)$, where

$$\hat{\mu}_k(\mathbf{x}) = \mu_k - \frac{1}{a_{kk}} \sum_{i \neq k} a_{ki}(x_i - \mu_i), \quad \hat{\sigma}_k^2 = \frac{1}{a_{kk}}.$$

We see that $\hat{\mu}_k(\mathbf{x})$ does not depend on x_k and

$$\hat{\mu}_k(\mathbf{x}) \in [\mu_k^-(\mathbf{x}), \mu_k^+(\mathbf{x})],$$

where

$$\begin{aligned} \mu_k^-(\mathbf{x}) &= \frac{1}{a_{kk}} \sum_{i=1}^d a_{ki} \mu_i - \frac{1}{a_{kk}} \left(\sum_{a_{ki} < 0, i \neq k} a_{ki} a_i + \sum_{a_{ki} \geq 0, i \neq k} a_{ki} b_i \right), \\ \mu_k^+(\mathbf{x}) &= \frac{1}{a_{kk}} \sum_{i=1}^d a_{ki} \mu_i - \frac{1}{a_{kk}} \left(\sum_{a_{ki} < 0, i \neq k} a_{ki} b_i + \sum_{a_{ki} \geq 0, i \neq k} a_{ki} a_i \right). \end{aligned}$$

Remark. Since $R_k(\mathcal{B}) > 0$ for all k , $\tau(t) > -\infty$ almost surely for all $t \in \mathbb{Z}$. This implies that the truncated multivariate normal case falls under the hypothesis of Theorem 1 and our algorithm works.

5.3. Back to one dimension. The minimum of truncated normals

We prove Proposition 1. To simplify the notation write $g_{a,b}(x; \mu)$ instead of $g_{a,b}(x; \mu, 1)$. Observing that

$$g_{a,b}(x; \mu, \sigma^2) = \frac{1}{\sigma} g_{a/\sigma, b/\sigma} \left(\frac{x}{\sigma}; \frac{\mu}{\sigma} \right),$$

it is sufficient to prove (10) for $\sigma = 1$. The proof is based on the following elementary lemmas.

Lemma 2. *Let $\mu_1 < \mu_2$. Then*

$$\min(g_{a,b}(x; \mu_1), g_{a,b}(x; \mu_2)) = g_{a,b}(x; \mu_1) \mathbf{1}\{x(\mu_1, \mu_2) \leq x\} + g_{a,b}(x; \mu_2) \mathbf{1}\{x(\mu_1, \mu_2) > x\},$$

where

$$x(\mu_1, \mu_2) = \frac{\mu_1 + \mu_2}{2} - \frac{1}{\mu_2 - \mu_1} \ln \left(\frac{A_{a,b}(\mu_1)}{A_{a,b}(\mu_2)} \right).$$

Lemma 3. *For all $\mu \in (\mu^-, \mu^+)$, it holds that*

$$\frac{\mu + \mu^+}{2} - \frac{1}{\mu^+ - \mu} \ln \left(\frac{A_{a,b}(\mu)}{A_{a,b}(\mu^+)} \right) > \frac{\mu^- + \mu}{2} - \frac{1}{\mu - \mu^-} \ln \left(\frac{A_{a,b}(\mu^-)}{A_{a,b}(\mu)} \right).$$

Proof of Proposition 1. Let $\mu \in (\mu^-, \mu^+)$. By Lemma 2 we have

$$\begin{aligned} g_{a,b}(x; \mu^-) \leq g_{a,b}(x; \mu) &\iff \frac{\mu^- + \mu}{2} - \frac{1}{\mu - \mu^-} \ln \left(\frac{A_{a,b}(\mu^-)}{A_{a,b}(\mu)} \right) \leq x, \\ g_{a,b}(x; \mu) \leq g_{a,b}(x; \mu^+) &\iff \frac{\mu + \mu^+}{2} - \frac{1}{\mu^+ - \mu} \ln \left(\frac{A_{a,b}(\mu)}{A_{a,b}(\mu^+)} \right) \leq x. \end{aligned}$$

Hence, $g_{a,b}(x; \mu) \leq \min(g_{a,b}(x; \mu^-), g_{a,b}(x; \mu^+))$ if and only if

$$x \in \left[\frac{\mu + \mu^+}{2} - \frac{1}{\mu^+ - \mu} \ln \left(\frac{A_{a,b}(\mu)}{A_{a,b}(\mu^+)} \right), \frac{\mu^- + \mu}{2} - \frac{1}{\mu - \mu^-} \ln \left(\frac{A_{a,b}(\mu^-)}{A_{a,b}(\mu)} \right) \right]. \tag{11}$$

But, by Lemma 3 the interval in (11) is empty for all $\mu \in (\mu^-, \mu^+)$. This implies that

$$\min(g_{a,b}(x; \mu^-), g_{a,b}(x; \mu^+)) < g_{a,b}(x; \mu),$$

from where

$$\inf_{\mu \in I} g_{a,b}(x; \mu, \sigma^2) = \min(g_{a,b}(x; \mu^-, \sigma^2), g_{a,b}(x; \mu^+, \sigma^2))$$

holds for $\sigma = 1$. The corresponding identity in (10) follows by applying Lemma 2 to $\mu_1 = \mu^-$ and $\mu_2 = \mu^+$.

6. Comparisons and comments on other methods

In this section we compare our method with the rejection method in some examples. The conclusion is that the rejection method may be better than ours in dimension two for some regions; but, ours becomes better and better as the dimension increases. Then we show why the method does not work when the region is not a box. This discards the following tempting

approach: multiply the target vector by a matrix to obtain a vector with i.i.d. coordinates. The transformed vector is much easier to simulate; the problem is that it is now conditioned to a transformed region. When the region is not a box, the conditioned vector is not an i.i.d. vector and a coupling must be performed. We show here that, in general, the corresponding coupling event has probability 0.

6.1. The rejection method

We compare our algorithm with the following rejection algorithm: simulate a uniformly distributed vector (\mathbf{x}, u) in $\mathcal{B} \times [0, \max_{\mathbf{x} \in \mathcal{B}} g(\mathbf{x})]$. If $u < g(\mathbf{x})$ then accept \mathbf{x} . We use $d + 1$ one-dimensional uniform random variables for each attempt of the rejection algorithm. Let N denote the expected number of one-dimensional uniform random variables generated by our perfect simulation algorithm, each counted the number of times that it is used. Let M denote the expected number of one-dimensional uniform random variables needed by the rejection algorithm to produce an accepted value. We divide the experiments into two parts, $d = 2$ and $d \geq 2$.

For the case in which $d = 2$, let $\mu = (0, 0)$ and

$$\Sigma = \begin{pmatrix} 1 & \frac{12}{5} \\ \frac{12}{5} & 9 \end{pmatrix}.$$

We consider the following two cases:

(i) boxes

$$\mathcal{B}(x_1, x_2) = [0, 1] \times [0, 1] + (x_1, x_2),$$

where $-4 \leq x_1 \leq 4$ and $0 \leq x_2 \leq 4$; and

(ii) boxes

$$\mathcal{B}(r) = [0, r] \times [0, r] \quad (\text{type 1}) \quad \text{and} \quad \mathcal{B}(r) = [-r, 0] \times [0, r] \quad (\text{type 2}),$$

where $r = \frac{1}{2}, 1, 2, 3, 4, 5, 6$.

The results are shown in Table 1 for (i) and in Figure 1 for (ii).

TABLE 1: The values of $M(x_1, x_2)$ and the mean number of one-dimensional uniform random variables used by the rejection method for the $d = 2$, case-(i) boxes. The corresponding number for our method, $N(x_1, x_2) \in [3.2, 3.3]$, is better for all boxes.

x_1	x_2				
	0	1	2	3	4
-4	65	76	88	100	112
-3	37	45	55	65	77
-2	16	22	28	36	45
-1	5	8	11	15	21
0	4	4	4	5	7
1	12	8	6	5	4
2	29	22	17	12	9
3	55	45	37	30	23
4	86	75	64	54	45

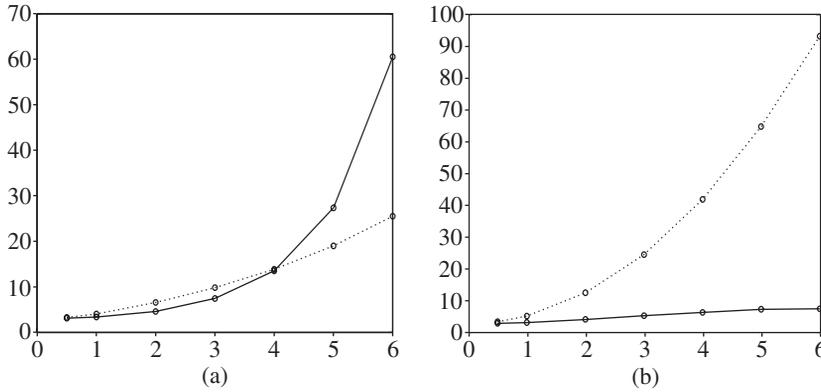


FIGURE 1: Plots of $(r, N(r))$ (solid line) and $(r, M(r))$ (dotted line) for $d = 2$: (a) case-(ii), type-1 boxes and (b) case-(ii), type-2 boxes. For type-2 boxes, our method is faster than the rejection method; but, for type-1 boxes, this is true when $r \leq 4$.

Now let us consider the case in which $d \geq 2$.

Example 1. Fix $\mu = (0, \dots, 0)$ and

$$\Sigma^{-1} = \frac{1}{2}I + \frac{1}{2}\mathbf{1}\mathbf{1}',$$

where I is the identity matrix and $\mathbf{1}' = (1, \dots, 1)$. We consider the following two cases:

- (i) boxes $\mathcal{B} = [0, \frac{1}{2}]^d$, $d = 2, \dots, 29$; and
- (ii) boxes $\mathcal{B} = [\frac{1}{2}, 1]^d$, $d = 2, \dots, 14$.

The results are shown in Table 2 for (ii) and in Figure 2 for (i).

TABLE 2: The values of N , the mean number of uniforms used by our method, and M , the mean number of uniforms used by the rejection method for the $d \geq 2$, case-(ii) boxes.

d	N	M
2	3	5
3	6	10
4	11	22
5	17	51
6	27	133
7	36	372
8	48	1 141
9	60	3 810
10	84	13 736
11	106	53 356
12	147	223 520
13	164	981 460
14	208	4 858 138

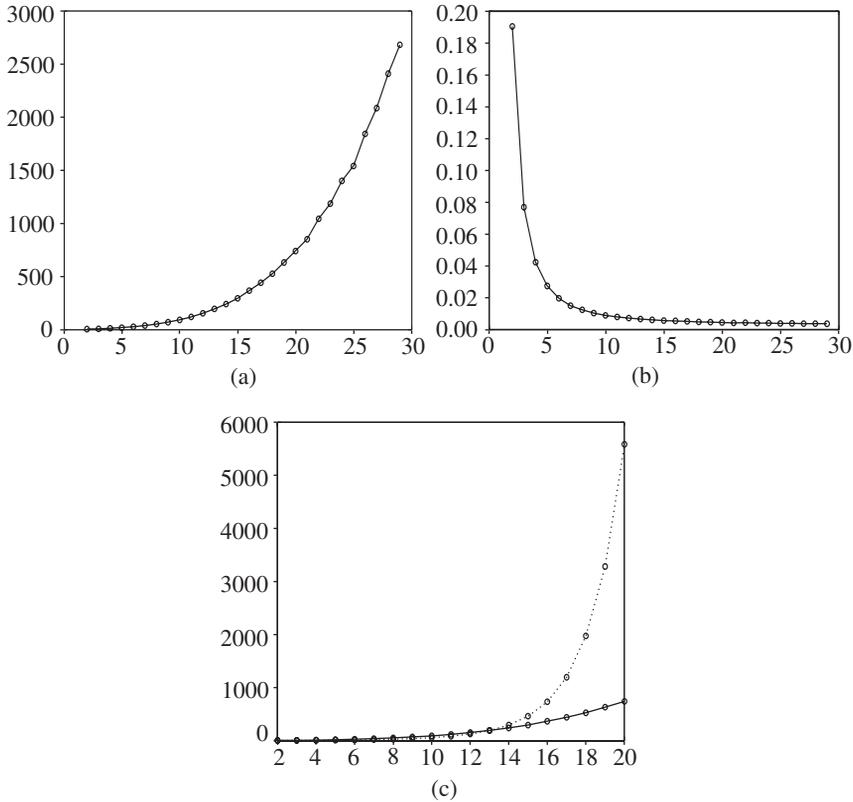


FIGURE 2: Plots of (a) $(d, N(d))$, (b) $(d, N(d)/d^4)$, and (c) $(d, N(d))$ (solid line) and $(d, M(d))$ (dotted line) for the $d \geq 2$, case-(i) boxes. It seems that $N(d)$ grows as $o(d^4)$ and $M(d)$ as $(d + 1)e^{ad}$ for some $a > 0$. For $d \geq 14$, our method is faster than the rejection method.

Example 2. Let $\mu = (0, \dots, 0)$ and Σ^{-1} be the matrix, with ‘neighbor structure’,

$$\Sigma_{i,j}^{-1} = \rho^{|j-i|} \mathbf{1}\{|j - i| \leq 1\},$$

where $\rho = \frac{1}{2}$. We consider boxes of the form $\mathcal{B} = [0, 1]^d$. The results are shown in Figure 3.

6.1.1. *Strongly correlated variables.* The algorithm slows down very fast with the dimension when the variables are strongly correlated. An Associate Editor proposes to consider a Gaussian vector truncated to the box $I^d = [0, 1]^d$, with mean $\mu = 0$ and covariance matrix

$$\Sigma = \varepsilon I + (1 - \varepsilon) \mathbf{1}\mathbf{1}',$$

for small values of ε . As a measure of the speed of the algorithm, we have computed the coefficient R , shown in Table 3, as a function of ε and the dimension d . For the case in which $d = 2$, the coupling time $\tau(0)$ for this example is a geometric with mean $1/R$, because it suffices that one of the uniforms is smaller than R to get the coupling in the next step. In general, as explained in Lemma 1, $\tau(0) \leq \tau^0(0)$, whose expectation is of the order $(1/R)^{d-1}$.

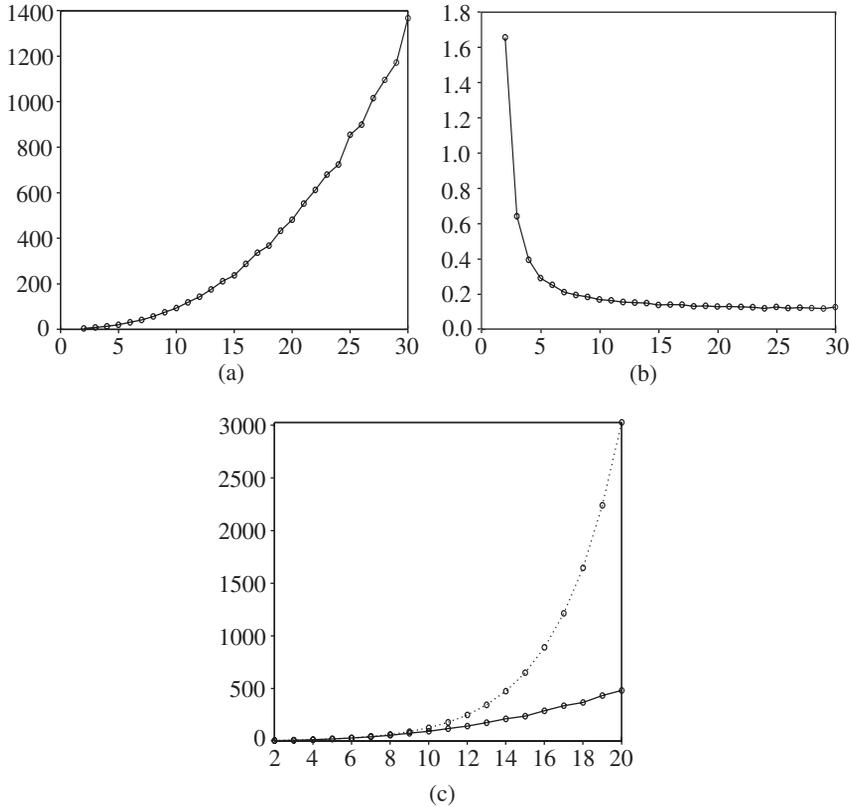


FIGURE 3: Plots of (a) $(d, N(d))$, (b) $(d, N(d)/(d \log d)^2)$, and (c) $(d, N(d))$ (solid line) and $(d, M(d))$ (dotted line) for the $d \geq 2$, Example 2 boxes. It seems that $N(d)$ grows as $(d \log d)^2$ and $M(d)$ as $(d + 1)e^{ad}$ for some $a > 0$. Our method is faster than the rejection method for all d .

TABLE 3: Coefficient R for $\mu = \mathbf{0}$, $\Sigma = \varepsilon \mathbf{I} + (1 - \varepsilon) \mathbf{1} \mathbf{1}'$, where \mathbf{I} is the identity matrix and $\mathbf{1}' = (1, \dots, 1)$, and the box $[0, 1]^d$.

d	ε	
	0.10	0.01
2	0.5139	0.8753×10^{-3}
4	0.3446	0.3121×10^{-4}
8	0.2792	0.5969×10^{-5}
16	0.2507	0.2615×10^{-5}
32	0.2375	0.1731×10^{-5}

6.1.2. *Other regions.* Let $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ be a d -dimensional multinormal random vector with zero mean and nonsingular covariance matrix Σ . Let $\mathcal{B} = [a_1, b_1] \times \dots \times [a_d, b_d]$. If we rotate and scale the coordinates to obtain i.i.d. $\mathcal{N}(0, 1)$ s, i.e. if we set $\mathbf{X} = \Sigma^{-1/2} \mathbf{Y}$, then the box constraints are of the form $\mathbf{a} \leq \Sigma^{1/2} \mathbf{x} \leq \mathbf{b}$, where $\mathbf{a} = (a_1, \dots, a_d)$ and $\mathbf{b} = (b_1, \dots, b_d)$. When the constraints are of this form, the conditional distribution of $X_k \mid \mathbf{X}^k = \mathbf{x}^k$ has a

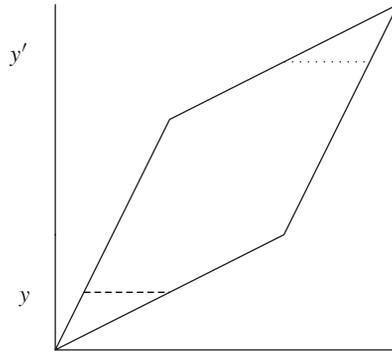


FIGURE 4: In a parallelogram region, the minimal value $\inf_x g_{X_1|X=x}(x_1)$ is 0. The truncated distributions conditioned to the second coordinate being y and y' respectively have disjoint supports \mathcal{F}^y and $\mathcal{F}^{y'}$, which are the projections on the horizontal axis of the dashed and dotted segments, respectively.

probability density function whose support depends on the condition \mathbf{x}^k . In such cases, the maximal coupling in our approach may have a coupling event with zero probability. Indeed, it could happen that

$$\inf_x g_{X_k|X=x}(x_k) = 0.$$

This implies that $R_k(x_k) = \int_{-\infty}^{x_k} \inf_x g_{X_k|X=x}(y) dy = 0$ for all x_k .

To see this for the case in which $d = 2$, observe that the transformation takes the rectangular box into a parallelogram \mathcal{F} with sides not in general parallel to the coordinate axes. We need to simulate the standard bidimensional normal \mathbf{X} conditioned to \mathcal{F} . It is still true that the law of X_1 conditioned on $X_2 = y$ is normal truncated to the slice of the parallelogram $\mathcal{F}^y := \{(x_1, x_2) \in \mathcal{F} : x_2 = y\}$. It is a matter of geometry that there are different values of y and y' such that $\mathcal{F}^y \cap \mathcal{F}^{y'} = \emptyset$; this implies that the infimum above is 0.

Take, for instance, $\mu = (0, 0)$,

$$\Sigma = \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix},$$

and \mathbf{Y} conditioned to the box $\mathcal{B} = [0, 1] \times [0, 1]$. Then

$$\Sigma^{-1/2} = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

and the transformed box is a parallelogram \mathcal{F} with vertices $(0, 0)$, $(\frac{2}{3}, \frac{1}{2})$, $(\frac{1}{3}, \frac{2}{3})$, and $(1, 1)$, as shown in Figure 4.

7. Final remarks

We have proposed a construction (simulation) of a multidimensional random vector conditioned to a bounded box using a version of the coupling-from-the-past algorithm of the Gibbs sampler dynamics. In the case of normal distributions, we have given an explicit construction, taking advantage of the bell shape of the density and the fact that the distribution of one coordinate, given the others, is a truncated normal with variance independent of the other coordinates. This allows us to define the coupling set C as a function of the extreme possible

values of the mean and to obtain positive probability for each coupling set C as soon as the box is bounded.

One of the consequences of the approach is to show that the Gibbs sampler is ergodic in the box if the coupling event has probability uniformly bounded below in each one of the coordinates. This corresponds to the condition $R_k(\mathcal{B}) > 0$ for all k th coordinates. When the uniform random variable used to update site k at some time falls below $R_k(\mathcal{B}) > 0$, this coordinate coincides for all processes.

The method can be used to show ergodicity of the process in an infinite volume when the inverse of the covariance matrix has a neighbor structure, for instance for tridiagonal matrices. As an example, consider the formal density in $[a, b]^{\mathbb{Z}^d}$

$$g(\mathbf{x}) \sim \exp\left(-\beta \sum_{|i-j|=1} (x_i - x_j)^2\right),$$

which corresponds to an infinite-volume Gaussian field, each coordinate truncated to $[a, b] \subset \mathbb{R}$; $\beta > 0$ is a parameter (known as the inverse temperature in statistical mechanics). The Gibbs sampler can be performed in continuous time, and the coupling can be carried out as in the finite case. However, the updating of site k depends only on the configurations in sites j such that $|j - k| = 1$. Each site gets a definite value (and not an interval) when the corresponding uniform random variable falls below a certain value $R = R(\beta, [a, b])$ (corresponding to $R_k(\mathcal{B})$, but now it is constant in the coordinates). To determine the value of site k at time t , we explore the process backwards; calling U the random variable used for the last time before t , say t' , to update site k , there are two cases to explore, $U < R$ and $U > R$. If $U < R$, we do not need to go further back, as the value is determined. If $U > R$, we need to know the values of $2d$ neighbors at time t' . Repeating the argument, we construct an 'oriented percolation' structure which will eventually finish if $R > 1 - 1/2d$. This value is obtained by dominating the percolation structure with a branching process which dies out with probability R and produces $2d$ offspring with probability $1 - R$. The value of R depends on the length of the interval $[a, b]$ and the strength of the interaction governed by β . We can imagine that, for small intervals and small β , things will work.

Acknowledgements

We thank Christian Robert and Havard Rue for discussions. This research was partially supported by FAPESP, CNPq, CAPES-SECyT, Facultad de Ingeniería de la Universidad de Buenos Aires.

References

- ALLENBY, G. M. AND ROSSI, P. (1999). Marketing models of consumer heterogeneity. *J. Econometrics* **89**, 57–78.
- DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York.
- GELFAND, A. E. AND SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409.
- GEMAN, S. AND GEMAN, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelligence* **6**, 721–741.
- GEWEKE, J. (1991). Efficient simulation from the Multivariate Normal and Student t-distribution subject to linear constraints. In *Computing Sciences and Statistics* (Proc. 23rd Symp. Interface), American Statistical Association, pp. 571–577.
- GEWEKE, J., KEANE, M. AND RUNKLE, D. (1997). Statistical inference in the multinomial multiperiod probit model. *J. Econometrics* **80**, 125–165.
- MØLLER, J. (1999). Perfect simulation of conditionally specified models. *J. R. Statist. Soc. B* **61**, 251–264.

- MURDOCH, D. J. AND GREEN, P. J. (1998). Exact sampling from a continuous state space. *Scand. J. Statist.* **25**, 483–502.
- PHILIPPE, A. AND ROBERT, C. (2003). Perfect simulation of positive Gaussian distributions. *Statist. Comput.* **13**, 179–186.
- PROPP, J. G. AND WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9**, 223–252.
- ROBERT, C. (1995). Simulation of truncated normal random variables. *Statist. Comput.* **5**, 121–125.
- RUBIN, D. B. (1987). *Multiple Imputations for Nonresponse in Surveys*. John Wiley, New York.
- TANNER, M. A. (1991). *Tools for Statistical Inference. Observed Data and Data Augmentation Methods* (Lecture Notes Statist. **67**). Springer, New York.
- TANNER, M. A. AND WONG, W. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82**, 528–549.
- THORISSON, H. (2000). *Coupling, Stationarity, and Regeneration*. Springer, New York.
- WILSON, D. B. (1998). Annotated bibliography of perfectly random sampling with Markov chains. In *Microsurveys in Discrete Probability* (DIMACS Ser. Discrete Math. Theoret. Comput. Sci. **41**), eds D. Aldous and J. Propp, American Mathematical Society, Providence, RI, pp. 209–220. Updated version available at <http://dimacs.rutgers.edu/~dbwilson/exact>.