

# A completion simulator for the two-sided truncated normal distribution

Jean-Louis FOULLEY

Station de génétique quantitative et appliquée,  
Institut national de la recherche agronomique,  
78352 Jouy-en-Josas Cedex, France

(Received 18 April 2000; accepted 11 July 2000)

**Abstract** – This paper presents simulation formulae of two-sided truncated normal random variables using a completion distribution and its two corresponding conditionals generated *via* a Gibbs sampler. This procedure extends formulae given by Robert and Casella for the one-sided case.

**simulation / normal distribution / truncation / completion / Gibbs**

**Résumé** – Un simulateur de «completion» pour la loi normale tronquée aux deux extrémités. Cet article présente des formules de simulation de variables aléatoires normales réduites tronquées aux deux extrémités à partir d'une distribution de «completion» et ses deux lois conditionnelles générées par un algorithme de Gibbs. Ces formules généralisent celles de Robert et Casella établies pour le cas d'une seule troncature.

**simulation / loi normale / troncature / «completion» / Gibbs**

## 1. INTRODUCTION

Monte Carlo statistical methods are becoming increasingly popular in animal breeding and quantitative genetics [1]. In the huge field of tools required by the implementation of such techniques [6], there is a need for simulation of truncated normal distributions such as those arising in Bayesian analysis of categorical or censored data [9].

Direct simulation from the normal distribution may be quite inefficient when the probability of the class considered is small due to the high probability of rejection. In this regard, accept-reject procedures have been proposed which significantly improve the efficiency of the simulation procedure (see, for example [7]).

This note describes an alternative approach known as “completion” which is based on “demarginalization” and Gibbs sampling of the corresponding full

conditionals. This technique was studied by Damien and Walker [3] and Robert and Casella [8] applied it to the one-sided truncated normal. Here, it will be extended to the general case of a two-sided truncation.

## 2. METHOD

The theory of completion, and more generally of the slice sampler, has been presented in several scientific articles [1, 3, 4] and also in text books [8]. In the case addressed here, the method can be viewed simply as follows.

Define  $h(x) = \exp(-x^2/2)$ , the two-sided truncated normal density on the interval  $[\mu^-, \mu^+]$  can be expressed as a function proportional to:

$$f(x) = h(x)I_{[\mu^- \leq x \leq \mu^+]} \quad (1)$$

where  $I_A$  is the usual indicator function for densities equal to 1 if  $x \in A$ , and 0 otherwise.

Let us consider the following function:

$$g(x, z) = I_{[\mu^- \leq x \leq \mu^+]} I_{[0 \leq z \leq h(x)]}. \quad (2)$$

Since  $h(x)$  can be written as  $\int I_{[0 \leq z \leq h(x)]} dz$ , the integral of  $g(x, z)$  with respect to  $z$  reduces to (1).

The  $g$  function so defined is known as a "completion" of the  $f$  function; formula (2) is especially attractive as it leads to a very easy Gibbs sampler implementation *via* the two uniform conditional distributions:

$$Z|x \sim U_{[0, h(x)]}, \quad (3)$$

$$X|z \sim U_{\{x; [h(x) \geq z] \cap [\mu^- \leq x \leq \mu^+]\}}, \quad (4)$$

where  $U_{[a, b]}$  symbolizes a uniform distribution on the interval  $[a, b]$ .

Specifying the conditions for the set in (4) gives:

If  $z \leq z_0$  where  $z_0 = \min [h(\mu^-); h(\mu^+)]$

$$X|z \sim U_{[\mu^-, \mu^+]}. \quad (5a)$$

Otherwise, for  $z > z_0$ ,

when

$$\text{i) } z_0 = h(\mu^-),$$

$$X|z \sim U_{[-\sqrt{-2 \ln z}, \min(+\sqrt{-2 \ln z}, \mu^+)]}, \quad (5b)$$

$$\text{ii) } z_0 = h(\mu^+),$$

$$X|z \sim U_{[\max(-\sqrt{-2 \ln z}, \mu^-), \sqrt{-2 \ln z}]}. \quad (5c)$$

Notice that the one sided truncation can be obtained as a special case of the one addressed here for  $\mu^+ \rightarrow +\infty$  so that  $z_0 = h(\mu^+) \rightarrow 0$ , and

$$X|z \sim U_{[\text{same as in (5c)}]}. \quad (6)$$

The same reasoning also applies to the standard  $N(0, 1)$  leading to

$$X|z \sim U_{[-\sqrt{-2 \ln z}, +\sqrt{-2 \ln z}]}. \quad (7)$$

Robert and Casella [8] studied the convergence of the empirical cdf generated from (7) and showed that this algorithm appears as a very competitive alternative to, for example, the Box-Muller procedure.

### 3. NUMERICAL ILLUSTRATION

The procedure is illustrated in Table I which displays the expected and empirical expectations and variances of truncated  $N(0, 1)$  for various combinations of the lower bound  $\mu^-$  and the length of the truncation range  $\mu^+ - \mu^-$ . The expectation is given by

$$m = \frac{[\phi(\mu^-) - \phi(\mu^+)]}{[\Phi(\mu^+) - \Phi(\mu^-)]}, \quad (8)$$

where  $\phi(\cdot)$ ,  $\Phi(\cdot)$  denote the pdf and cdf respectively of the  $N(0, 1)$  distribution.

The value of the variance ( $v$ ) is obtained indirectly using the expressions for the one-sided truncated normal [2] *i.e.*

$$E(X|X > \mu^+) = i = \frac{\phi(\mu^+)}{[1 - \Phi(\mu^+)]}, \quad (9a)$$

$$\text{Var}(X|X > \mu^+) = 1 - i(i - \mu^+) \quad (9b)$$

and the expression for the variance of a mixture:

$$f(x; \mu, \sigma^2) = \sum_{j=1}^J \pi_j f_j(x; \mu_j, \sigma_j^2)$$

of  $J$  components having mean and variance  $(\mu_j, \sigma_j^2)$  with weights  $\pi_j$ :

$$\sigma^2 = \sum_{j=1}^J \pi_j [\sigma_j^2 + (\mu_j - \mu)^2]. \quad (10)$$

Formula (10) is applied here with  $\mu = 0$ ,  $\sigma^2 = 1$ ;  $\pi_1 = \Phi(\mu^-)$ ,  $\pi_2 = \Phi(\mu^+) - \Phi(\mu^-)$ ,  $\pi_3 = 1 - \Phi(\mu^+)$ ,  $\mu_3, \nu_3$  obtained from (9ab),  $\mu_1, \nu_1$  using the same expression applied to  $X \leq \mu^- \Leftrightarrow X^* = -X \geq -\mu^-$ ,  $\mu_2 = m$  in (8) and  $\sigma_2^2 = \nu$  as the unknown.

**Table I.** Expectation ( $m$ ) and variance ( $v$ ) of truncated normal distributions.a) True values ( $m$ : above,  $v$ : below).

$\mu^-$	$\mu^+ - \mu^-$			
	+0.5	+1.5	+2.5	+3.5
-3.0	-2.6949	-1.9110	-1.1317	-0.5037
	0.0188	0.1131	0.2491	0.4719
-2.0	-1.7143	-1.0430	-0.4457	-0.0830
	0.0199	0.1503	0.3766	0.6611
-1.0	-0.7345	-0.2066	0.1452	0.2687
	0.0205	0.1728	0.4157	0.5856
0.0	0.2449	0.6220	0.7724	0.7965
	0.0206	0.1647	0.3146	0.3594

b) Empirical values ( $N = 10^6$  observations).

$\mu^-$	$\mu^+ - \mu^-$			
	+0.5	+1.5	+2.5	+3.5
-3.0	-2.6947	-1.9111	-1.1317	-0.5033
	0.0188	0.1132	0.2486	0.4715
-2.0	-1.7142	-1.0418	-0.4465	-0.0827
	0.0199	0.1500	0.3772	0.6606
-1.0	-0.7347	-0.2070	0.1449	0.2673
	0.0205	0.1727	0.4155	0.5848
0.0	0.2448	0.6219	0.7720	0.7955
	0.0206	0.1647	0.3144	0.3600

Empirical values of the mean and variance given in Table I based on  $N = 10^6$  cycles are in very good agreement with expected values. The largest difference occurred for a range of  $\mu^+ - \mu^- = 3.5$  but empirical values of the mean remained within the range of  $m \pm 2\sqrt{v}$ . For instance for  $N = 10^7$ ,  $m = -0.282786$  and  $v = 0.616142$  whereas empirical values are  $\hat{m} = -0.282667$  and  $\hat{v} = 0.616251$  with a 95 percent confidence interval based on  $m$  and  $v$  equal to  $[-0.283283, -0.282289]$

As indicated by Robert and Casella [8], the algorithm displays nice ergodic properties even for small chains. Therefore, it can be used for generating few or many observations. However, it will be particularly interesting when a large number of values have to be simulated; if a single observation is needed, accept-reject procedures as described in Robert [7] may be preferred to this method. Robert's procedure is exact for a single value simulation whereas MCMC methods only provide asymptotic approximations.

**REFERENCES**

- [1] Besag J., Green P.J., Spatial statistics and Bayesian computation, *J. R. Stat. Soc. B* 55 (1993) 25–37.
- [2] Cochran W.G. Improvement by means of selection, in: *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. University of California Press, Berkeley, 1951, pp. 449–470.
- [3] Damien P., Walker S., Sampling probability densities via uniform random variables and a Gibbs sampler. Technical report, Business School, University of Michigan, 1996.
- [4] Damien P., Wakefield J., Walker S., Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables, *J. R. Stat. Soc. B* 61 (1999) 331–344.
- [5] Gianola D., Statistics in animal breeding, *J. Am. Stat. Assoc.* 95 (2000) 296–299.
- [6] Janss L.L.G., MaGGic: a package of subroutines for genetic analysis with Gibbs sampling, in: *Proceedings of the 6th World Congress on Genetics applied to Livestock Production*, Armidale, 27, 1998, pp. 459–460.
- [7] Robert C.P., Simulation of truncated normal variables, *Stat. Comput.* 5 (1995) 121–125.
- [8] Robert C.P., Casella G., *Monte Carlo Statistical Methods*, Springer, Berlin, 1999.
- [9] Sorensen D.A., Andersen S., Gianola D., Korsgaard I., Bayesian inference in threshold models using Gibbs sampling, *Genet. Sel. Evol.* 26 (1994) 333–360.