

Bayesian Analysis for Normal Data with Skew-Normal Priors

*Analisi Bayesiana Per Dati Normali Quando La
Distribuzione A Priori E' Normale Asimmetrica*

Nicola Loperfido

Università di Urbino, facoltà di Economia

Via Saffi 42, 61029 Urbino (PU), ITALY

nicola@econ.uniurb.it

Riassunto: Il lavoro considera un semplice modello gerarchico normale dove la distribuzione a priori per la media della popolazione è di tipo normale asimmetrico. Si dimostra che anche la distribuzione marginale delle osservazioni è normale asimmetrica e che la distribuzione predittiva è esprimibile analiticamente.

Keywords: skew-normal distribution, Bayesian prediction

Introduction

O'Hagan and Leonard (1976) discussed a Bayesian approach to inference on the mean θ of a normal distribution with a known variance. They modeled this uncertainty hierarchically, assuming that the prior distribution on θ was conditionally normal, given the mean ζ of θ itself. The distribution of ζ , moreover, was chosen to be normal truncated at origin. Hence the marginal distribution of θ had the following form:

$$\pi_{\theta} = \frac{1}{\sigma \Phi(\tau_x)} \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left[\frac{\theta - \mu}{\sigma} \frac{\delta / \sigma + \tau}{\sqrt{1 + \delta^2}}\right]$$

(1)

where $\phi(x)$ and $\Phi(x)$ are the probability density function and cumulative distribution function of a standard normal distribution respectively. We shall refer to the above distribution as generalized skew-normal with location parameter μ , scale parameter σ , shape parameter δ and truncation parameter τ : GSN(μ ; σ ; δ ; τ). The word “generalized” reminds that the skew-normal SN(μ ; σ ; α) (Azzalini, 1985) is a generalized skew-normal distribution with truncation parameter set equal to zero.

Skew normal distribution, which appears in different areas, such as econometrics (Aigner et al., 1977), Biometrics (Roberts, 1966) and finance (Bartolucci et al., 2000) was generalized to the multivariate case (Azzalini and Dalla Valle, 1996) as follows:

$$f(\mathbf{x}) = 2\phi_n(\mathbf{x} | \mu; \Sigma) \Phi(\alpha' \mathbf{x} - \mu)$$

(2)

where $\phi_n(\mathbf{x}|\mu;\Sigma)$ is the p.d.f. of a n-dimensional normal distribution with mean μ and variance Σ . We shall write $SN(\mu;\Sigma;\alpha)$ to denote that the distribution (2) is n-dimensional skew-normal with location vector μ , scale matrix Σ and shape vector α .

Applications of the multivariate skew-normal distribution include inference from non random samples (Copas and Li, 1997) as well as discriminant analysis and graphical models (Azzalini and Capitanio, 1999).

The above applications are frequentist in nature, but distribution (2) plays a relevant role within the O'Hagan and Leonard model, too. The joint distribution of the data is multivariate skew-normal, when the prior distribution is skew-normal. Moreover under the same assumptions, the predictive distribution is generalized skew-normal.

The present paper deals with the univariate case only (i.e. the mean is a scalar). Extension to the multivariate case, where the mean is a vector subject to linear constraints, is dealt with in a work which is currently in progress.

Marginal distribution

Theorem 1: Let $X_1, X_2, \dots, X_n$ be a random sample from the conditional distribution $N(\theta; \sigma^2)$, where Θ is $SN(m; v; \lambda)$. Then the marginal density of $X_1, X_2, \dots, X_n$ is multivariate skew-normal:

$$\begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix} \equiv SN \left(m \mathbf{1}_n; \sigma^2 \mathbf{I}_n + v^2 \mathbf{1}_n \mathbf{1}_n' \frac{\lambda v \mathbf{1}_n}{\sqrt{v^2 n + \sigma^2} \sqrt{v^2 n + \sigma^2 + v^2 \lambda^2 \sigma^2}} \right) \quad (3)$$

Sketch of proof: The marginal density of $X_1, X_2, \dots, X_n$ is:

$$f_x = \int_{-\infty}^{+\infty} \frac{2}{v \sigma^n \sqrt{2\pi}^{n+1}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left[\left(\frac{x_i - \theta}{\sigma} \right)^2 + \left(\frac{\theta - m}{v} \right)^2 \right] \right\} \Phi \left(\lambda \frac{\theta - m}{v} \right) d\theta \quad (4)$$

Define now the integrand variable ω :

$$\omega = \sqrt{\frac{v^2 n + \sigma^2}{v^2 \sigma^2}} \left(\theta - \frac{m \sigma^2 + v^2 \sum x_i}{v^2 n + \sigma^2} \right) \quad (5)$$

Then (4) is equivalent to

$$f(\mathbf{x}) = 2 \phi_n(\mathbf{x} | m \mathbf{1}_n; \sigma^2 \mathbf{I}_n + v^2 \mathbf{1}_n \mathbf{1}_n') \int_{-\infty}^{+\infty} \phi_1(\omega) \Phi \left[\frac{\lambda \omega \sigma}{\sqrt{v^2 n + \sigma^2}} + \frac{\lambda v \mathbf{1}_n' (\mathbf{x} - \mathbf{1}_n m)}{v^2 n + \sigma^2} \right] d\omega \quad (5)$$

Apply now a well known property of normal integrals (Owen, 1980) and write:

$$f_{\mathbf{x}} = 2\phi_n(\mathbf{x} | m\mathbf{1}_n; \sigma^2 \mathbf{I}_n + v^2 \mathbf{1}_n \mathbf{1}_n') \Phi \left[\frac{\lambda v \mathbf{1}_n' \mathbf{x} - \mathbf{1}_n m}{\sqrt{v^2 n + \sigma^2} \sqrt{v^2 n + \sigma^2 + \lambda^2 \sigma^2}} \right]$$

(6)

By definition, the above density is equivalent to (4) and this completes the proof.

Predictive distribution

Theorem 2: Let $x_1, \dots, x_n$ be a random sample from the conditional distribution $N(\theta; \sigma^2)$, where Θ is $SN(m; v; \lambda)$. Then the predictive distribution of X_{n+1} (assumed conditionally independent of, and with the same distribution of $X_1, \dots, X_n$) is generalized skew-normal. That is:

$$X_{n+1} | x_1, \dots, x_n \equiv GSN(\mu_x; \sigma_n; \delta_n; \tau_x)$$

where

$$\mu_x = m + \frac{v^2 \sum x_i - nm}{nv^2 + \sigma^2} \quad \tau_x = \frac{\lambda v \sum x_i - nm}{\sqrt{v^2 n + \sigma^2} \sqrt{v^2 n + \sigma^2 + \lambda^2 v^2 \sigma^2}}$$

$$\sigma_n^2 = \sigma^2 + \frac{v^2}{nv^2 + \sigma^2} \quad \delta_n = \lambda v \sqrt{\frac{nv^2 \sigma^2 + \sigma^4 + v^2}{v^2 n + \sigma^2 + \lambda^2 \sigma^2}}$$

(7)

Sketch of proof: from theorem 1 we know that the joint distribution of $X_1, X_2, \dots, X_{n+1}$ is multivariate skew-normal, so that

$$f_{\mathbf{x}} = \frac{\phi_n(x; \mathbf{x} | m\mathbf{1}_{n+1}; \sigma^2 \mathbf{I}_{n+1} + v^2 \mathbf{1}_{n+1} \mathbf{1}_{n+1}')}{\phi_n(\mathbf{x} | m\mathbf{1}_n; \sigma^2 \mathbf{I}_n + v^2 \mathbf{1}_n \mathbf{1}_n')} \frac{\Phi \left[\frac{\lambda v \sum x_i + x - m - nm}{\sqrt{v^2 n + v^2 + v^2 \sigma^2} \sqrt{v^2 n + v^2 + \sigma^2 + \lambda^2 \sigma^2}} \right]}{\Phi \left[\frac{\lambda v \mathbf{1}_n' \mathbf{x} - \mathbf{1}_n m}{\sqrt{v^2 n + \sigma^2} \sqrt{v^2 n + \sigma^2 + \lambda^2 \sigma^2}} \right]}$$

(8)

From the definitions of μ_x , σ_n , δ_n and τ_x we can rewrite (11) as follows:

$$f_{\mathbf{x}} = \frac{\phi_n(x; \mathbf{x} | m\mathbf{1}_{n+1}; \sigma^2 \mathbf{I}_{n+1} + v^2 \mathbf{1}_{n+1} \mathbf{1}_{n+1}')}{\phi_n(\mathbf{x} | m\mathbf{1}_n; \sigma^2 \mathbf{I}_n + v^2 \mathbf{1}_n \mathbf{1}_n')} \cdot \frac{\Phi \left[\frac{\delta_n}{\sqrt{1 + \delta_n^2}} \frac{x - \mu_x}{\sigma_n} + \frac{\tau_x}{\sqrt{1 + \delta_n^2}} \right]}{\Phi(\tau_x)}$$

(9)

The first ratio in the above equation is the conditional density of X_{n+1} given $X_1, X_2, \dots, X_n$ when $X_1, X_2, \dots, X_{n+1}$ are assumed to be exchangeable and normally distributed with $E(X_i) = m$, $V(X_i) = v^2 + \sigma^2$, $C(X_i; X_j) = v^2$. We can therefore apply standard results of equicorrelation matrices

$$f(x|\mathbf{x}) = \frac{\phi(x|\mu_x; \sigma_n)}{\Phi(\tau_x)} \cdot \Phi \left[\frac{\delta_n}{\sqrt{1+\delta_n^2}} \frac{x-\mu_x}{\sigma_n} + \frac{\tau_x}{\sqrt{1+\delta_n^2}} \right] \quad (10)$$

Recall now the definition of generalized skew-normal distribution and the proof is complete.

Bayesian prediction

Theorem 3: Let $x_1, \dots, x_n$ be a random sample from the conditional distribution $N(\theta; \sigma^2)$, where θ is $SN(m; v; \lambda)$. Then the Bayesian prediction, with respect to the square loss function, and the corresponding expected loss, are

$$E(X_{n+1} | x_1, \dots, x_n) = \mu_x + \sigma_n \delta_n \frac{\phi(\tau_x)}{\Phi(\tau_x)}$$

$$V(X_{n+1} | x_1, \dots, x_n) = \sigma^2 + \sigma_n^2 \delta_n^2 \frac{\phi(\tau_x)}{\Phi(\lambda \tau)} \left[\gamma_z - \frac{\phi(\tau_x)}{\Phi(\tau_x)} \right] \quad (11)$$

The proof for this theorem easily follows from standard integration techniques.

References

- Aigner D.J., Lovell, C.A.K. , Schmid, P. (1977) Formulation and estimation of stochastic frontier production function model. *Journal of Econometrics*, 6, 21-37.
- Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178
- Azzalini, A. , Dalla Valle, A. (1996): The multivariate skew-normal distribution. *Biometrika*, 83, 715-726.
- Azzalini, A. , Capitanio, A. (1999) Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, B*, 61, 579-602.
- Bartolucci F., De Luca G., Loperfido N. (2000) A generalization for skewness in the basic stochastic volatility model, in *Proceedings of the 15th International workshop on Statistical Modelling*, 140-145
- Copas, J.b., Li H.G. (1997) Inference from nonrandom samples, *Journal of the Royal Statistical society B*, 59, 55-95.
- O'Hagan A. Leonard T. (1976) Bayes estimation subject to uncertainty about parameter constraints. *Biometrika* 63, 201-202.
- Roberts C. (1966) A correlation model useful in the study of twins, *Journal of the American statistical Association*, 61.
- Owen D. B. (1989) A table of normal integrals. *Communications in Statistics B*, 9, 389-419.