

Lecture V: Kernels and the Linear Normal Model

Charles B. Moss

May 6, 2013

I. Kernels

A. Returning to some of the results from Lecture 4.

1. **Definition 2.1.1** The *marginal likelihood* of the model A is $p(y^0 | A)$
2. **Definition 2.1.2** The *posterior density* of the vector of unobservables θ_A in model A is

$$p(\theta_A | y^0, A) = \frac{p(\theta_A | A) \times p(y^0 | \theta_A, A)}{p(y^0 | A)} \quad (1)$$

3. The posterior density can then be expressed as

$$\begin{aligned} p(\alpha_1 | \{0.1532, 0.1620, 0.1548\}, A) = \\ \frac{K \exp \left[-\frac{((0.1532 - \alpha_1)^2 + (0.1620 - \alpha_1)^2 + (0.1548 - \alpha_1)^2)}{2 \times 0.1125} - \frac{(\alpha_1 - 0.25)^2}{2 \times 0.125} \right]}{K \times 0.786801} \end{aligned} \quad (2)$$

Notice that the K term can be cancelled out of the expression

$$\begin{aligned} p(\alpha_1 | \{0.1532, 0.1620, 0.1548\}, A) = \\ \frac{\exp \left[-\frac{((0.1532 - \alpha_1)^2 + (0.1620 - \alpha_1)^2 + (0.1548 - \alpha_1)^2)}{2 \times 0.1125} - \frac{(\alpha_1 - 0.25)^2}{2 \times 0.125} \right]}{0.786801} \end{aligned} \quad (3)$$

Thus, all the "relevant" information is contained in the exponential term

$$p(\{s_{11}, s_{12}, s_{13}\}, \alpha_1 | A) \propto \exp \left[-\frac{1}{2k} ((s_{11} - \alpha_1)^2 + (s_{12} - \alpha_1)^2 + (s_{13} - \alpha_1)^2) \right] \quad (4)$$

(where k is a known constant in this case by assumption).

B. **Definition 2.1.3** Any nonnegative function $k(x)$ proportional to a probability density function $p(x)$ is a *kernel* of $p(x)$.

1. The result presented in Equation 4 is such a kernel

$$k(\{s_{11}, s_{12}, s_{13}\}, \alpha_1 | A) = \exp \left[-\frac{1}{2k} ((s_{11} - \alpha_1)^2 + (s_{12} - \alpha_1)^2 + (s_{13} - \alpha_1)^2) \right] \quad (5)$$

2. We can refer to

$$p(\theta_A | y^0, A) = \frac{p(\theta_A | A) \times p(y^0 | \theta_A, A)}{p(y^0 | A)} \quad (6)$$

as the *normalized posterior density* while

$$p(\theta_A | y^0, A) \propto p(\theta_A | A) \times p(y^0 | \theta_A, A) \quad (7)$$

is the *posterior density kernel in standard form*. Explicitly, since $p(y^0 | A)$ is a constant all we need to know about the posterior is the numerator. Thus from Equation 3

$$k(\alpha_1 | \{s_{11}, s_{12}, s_{13}\}, A) = \exp \left[-\frac{((0.1532 - \alpha_1)^2 + (0.1620 - \alpha_1)^2 + (0.1548 - \alpha_1)^2)}{2 \times 0.1125} - \frac{(\alpha_1 - 0.25)^2}{2 \times 0.125} \right] \quad (8)$$

C. **Definition 2.1.4** A *complete model* A consists of three components: the observed density $p(y | \theta_A, A)$, the prior density $p(\theta_A | A)$, and the vector of interest density $p(\omega | y, \theta_A, A)$.

1. "The objective of inference can generally be expressed as the posterior density of the vector of interest ω (Geweke)." For example, we may be interested in the elasticity of an input demand function with respect to the price of the input.
2. For a particular model A

$$p(\omega | y^0, A) = \int_{\Theta_A} p(\omega | y^0, \theta_A, A) p(\theta_A | y^0, A) dv(\theta_A) \quad (9)$$

3. "If a hypothesis restricts θ_A to a set $\Theta_{A0} \subset \Theta_A$, then by taking $h(\omega) = I_{\Theta_{A0}}(\theta_A)$, we have $E[h(\omega) | y^0, A] = P(\theta_A \in \Theta_{A0} | y^0, A)$, the posterior probability that the hypothesis is true in the context of model A (Geweke)."

II. Example 2.1.2 Normal Linear Regression Model

- A. Setup – we start with a vector of dependent variables $y \in M_{T \times 1}$ and a matrix of independent variables $X \in M_{T \times k}$. In the standard Ordinary Least Squares setup

$$y = X\beta + \epsilon \Rightarrow \beta_{OLS} = (X'X)^{-1} X'y. \quad (10)$$

Under standard Ordinary Least Squares assumptions $T \gg k$ so that $|X'X| > 0$ and is invertible.

- B. Under the standard model

$$y | (\beta, h, X, A) \sim N(X\beta, h^{-1}I_T) \quad \text{rank}(X) = k \quad (11)$$

so y is distributed normally with a mean of $X\beta$ and a variance of $h^{-1}I_T$ (I_T is an identity matrix of order T).

1. Under standard assumptions the independent variables X are fixed.
2. "The parameter h is the *precision* of each of the i.i.d. disturbances ϵ_t ; it is the inverse of the variance $\text{var}(\epsilon_t) = \sigma^2$ (Geweke)." Under Ordinary Least Squares we typically assume that $y \sim N(X\beta, \sigma^2 I_T)$. Thus, if $h = 1/\sigma^2 \Rightarrow h^{-1} = \sigma^2$. Hence, $h^{-1}I_T \Rightarrow \sigma^2 I_T$.

Table 1: Ten Years of KLEM Capital Data

	Dependent Variable	Independent Variable			
		Log of Output	Log of Input Prices		
	Capital Share		Capital	Labor	Energy
1960	0.1532	11.9413	-0.2894	-0.9707	-0.4768
1961	0.1620	11.9593	-0.2144	-0.9253	-0.4794
1962	0.1548	11.9544	-0.2646	-0.8764	-0.5117
1963	0.1551	11.9882	-0.2523	-0.8360	-0.5148
1964	0.1343	11.9818	-0.4185	-0.7026	-0.5397
1965	0.1534	12.0095	-0.2654	-0.6403	-0.5484
1966	0.1587	12.0039	-0.2324	-0.6057	-0.5884
1967	0.1501	12.0405	-0.3064	-0.5192	-0.5632
1968	0.1459	12.0533	-0.3671	-0.4424	-0.5920
1969	0.1818	12.0633	-0.0874	-0.4258	-0.6404

3. The parameter vector from our previous discussion can then be expressed as $\theta'_A = (\beta', h)$.
 - a. The β coefficient and the precision coefficient h are independent in the prior.
 - b. The prior distribution of β is

$$\beta | A \sim N(\tilde{\beta}, \tilde{H}^{-1}) \quad (12)$$

where $\tilde{\beta}$ is our expected value of β and $\tilde{H}$ is the inverse of the covariance matrix for those expectations.

- c. To nail these concepts down a little more, consider the simple share equation for the share of capital used in agricultural production for 1960 – 1969

$$s_{1t} = \alpha_1 + \Gamma_{11} \ln(q_t) + A_{11} \ln(w_{1t}) + A_{12} \ln(w_{2t}) + A_{13} \ln(w_{3t}) + \epsilon_t \quad (13)$$

using the data presented in Table 1.

- c. The question of a prior would seem to be of tremendous importance. In the case of the input share equation, I think a good prior would be that the translog formulation

was uninformed – so that the only nonzero coefficient is the constant term α_1 . Using the sample data $\hat{\alpha}_1 = 0.1549$. The prior for the other coefficients would then be zero. The prior

$$\tilde{\beta} = \begin{bmatrix} 0.1549 \\ 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.0000 \end{bmatrix} \quad (14)$$

- d. The next question is the prior of the variance matrix. The variance of the mean (i.e., the variance of $\hat{\alpha}_1$ can be derived from the sample as

$$V[\hat{\alpha}_1] = \frac{\sigma^2}{n} = 0.000014689 \quad (15)$$

One approach would be to set the prior variance to a multiple of this value

$$\tilde{H}^{-1} = aV[\hat{\alpha}_1] I_k = 0.000014689a \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (16)$$

Therefore

$$\tilde{H} = \frac{1}{0.000014689a} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (17)$$

- e. The prior distribution for h is

$$s^2 h | A \sim \chi^2(\tilde{\nu}) \quad (18)$$

hence $E[s^2 h] \rightarrow 1$ or the normalized s^2 limits to a χ^2 distribution.

4. Building the prior, we start with the prior for β

$$p(\beta|A) = \left(\frac{1}{\sqrt{2\pi}}\right)^k |\tilde{H}|^{1/2} \exp \left[-(\beta - \tilde{\beta})' \tilde{H} (\beta - \tilde{\beta}) / 2 \right]. \quad (19)$$

Working on the prior for h we start by considering the χ^2 distribution

$$p(w) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} w^{(\nu-2)/2} \exp \left[-\frac{w}{2} \right] \quad (20)$$

where $E(w) = \nu$ and $V[w] = 2\nu$. The prior mean of h is $E[h|A] = \tilde{\nu}/\tilde{s}^2$ and the variance is $V[h|A] = 2\tilde{\nu}/\tilde{s}^4$. The probability density function for h is then

$$p(h|A) = \frac{1}{2^{\tilde{\nu}/2} \Gamma(\tilde{\nu}/2)} (\tilde{s}^2)^{\tilde{\nu}/2} h^{(\tilde{\nu}-2)/2} \exp \left[-\frac{\tilde{s}^2 h}{2} \right] \quad (21)$$

5. The distribution for the observable variables is then

$$p(y|\beta, h, X, A) = \left(\frac{1}{\sqrt{2\pi}}\right)^T h^{T/2} \exp \left[-h(y - X\beta)'(y - X\beta) / 2 \right] \quad (22)$$

6. Taking the development a piece at time – let us multiply Equation 19 times Equation 22

$$\begin{aligned} p(\beta|A) \times p(y|\beta, h, X, A) &= p(y, \beta|h, X, A) = \\ &\left(\frac{1}{\sqrt{2\pi}}\right)^k \left(\frac{1}{\sqrt{2\pi}}\right)^T |\tilde{H}|^{1/2} h^{T/2} \\ &\exp \left[-\left\{ (\beta - \tilde{\beta})' \tilde{H} (\beta - \tilde{\beta}) + h(y - X\beta)'(y - X\beta) \right\} / 2 \right] \end{aligned} \quad (23)$$

which is Equation 2.18d in Geweke. Note that $|\tilde{H}|^{1/2}$ is a constant once the prior is set. Hence, Equation 23 can be rewritten as

$$p(y, \beta | h, X, A) = K_{\tilde{H}} h^{T/2} \exp \left[- \left\{ (\beta - \tilde{\beta})' \tilde{H} (\beta - \tilde{\beta}) + h(y - X\beta)'(y - X\beta) \right\} / 2 \right] \quad (24)$$

where $K_{\tilde{H}}$ is a constant.

1. Completing the matrix operation in the exponent of Equation 24

$$\begin{aligned} & (\beta - \tilde{\beta})' \tilde{H} (\beta - \tilde{\beta}) + h(y - X\beta)'(y - X\beta) \\ &= \beta' \tilde{H} \beta - \beta' \tilde{H} \tilde{\beta} - \tilde{\beta}' \tilde{H} \beta + \tilde{\beta}' \tilde{H} \tilde{\beta} + hy'y - hy'X\beta - h\beta'X'y + h\beta'X'X\beta \end{aligned} \quad (25)$$

2. Start by separating the $\beta'K\beta$ terms

$$\beta' \tilde{H} \beta + \beta' X' X X \beta \Rightarrow \beta' [\tilde{H} + hX'X] \beta \quad (26)$$

here we let $\bar{H} = \tilde{H} + hX'X$ - note that $\bar{H}(h) = \tilde{H} + hX'X$ because we if assume that X is fixed in the sample the choice value or actually the value that we base the integral on is h .

3. Next, consider the $\beta'K\tilde{\beta}$ terms

$$\beta' \tilde{H} \tilde{\beta} + h\beta'X'y \Rightarrow \beta' [\tilde{H}\tilde{\beta} + hX'y] \quad (27)$$

using this expression we can define $\bar{\beta}' = \bar{H}^{-1} [\tilde{H}\tilde{\beta} + hX'Xb]$.

In part this substitution is because $b = (X'X)^{-1} X'y \Rightarrow (X'X)b = X'y$.

4. Collecting the left-over terms

$$Q = hy'y + \tilde{\beta}' \tilde{H} \tilde{\beta} + \bar{\beta}' \bar{H} \bar{\beta}. \quad (28)$$

5. Collecting the terms and assuming that we have a sample $\Rightarrow \{y^0, X^0\}$.
 - a. $\bar{H}(h) = \tilde{H} + hX^{0'}X^0$.

- b. $\bar{\beta}(h) = \bar{H}^{-1}(h) [\tilde{H}\tilde{\beta} + hX^{0'}X^0b]$ – where $b = (X^{0'}X^0)^{-1}X^{0'}y$.
 - c. $p(\beta|h, y^0, X^0, A) \propto \exp \left[-(\beta - \bar{\beta}(h))' \bar{H}(h) (\beta - \bar{\beta}(h)) \right]$.
 - d. Hence $\beta|(h, y^0, X^0, A) \sim N(\bar{\beta}(h), \bar{H}^{-1}(h))$.
 - e. The point is then taking the expectations of $\bar{\beta}(h)$ and $\bar{H}(h)$ with respect to the χ^2 distribution of h .
- C. Returning to the cost share example, let $a = 1,000$. The prior of the variance matrix is then

$$\tilde{H} = \begin{bmatrix} 68.0766 & 0 & 0 & 0 & 0 \\ 0 & 68.0766 & 0 & 0 & 0 \\ 0 & 0 & 68.0766 & 0 & 0 \\ 0 & 0 & 0 & 68.0766 & 0 \\ 0 & 0 & 0 & 0 & 68.0766 \end{bmatrix}. \quad (29)$$

Setting $\nu = 1$ and $s^2 = 0.0000146893$, the modified χ^2 distribution becomes

$$f(h) = \frac{0.001529101}{\sqrt{h}} \exp[-0.00000724467h] \quad (30)$$

Given $b = (-0.35380 \ 0.044222 \ 0.129403 \ -0.007595 \ -0.014154)'$ the expectation becomes

$$E[\bar{\beta}(h)] = \int_0^\infty f(h) \bar{\beta}(h) dh = \begin{bmatrix} 0.15570 \\ 0.00186 \\ 0.05253 \\ 0.00350 \\ -0.00531 \end{bmatrix} \quad (31)$$

$$\begin{pmatrix} 680834. & 8.16888 \times 10^6 & -183664. & -472751. & -371344. \\ 8.16888 \times 10^6 & 9.80241 \times 10^7 & -2.20346 \times 10^6 & -5.66775 \times 10^6 & -4.4572 \times 10^6 \\ -183664. & -2.20346 \times 10^6 & 54478.9 & 128363. & 99396. \\ -472751. & -5.66775 \times 10^6 & 128363. & 352875. & 251813. \\ -371344. & -4.4572 \times 10^6 & 99396. & 251813. & 204299. \end{pmatrix} \quad (32)$$

D. Finally, the updated h distribution is computed by the fact that

$$p(h \mid \beta, y^0, X, A) \propto h^{(T+\tilde{\nu}-2)/2} \exp \left[- \left[\tilde{s}^2 + (y^0 + X\beta)' (y^0 - X\beta) \right] h/2 \right]. \quad (33)$$

Hence.

$$\bar{s}^2 h \mid (\beta, y^0, X, A) \sim \chi^2(\bar{\nu}) \quad (34)$$

where

$$\begin{aligned} \bar{s}^2 &= \tilde{s}^2 + (y^0 - X\beta)' (y^0 - X\beta) \quad \text{and} \\ \bar{\nu} &= T + \tilde{\nu} \end{aligned} \quad (35)$$