A Note on Efficient Conditional Simulation of Gaussian Distributions

Arnaud Doucet
Departments of Computer Science and Statistics,
University of British Columbia,
Vancouver, BC, Canada

April 2010

Abstract. Consider a multivariate Gaussian random vector which can be partitioned into observed and unobserved components.We review a technique proposed almost twenty years ago in the astrophysics literature to sample from the posterior Gaussian distribution of the unobserved components given the observed components [6]. This technique can be computationally cheaper than the standard approach which requires computing the Cholesky decomposition of the posterior covariance matrix. This useful method does not appear to be widely known and has been rediscovered independently in various publications.

Keywords: forward filtering backward sampling, Gaussian processes, Kalman filter and smoother, multivariate normal distribution, state-space models

Preliminary Remark. This note contains no original material and will never be submitted anywhere for publication. However it might be of interest to people working with Gaussian random fields/processes so I am making it publicly available.

1. Problem Statement

Let Z be a $\mathbb{R}^n-$valued Gaussian random vector such that

$$Z = \left(\begin{array}{c} X \\ Y \end{array} \right)$$

where X takes values in $\mathbb{R}^{n_x}$ and Y in $\mathbb{R}^{n_y}$. We assume that Z follows a multivariate normal distribution of mean m and covariance Σ

$$Z \sim \mathcal{N}(m, \Sigma)$$

with

$$m = \left(\begin{array}{c} m_x \\ m_y \end{array} \right) \text{ and } \Sigma = \left(\begin{array}{cc} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^{\mathrm{T}} & \Sigma_{yy} \end{array} \right)$$

where $m_x = E(X)$, $m_y = E(Y)$, $\Sigma_{xx} = cov(X)$, $\Sigma_{yy} = cov(Y)$ and $\Sigma_{xy} = cov(X, Y)$.

It is easy to establish that given $Y = y$, we have

$$X|\,(Y = y) \sim \mathcal{N}\left(m_{x|y}, \Sigma_{x|y}\right)$$

where

$$\begin{aligned} m_{x|y} &= m_x + \Sigma_{xy}\Sigma_{yy}^{-1}\,(y - m_y)\,, \\ \Sigma_{x|y} &= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^{\mathrm{T}}. \end{aligned}$$

Assume we are interested here in sampling from $\mathcal{N}\left(m_{x|y}, \Sigma_{x|y}\right)$. The standard approach consists of computing the Cholesky decomposition of $\Sigma_{x|y}$ denoted here $\sqrt{\Sigma_{x|y}}$ and using

$$\overline{X} = m_{x|y} + \sqrt{\Sigma_{x|y}}\; U$$

where $U \sim \mathcal{N}\,(0, I)$ is a n_x-dimensional vector of independent standard normal random variables. It is indeed easy to check that $\overline{X} \sim \mathcal{N}\left(m_{x|y}, \Sigma_{x|y}\right)$. However, it might too expensive to compute this Cholesky decomposition if $n_x \gg 1$.

2. Methodology

2.1. Algorithm. The algorithm proposed in [6] to sample $\mathcal{N}\left(m_{x|y}, \Sigma_{x|y}\right)$ can be summarized as follows.

- Sample $Z = \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\,(m, \Sigma)\,.$
- Return $\overline{X} = X + \Sigma_{xy}\Sigma_{yy}^{-1}\,(y - Y)\,.$

Compared to the standard method, this algorithm bypasses the computation of the posterior covariance $\Sigma_{x|y}$ and of its Cholesky decomposition. Contrary to the standard method, it requires being able to simulate a random vector from the prior and to use a standard regression update. In many applications, it is computationally much cheaper and easier to implement this algorithm than the standard method.

2.2. Validity of the algorithm. To establish that $\overline{X} \sim \mathcal{N}\left(m_{x|y}, \Sigma_{x|y}\right)$, we note that $\overline{X}$ satisfies

$$\overline{X} = m_{x|y} + X - E\,(X|Y)\,. \tag{1}$$

It follows that

$$E\left(\overline{X}\middle|\, Y\right) = m_{x|y} + E\,(X|Y) - E\,(X|Y) = m_{x|y}$$

Hence, we have

$$E\left(\overline{X}\right) = E\left(E\left(\overline{X}\middle|\, Y\right)\right) = m_{x|y}.$$

We also have

$$cov\left(\overline{X}|Y\right) = cov\left(\left(X - E\left(X|Y\right)\right)|Y\right) = \Sigma_{x|y}$$

as the posterior covariance is independent of the specific realization of the observations. Hence, we obtain

$$cov\left(\overline{X}\right) = \Sigma_{x|y}$$

which establishes the validity of the sampling method.

3. Applications

To the best of our knowledge, this algorithm first appeared in astrophysics where it was applied to Gaussian random fields [6]; see [7] for a recent review. In this context, n_x is so large that it is virtually impossible to compute $\Sigma_{x|y}$ and its Cholesky decomposition. This algorithm might also prove useful for Gaussian processes applications arising in spatial statistics [2] and machine learning [8].

We present here two different applications of this algorithm which have been derived independently from [6].

3.1. Ensemble Kalman filter. Consider a linear Gaussian state-space model satisfying for $n \geq 1$

$$\begin{aligned} X_n &= AX_{n-1} + V_n, \\ Y_n &= CX_n + W_n, \end{aligned}$$

where $X_0 \sim \mathcal{N}\left(0, \Sigma_0\right)$, $V_n \sim \mathcal{N}\left(0, \Sigma_v\right)$ and $W_n \sim \mathcal{N}\left(0, \Sigma_w\right)$. For any generic sequence $\left\{z_k\right\}_{k\geq 0}$, let us denote $z_{i:j} = \left(z_i, z_{i+1}, \ldots, z_j\right)$. We are interested in the posterior densities $\left\{p\left(x_n|\, y_{1:n}\right)\right\}_{n\geq 1}$. These posterior densities are Gaussian and their statistics $m_{x,n|n} = E\left(X_n|\, y_{1:n}\right)$ and $\Sigma_{xx,n|n} = cov\left(X_n|\, y_{1:n}\right)$ can be computed using the Kalman filter. However if the dimension n_x of the state X_n is very high, then it is not possible to implement the Kalman filter equations. This has motivated the development of approximation techniques in geosciences.

A very popular approach in this field is known as the Ensemble Kalman filter [4]. In the ensemble Kalman filter, the posterior distributions are approximated by random samples. Assume you have at time $n-1$, N samples $\overline{X}_{n-1}^{(i)} \sim \mathcal{N}\left(\widehat{m}_{x,n-1|n-1}, \widehat{\Sigma}_{xx,n-1|n-1}\right)$ $(i = 1, ..., N)$ where $\left(\widehat{m}_{x,n-1|n-1}, \widehat{\Sigma}_{xx,n-1|n-1}\right)$ are estimates of $\left(m_{x,n|n}, \Sigma_{xx,n|n}\right)$. Then at time n, the algorithm proceeds as follows.

- Sample $X_n^{(i)} \sim \mathcal{N}\left(A\overline{X}_{n-1}^{(i)}, \Sigma_v\right)$ and $Y_n^{(i)} \sim \mathcal{N}\left(CX_n^{(i)}, \Sigma_w\right)$.
- Compute $\widehat{m}_{x,n|n-1} = \frac{1}{N}\sum_{i=1}^N X_n^{(i)}$, $\widehat{m}_{y,n|n-1} = \frac{1}{N}\sum_{i=1}^N Y_n^{(i)}$, $\widehat{\Sigma}_{xy,n|n-1} = \frac{1}{N}\sum_{i=1}^N X_n^{(i)}\left(Y_n^{(i)}\right)^{\mathrm{T}} - \widehat{m}_{x,n|n-1}\,\widehat{m}_{y,n|n-1}^{\mathrm{T}}$, $\widehat{\Sigma}_{yy,n|n-1} = \frac{1}{N}\sum_{i=1}^N Y_n^{(i)}\left(Y_n^{(i)}\right)^{\mathrm{T}} - \widehat{m}_{y,n|n-1}\,\widehat{m}_{y,n|n-1}^{\mathrm{T}}$.

- Compute $\overline{X}_n^{(i)} = X_n^{(i)} + \widehat{\Sigma}_{xy,n|n-1}\widehat{\Sigma}_{yy,n|n-1}^{-1}\left(y_n - Y_n^{(i)}\right)$.
- Compute $\widehat{m}_{x,n|n} = \frac{1}{N}\sum_{i=1}^N \overline{X}_n^{(i)}$ and $\widehat{\Sigma}_{xx,n|n} = \frac{1}{N}\sum_{i=1}^N \overline{X}_n^{(i)}\left(\overline{X}_n^{(i)}\right)^{\mathrm{T}} - \widehat{m}_{x,n|n}\,\widehat{m}_{x,n|n}^{\mathrm{T}}$.

As N goes to infinity, it follows directly from the previous developments that $\overline{X}_n^{(i)} \sim \mathcal{N}\left(m_{x,n|n}, \Sigma_{xx,n|n}\right)$.

3.2. Posterior simulation in Gaussian state-space models. Consider again a linear Gaussian state-space model

$$X_n = AX_{n-1} + V_n, \tag{2}$$
$$Y_n = CX_n + W_n, \tag{3}$$

where $X_0 \sim \mathcal{N}(0, \Sigma_0)$, $V_n \sim \mathcal{N}(0, \Sigma_v)$ and $W_n \sim \mathcal{N}(0, \Sigma_w)$. Let us denote $y_{1:n} = (y_1, y_2, \ldots, y_n)$. When implementing a Markov chain Monte Carlo (MCMC) algorithm to estimate the hyperparameters of this model, it is usually necessary to sample from $p(x_{0:n}|\, y_{1:n})$. This is typically achieved using the Forward Filtering Backward Sampling (FFBS) technique [1], [5]. An alternative to this well-known technique is given by the following algorithm [3].

- Sample $X_{0:n}, Y_{1:n}$ using Eq. (2)-(3).
- Use the Kalman smoother to compute both $E(X_{0:n}|\, Y_{1:n})$ and $E(X_{0:n}|\, y_{1:n})$.
- Return $\overline{X}_{0:n} = E(X_{0:n}|\, y_{1:n}) + X_{0:n} - E(X_{0:n}|\, Y_{1:n})$.

The fact that $\overline{X}_{0:n} \sim p(x_{0:n}|\, y_{1:n})$ follows directly from Eq. (1). A minor advantage of this method over the FFBS approach is that it only relies on standard Kalman smoothing code. Actually, the algorithm discussed in [3] is slightly different. In this paper, the authors propose to sample from $p(x_0, v_{1:n}|\, y_{1:n})$ instead of $p(x_{0:n}|\, y_{1:n})$ using the disturbance smoother $E(V_{1:n}|\, Y_{1:n})$. The rationale for sampling from $p(x_0, v_{1:n}|\, y_{1:n})$ is that Σ_v is typically a low-rank matrix.

References

[1] C.K. Carter and R. Kohn, "On Gibbs sampling for state space models", *Biometrika*, vol. 81, pp. 541-553, 1994.

[2] N. Cressie, *Statistics for Spatial Data*, New York: Wiley, 1993.

[3] J. Durbin and S.J. Koopman, "A simple and efficient simulation smoother for state space time series analysis", *Biometrika*, vol. 89, pp. 603-616, 2002.

[4] G. Eversen, *Data assimilation: The Ensemble Kalman Filter*, 2nd ed., Springer, 2009.

[5] S. Frühwirth-Schnatter, "Data augmentation and dynamic linear models", *J. Time Series Analysis*, vol. 15, pp. 183–202, 1994.

[6] Y. Hoffman and E. Ribak, "Constrained realizations of Gaussian fields - a simple algorithm", *The Astrophysical Journal*, vol. 380, pp. L5–L8, Oct. 1991.

[7] Y. Hoffman, "Gaussian fields and constrained simulations of the large-scale structure", in *Data Analysis in Cosmology*, Lecture Notes in Physics, Berlin: Springer-Verlag, pp. 565-583, 2009.

[8] C.E. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.