The Korean Communications
in Statistics Vol. 8, No. 3, 2001
pp. 797-803

Bayes Prediction Density in Linear Models

S. H. Kim[1)]

Abstract

This paper obtained Bayes prediction density for the spatial linear model with non-informative prior. It showed the results that predictive inferences is completely unaffected by departures from the normality assumption in the direction of the elliptical family and the structure of prediction density is unchanged by more than one additional future observations.

Keywords : Prediction density; Non-informative priors; Linear Regression Model; Elliptical Errors; Mixture of multivariate t density.

1. Introduction

Let's consider the problem of prediction on Z_2 based on $p(Z_1)$, where $Z_1 = (Z_1(x_1), \ldots, Z_{n_1}(x_{n_1}))'$, $Z_2 = (Z_{n_1+1}(x_{n_1+1}), \ldots, Z_n(x_n))'$ with $n = n_1 + n_2$ and $x_1, x_2, \ldots, x_{n_1}, x_{n_1+1}, \ldots, x_n$ are known locations on research region R. On a given loss function $L(Z_2, p(Z_1))$, we could minimize $E[L(Z_2, p(Z_1))]$ with respect to predictor p. Regardless of which loss function is specified, a well-known result of Bayesian decision theory shows that the optimal predictor p^* is derived by minimizing with respect to p, $E[L(Z_2, p(Z_1))|Z_1]$.

Thus the optimal predictor p^* should be depends on the predictive density $f^B(z_2 \mid Z_1)$. Hence, the predictive density is very important role in spatial statistics.

Suppose that $Z(x)$ is a real-valued stationary continuous random surfaces at location x with some smooth trend surfaces plus a spatially correlated residual process such as,

$$Z(x) = f(x)\beta + \varepsilon(x),$$

where $f(x) = (f_1(x), f_2(x), \ldots, f_k(x))$ is a known vector and β is a $k \times 1$ vector of

1) Associate Professor, Department of Statistics and Information, Anyang University, Anyang-shi 430-714, Korea.
E-mail : shkim@aycc.anyang.ac.kr

unknown regression coefficients.

The mean of $Z(x)$ is $E[Z(x)] = f(x)\beta$ and the variance is σ^2.

In general, we observe $Z_1 = (Z(x_1), \ldots, Z(x_{n_1}))'$ and will predict the future observations, $Z_2 = (Z(x_{n_1+1}), \ldots, Z(x_n))'$. Then we have,

$$Z_n = F\beta + \varepsilon_n,$$

, where F= $\{f_j(x_l)\}_{n\times k}$ $j = 1, \ldots, k$, $l = 1, \ldots, n$ and $\varepsilon_n = (\varepsilon(x_1), \ldots, \varepsilon(x_n))'$. And we can partition , $Z_n = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$, $F = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}$ and $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$ with Z_i , $n_i \times 1$ vector, F_i , $n_i \times k$ matrix and ε_i , $n_i \times 1$ vector.

For the following discussion, let the variance-covariance matrix be

$$Var(Z_n) = \sigma^2 \Sigma(\theta) = \sigma^2 \begin{bmatrix} \Sigma_{11}(\theta) & \Sigma_{12}(\theta) \\ \Sigma_{21}(\theta) & \Sigma_{22}(\theta) \end{bmatrix}.$$

Let θ be a $p\times 1$ structured parameter vector of the variance and covariance matrix. And for the following discussion, let $\Sigma_{ij}(\theta) = \Sigma_{ij}$.

The Bayes prediction density is defined as,

$$f^B(z_2|z_1) = \int f(z_2|z_1, \tau)\pi(\tau|z_1)d\tau, \tag{1.1}$$

where the integral is finite. The optimal predictor of Z_2 based on z_1 should be depends on $f^B(z_2 | z_1)$ with respect to Bayesian point of view. And τ is the parameter vector that indexes the pdf of z and $\pi(\tau)$ is the prior. In our case, $\tau = (\sigma^2, \theta, \beta)$. For $\pi(\tau)$, because β is a location parameter, we expect that our prior about β has no relationship with those about σ^2 and θ and a prior might expect σ^2 and β to be independent, leading to the use of Jeffrey's prior. Hence suppose that the non-informative prior of $(\beta, \sigma^2, \theta)$ is

$$\pi(\beta, \sigma^2, \theta) \propto \pi(\theta)/\sigma^2, \text{ where } \beta \in \mathrm{R}^k, \sigma^2 > 0, \theta \in \mathrm{R}^p.$$

The computational formulas for the proof of Section 3 will be contained on Section 2. We will investigate the Bayes prediction density under elliptical errors assumption, which is more robust assumption than normal errors in Section 3.

2. Some Quadratic Forms

Firstly, we state here the computational formula for the difference between recursive updatings of weighted least squares estimates in the general linear model when more than one additional observations become available. Let $\widehat{\beta} = (F'\Sigma^{-1}F)^{-1}F'\Sigma^{-1}z_n$ and $\widehat{\beta}_1 = (F'_1\Sigma_{11}^{-1}F_1)^{-1}F_1'\Sigma_{11}^{-1}z_1$.

[Proposition 2.1] $(\widehat{\beta}_1 - \widehat{\beta}) = -(F'\Sigma^{-1}F)^{-1}A'\Sigma_{22.1}^{-1}(z_2 - B)$,

where $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, $A = (F_2 - \Sigma_{21}\Sigma_{11}^{-1}F_1)$ and $B = A\widehat{\beta}_1 + \Sigma_{21}\Sigma_{11}^{-1}z_1$.

Secondly, we consider the partition of residual sum of squares, which is useful for Section 3. The proof can be obtained with the result of Propostion 2.1.

Let $SSE = (z_n - F\widehat{\beta})'\Sigma^{-1}(z_n - F\widehat{\beta})$ and $SSE_1 = (z_1 - F_1\widehat{\beta}_1)'\Sigma_{11}{}^{-1}(z_1 - F_1\widehat{\beta}_1)$.

[Proposition 2.2] $SSE = SSE_1 + (z_2 - B)'[\Sigma_{22.1} + A(F_1'\Sigma_{11}{}^{-1}F_1)^{-1}A']^{-1}(z_2 - B)$.

Note: Here we can find very interesting facts that B is a universal Kriging and $\sigma^2[\Sigma_{22.1} + A(F_1'\Sigma_{11}^{-1}F_1)^{-1}A']$ is the squared error of the universal Kriging. Kriging is just the best linear unbiased predictor under the general linear model. See Cressie(1991) and Ripley(1981).

3. Elliptical Errors Model

The linear regression model with non-random regressors and elliptical errors,

$$Z_n = F\beta + \Psi(T)\,u_n,$$

where $Z_n \in R^n$, $\beta \in R^k$, $u_n \sim N(0, \sigma^2\Sigma)$ and Ψ is a positive function and T is a positive random variable with distribution G, independent of u_n. The above model implies that conditional on T, $(Z_n | T, \beta, \sigma^2, \Sigma) \sim N(F\beta, (\sigma^2\Sigma \cdot \Psi^2(T)))$.

And the distribution of $Z_n | \beta, \sigma^2, \Sigma$ is

$f(z_n | \beta, \sigma^2, \Sigma) \propto$

$$\int_0^{\infty} |\sigma^2 \Sigma|^{-1/2} \{ \Psi^{-2}(T) \}^{n/2} \exp\left[\frac{-1}{2\sigma^2} \Psi^{-2}(T)\, (z_n - F\beta)' \Sigma^{-1} (z_n - F\beta) \right] dG(T).$$

From the above expression, the various distributions including the ε-contaminated normal and multivariate exponential can be generated (Muirhead(1983, pp. 32-34)).

[Theorem] Let $\pi(\beta, \sigma^2, \theta) \propto \pi(\theta) / \sigma^2$. Then the Bayes prediction density of Z_2, under the above elliptical errors model, is for any T,

$$f^B(z_2 \mid z_1) \propto \int_{\theta \in R^p} f_{n_2}(z_2; B, SSE_1 V_\theta / (n_1 - k)) \cdot |F_1' \Sigma_{11}^{-1} F_1|^{-\frac{1}{2}} |\Sigma_{11}|^{-\frac{1}{2}} (SSE_1)^{-(\frac{n_1 - k}{2})} \pi(\theta)\, d\theta,$$

where $V_\theta = [\Sigma_{22.1} + A(F_1' \Sigma_{11}^{-1} F_1)^{-1} A']$ and $f_{n_2}(z_2; B, SSE_1 V_\theta / (n_1 - k))$ is a $(n_2 \times 1)$ multivariate t pdf with the mean vector, B, the variance-covariance matrix, $SSE_1 V_\theta / (n_1 - k)$, and $n_1 - k$ degrees of freedom.

[Proof] We consider the prediction density under the spatial linear model with elliptical errors. The Bayes prediction density can be expressed by (1.1) as follows,

$$f^B(z_2 \mid z_1) = \int f(z_2 \mid z_1, \theta)\, \pi(\theta \mid z_1)\, d\theta \tag{3.1}$$

$$, \text{ where } f(z_2 \mid z_1, \theta) = \int f(z_2 \mid z_1, \theta, T)\, dG(T). \tag{3.2}$$

Consider the integrand of (3.2). By the Bayes rule, it can be expressed by,

$$f(z_2 \mid z_1, \theta, T) = \frac{f(z_1, z_2 \mid \theta, T)}{f(z_1 \mid \theta, T)}. \tag{3.3}$$

The nominator and denominator of (3.3) can be expressed with respect to elliptical error model as follows,

$$f(z_1, z_2 \mid \theta, T) = \int_{\sigma^2 > 0} \int_{\beta \in R^k} f(z_1, z_2 \mid \theta, \beta, \sigma^2, T)\, \pi(\sigma^2, \beta \mid \theta, T)\, d\beta\, d\sigma^2 \text{ and}$$

$$f(z_1 \mid \theta, T) = \int_{\sigma^2 > 0} \int_{\beta \in R^k} f(z_1 \mid \theta, \beta, \sigma^2, T)\, \pi(\sigma^2, \beta \mid \theta, T)\, d\beta\, d\sigma^2, \tag{3.4}$$

where $\pi(\sigma^2, \beta | \theta, T) \propto \frac{1}{\sigma^2}$.

Since $Z_1, Z_2 | \beta, \sigma^2, \theta, T \sim N(F\beta, \sigma^2\Sigma\, \Psi^2(T))$,

$$f(z_1, z_2 | \theta, T) \propto \int\int \frac{1}{|\sigma^2\Sigma\,\Psi^2(T)|^{1/2}} \cdot \exp\left\{-\frac{1}{2\sigma^2\Psi^2(T)}(z_n - F\beta)'\Sigma^{-1}(z_n - F\beta)\right\} \frac{1}{\sigma^2}\, d\beta\, d\sigma^2. \tag{3.5}$$

Consider $(z_n - F\beta)'\Sigma^{-1}(z_n - F\beta)$. Then this quadratic form can be partitioned into two parts as follows,

$$(z_n - F\beta)'\Sigma^{-1}(z_n - F\beta) = SSE + (\beta - \hat{\beta})'(F'\Sigma^{-1}F)(\beta - \hat{\beta}). \tag{3.6}$$

By (3.6), (3.5) can be expressed by,

$$f(z_1, z_2 | \theta, T) \propto \int_{\sigma^2>0} \frac{1}{|\sigma^2\Sigma\,\Psi^2(T)|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2\Psi^2(T)} SSE\right\} \frac{1}{\sigma^2} \cdot \left[\int_{\beta\in R^k} \exp\left\{-\frac{1}{2\sigma^2\Psi^2(T)}(\hat{\beta} - \beta)'(F'\Sigma^{-1}F)(\hat{\beta} - \beta)\right\} d\beta\right] d\sigma^2. \tag{3.7}$$

By the well-known fact, $\hat{\beta} | \sigma^2, \theta, T \sim N(\beta, \sigma^2\Psi^2(T)(F'\Sigma^{-1}F)^{-1})$, (3.7) is

$$f(z_1, z_2 | \theta, T) \propto \int_{\sigma^2>0} \frac{1}{|\sigma^2\Sigma\,\Psi^2(T)|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2\Psi^2(T)} \cdot SSE\right\} \cdot |\sigma^2\Psi^2(T)(F'\Sigma^{-1}F)^{-1}|^{\frac{1}{2}} \frac{1}{\sigma^2}\, d\sigma^2. \tag{3.8}$$

Using both the technique of change of variable and the property of Gamma function, (3.8) can be expressed by,

$$f(z_1, z_2 | \theta, T) \propto |F'\Sigma^{-1}F|^{-\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} (SSE)^{-(\frac{n-k}{2})}. \tag{3.9}$$

Similarly, it can be shown that

$$f(z_1 \mid \theta, T) \propto |F_1' \Sigma_{11}^{-1} F_1|^{-\frac{1}{2}} |\Sigma_{11}|^{-\frac{1}{2}} (SSE_1)^{-(\frac{n_1-k}{2})} . \tag{3.10}$$

Since $f(z_2 | z_1, T, \theta)$ is the ratio between (3.9) and (3.10), we have

$$f(z_2 | z_1, T, \theta) \propto \frac{|F' \Sigma^{-1} F|^{-\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} (SSE)^{-(\frac{n-k}{2})}}{|F_1' \Sigma_{11}{}^{-1} F_1|^{-\frac{1}{2}} |\Sigma_{11}|^{-\frac{1}{2}} (SSE_1)^{-(\frac{n_1-k}{2})}} . \tag{3.11}$$

Because (3.11) is not dependent on T, (3.11) is a same density as $f(z_2 | z_1, \theta)$ with respect to (3.2). The second integrand of (3.1), $\pi(\theta | z_1)$ can be expressed by Bayes rule as follow,

$$\pi(\theta | z_1) \propto f(z_1 \mid \theta) \pi(\theta) . \tag{3.12}$$

Since the posterior density of Z_1, $f(z_1 \mid \theta)$ can be obtained by (3.10), hence (3.12) is,

$$\pi(\theta | z_1) \propto |F_1' \Sigma^{-1} F_1|^{-\frac{1}{2}} |\Sigma_{11}|^{-\frac{1}{2}} (SSE_1)^{-(\frac{n_1-k}{2})} \pi(\theta) . \tag{3.13}$$

When we put the results of (3.11) and (3.12) into the integrands of (3.1), the prediction density $f^B(z_2 \mid z_1)$ can be written by,

$$f^B(z_2 \mid z_1) \propto \int_{\theta \in R^p} (SSE)^{-(\frac{n-k}{2})} |\Sigma|^{-\frac{1}{2}} | F' \Sigma^{-1} F|^{-\frac{1}{2}} \pi(\theta)\, d\theta . \tag{3.14}$$

Now we consider the integrands of (3.14). First, consider $(SSE)^{-\frac{(n-k)}{2}}$. We can express this by Proposition 2.2 as follows,

$$SSE^{-\frac{(n-k)}{2}} = \frac{1}{[1 + \frac{1}{n_1-k} (z_2 - B)' (SSE_1 V_\theta / (n_1-k))^{-1} (z_2 - B)]^{\frac{(n-k)}{2}}} \cdot (SSE_1)^{-\frac{n_2}{2}} (SSE_1)^{-\frac{n_1-k}{2}} . \tag{3.15}$$

Secondly, using the simple property of determinant function, we can express,

$$|\Sigma|^{-\frac{1}{2}} = |\Sigma_{22.1}|^{-\frac{1}{2}} \cdot |\Sigma_{11}|^{-\frac{1}{2}} . \tag{3.16}$$

Finally, we consider the determinant function such as

$$|F' \Sigma^{-1} F|^{-\frac{1}{2}} = |F_1' \Sigma_{11}^{-1} F_1|^{-\frac{1}{2}} |I + A' \Sigma_{22.1}^{-1} A (F_1' \Sigma_{22.1} F_1)^{-1}|^{-\frac{1}{2}}. \tag{3.17}$$

Using the simple property of determinant function, we can derive

$$|\Sigma_{22.1} + A(F_1' \Sigma_{22.1} F_1)^{-1} A'|^{-\frac{1}{2}} = |\Sigma_{22.1}|^{-\frac{1}{2}} |I + A' \Sigma_{22.1}^{-1} A (F_1' \Sigma_{22.1} F_1)^{-1}|^{-\frac{1}{2}}. \tag{3.18}$$

Using from (3.15) to (3.18), the integrand of (3.14) can be expressed by,

$$\begin{aligned} (SSE)^{-(\frac{n-k}{2})} |\Sigma|^{-\frac{1}{2}} |F' \Sigma^{-1} F|^{-\frac{1}{2}} \pi(\theta) \propto\ & |SSE_1 V_\theta / (n_1 - k)|^{-\frac{1}{2}} \\ & \cdot \frac{1}{[1 + \frac{1}{n_1 - k} (z_2 - B)' (SSE_1 V_\theta / (n_1 - k))^{-1} (z_2 - B)]^{\frac{(n-k)}{2}}} \\ & \cdot |F_1' \Sigma_{11}^{-1} F_1|^{-\frac{1}{2}} |\Sigma_{11}|^{-\frac{1}{2}} (SSE_1)^{-(\frac{n_1 - k}{2})} \pi(\theta). \end{aligned} \tag{3.19}$$

When we put (3.19) onto (3.14), the prediction density $f^B(z_2 \mid z_1)$ can be obtained. ■

Remark: We showed that the prediction density of Z_2 based on z_1 is the mixture of multivariate t density under elliptical errors. It is noted that Handcock and Stein(1993) obtained the same prediction density under multivariate normal errors for only one future observation. Hence our result extended theirs such that the prediction density is unaffected by a change in the error distribution from multivariate normal to elliptical and the structure of density is unchanged by more than one additional future observations. Handcock and Wallis (1994) used the Handcock and Stein (1993)'s result too.

References

[1] Cressie, N.A.C.(1991). *Statistics for Spatial Data*, John Wiley, New York.

[2] Handcock, M.S. and Stein, M.L.(1993). A Bayesian Analysis of Kriging, *Technometrics*, 35, 403-410.

[3] Handcock, M.S. and Wallis, J.R.(1994). An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields, *Journal of the American Statistical Association*, 89, 368-378.

[4] Lee, S.E.(2001). Bayes Prediction for Small Area Estimation, *The Korean Communication in Statistics*, 8(2), 407-416.

[5] Muirhead, R.J.(1983). *Aspects of multivariate statistical theory*, Wiley, New York.

[6] Ripley, B.D.(1981) *Spatial Statistics*, John Wiley, New York.