CHAPTER 5

Expansion of the Posterior, Bayes Estimate and Bayes Risk

5.1. Expansion of the posterior. Let $P(\theta)$ as well as $\pi(\theta)$ stand for a prior probability density and, deviating slightly from Ghosh, Sinha and Joshi [(1982), page 422]

$$b = -\frac{1}{n}\frac{d^2 \log p(X_1, X_2, \dots, X_n|\theta)}{d\theta^2}\bigg|_{\hat{\theta}}. \tag{5.1}$$

Let

$$F_n(h) = P\left\{\sqrt{nb}\,(\theta - \hat{\theta}) < h|X_1, X_2, \dots, X_n\right\} \tag{5.2}$$

be the posterior distribution function of the normalized θ. Under various conditions, $F_n(h)$ is approximately $\Phi(h)$, where Φ is the standard normal distribution function.

Here is a typical result. Assume regularity conditions on $p(x|\theta)$ and let $\pi(\theta)$ be continuous and positive at a fixed point θ_0. Then

$$\lim_{n\to\infty} \sup_h |F_n(h) - \Phi(h)| \to 0 \quad \text{a.s.}\left(P_{\theta_0}\right). \tag{5.3}$$

Le Cam (1958) has a similar theorem under P_π, where P_π is the marginal distribution of $\{X_n\}$ under $\pi \otimes P_\theta$. Here $\pi \otimes P_\theta$ stands for the joint distribution of θ and X's under which θ has density $\pi(\theta)$ and, given θ, X's have the joint distribution P_θ. Under P_θ, X's are i.i.d. $p(x|\theta)$.

Under stronger conditions on $p(x|\theta)$ and the assumption of $(k+1)$ times continuous differentiability of $\pi(\theta)$ at θ_0 and $P(\theta_0) > 0$, Johnson (1970) proves the following rigorous and precise version of a refinement due to Lindley (1961).

Fix positive integers r and k. Then under regularity conditions depending on r and k,

$$P_{\theta_0}\left\{\sup_h |F_n(h) - \Phi(h) - \sum_{j=1}^{k} (\cdot) n^{-j/2}| \le M n^{-(k+1)/2}\right\} = 1 - O(n^{-r}), \tag{5.4}$$

where $(\cdot)$ are terms similar to those appearing in Edgeworth expansions, namely, each is

$$\phi(h)\{\text{polynomial in } h \text{ with coefficients depending on } X_1, X_2, \ldots, X_n\}.$$

An explicit result is given in (5.4f). The above result is a reformulation of Johnson's theorem and is taken from Ghosh, Sinha and Joshi (1982). Under the same assumptions one can get a similar theorem involving the L_1 distance between the posterior density and an approximation. If we take $r > 1$ we can immediately get, by the Borel–Cantelli lemma, an a.s. version which is similar to (5.3).

The proof of (5.4) is similar to the derivation of the formal Edgeworth expansion in Chapter 2, except that Taylor expansion of log likelihood and prior takes on the role of expansion of f, the log characteristic function, and hence no inversion like (2.8) is needed. Since no inversion is needed, rigorous justification is much easier than that for Edgeworth expansions, and consists of essentially two steps. The first step is to show that the tails of the posterior are negligible. The second step is to expand by Taylor's theorem in the remaining part, that is, for, say, $|\theta - \hat{\theta}| < (\log n)/\sqrt{n}$. Note that the first term in the expansion is zero because $\hat{\theta}$ satisfies the likelihood equation, the second term is a quadratic $-nb(\theta - \hat{\theta})^2$ which leads to posterior normality as in (5.3) and the subsequent terms provide the refinement in (5.4). A "formal" argument showing how the terms are calculated is presented below

Let

$$a_i = \frac{1}{n} \frac{d^i \log p(X_1, X_2, \ldots, X_n|\theta)}{d\theta^i}\bigg|_{\hat{\theta}}, \tag{5.4a}$$

so that $b = -\alpha_2$. Let $h_1 = \sqrt{n}(\theta - \hat{\theta}) = h/\sqrt{b}$. Then

$$\pi(\hat{\theta} + n^{-1/2}h_1) = \pi(\hat{\theta})\left[1 + n^{-1/2}h_1 \frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} + \frac{1}{2} n^{-1} h_1^2 \frac{\pi''(\hat{\theta})}{\pi(\theta)}\right] + o(n^{-1}). \tag{5.4b}$$

Let $L(\theta) = \log p(X_1, X_2, \ldots, X_n|\theta)$. Then

$$L(\hat{\theta} + n^{-1/2}h_1) - L(\hat{\theta}) = -\tfrac{1}{2}h_1^2 b + \tfrac{1}{6} n^{-1/2} h_1^3 a_3 + \tfrac{1}{24} n^{-1} h_1^4 a_4 + o(n^{-1}). \tag{5.4c}$$

Hence,

$$\pi(\hat\theta + n^{-1/2}h_1)\exp\left[L(\hat\theta + n^{-1/2}h_1) - L(\hat\theta)\right]$$

$$= \pi(\hat\theta)\left[\exp\left\{-\frac{1}{2}h_1^2 b\right\}\right]$$

$$\times\left[1 + n^{-1/2}\left\{\frac{1}{6}h_1^3 a_3 + h_1\frac{\pi'(\hat\theta)}{\pi(\hat\theta)}\right\}\right. \tag{5.4d}$$

$$\left. + n^{-1}\left\{\frac{1}{24}h_1^4 a_4 + \frac{1}{72}h_1^6 a_3^2 + \frac{1}{2}h_1^2\frac{\pi''(\hat\theta)}{\pi(\hat\theta)} + \frac{1}{6}h_1^4 a_3\frac{\pi'(\hat\theta)}{\pi(\hat\theta)}\right\}\right],$$

$$\int \pi(\hat\theta + n^{-1/2}h_1)\exp\left[L(\hat\theta + n^{-1/2}h_1) - L(\hat\theta)\right] dh_1$$

$$= \pi(\hat\theta)\sqrt{\frac{2\pi}{b}}\left[1 + n^{-1}\left\{\frac{a_4}{8b^2} + \frac{15}{72b^6}a_3^2 + \frac{1}{2b}\frac{\pi''(\hat\theta)}{\pi(\hat\theta)}\right.\right. \tag{5.4e}$$

$$\left.\left. + \frac{1}{2b^2}a_3\frac{\pi'(\hat\theta)}{\pi(\hat\theta)}\right\}\right] + o(n^{-1}).$$

The posterior density of $h_1 = \sqrt{n}(\theta - \hat\theta)$ is the ratio of the above two expressions and equals

$$\pi(h_1|X_1, X_2, \dots, X_n)$$

$$= \sqrt{\frac{b}{2\pi}}\,e^{-h_1^2/2}\left[1 + n^{-1/2}\left\{\frac{1}{6}h^3 a_3 + h\frac{\pi'(\hat\theta)}{\pi(\hat\theta)}\right\}\right.$$

$$+ n^{-1}\left\{\left(\frac{1}{24}h^4 a_4 + \frac{1}{72}h^6 a_3^2\right.\right. \tag{5.4f}$$

$$\left. + \frac{1}{2}h^2\frac{\pi''(\hat\theta)}{\pi(\hat\theta)} + \frac{1}{6}h^4 a_3\frac{\pi'(\hat\theta)}{\pi(\hat\theta)}\right)$$

$$\left.\left. - \left(\frac{a_4}{8b^2} + \frac{15}{72b^6}a_3^2 + \frac{1}{2b}\frac{\pi''(\hat\theta)}{\pi(\hat\theta)} + \frac{1}{2b^2}a_3\frac{\pi'(\hat\theta)}{\pi(\hat\theta)}\right)\right\}\right] + o(n^{-1}).$$

Transforming to $h = (\sqrt{b})h_1$, we get the expansion for the posterior density of $\sqrt{nb}(\theta - \hat\theta)$, and integrating that from $-\infty$ to h, we get the terms in (5.4).

Integrating h_1 with respect to $\pi(h_1|X_2, X_2, \dots, X_n)$, we get a formal expansion for the posterior mean:

$$B_n = \hat\theta + n^{-1}\left\{\frac{a_3}{2b} + \frac{1}{b}\frac{\pi'(\hat\theta)}{\pi(\hat\theta)}\right\} + o(n^{-3/2}). \tag{5.4g}$$

The constant M appearing in (5.4) can be somewhat misleading. See the first section of Ghosh, Sinha and Joshi (1982) for more details on M and the possibility of drawing wrong conclusions from (5.4). We note that M depends on θ_0 through the values of $p(\theta)$ and its derivatives at θ_0. Moreover, if $\pi(\theta)$ is, say, the uniform density on (0, 1), then the assumption of continuity at θ_0, needed even for (5.3), fails at $\theta_0 = 0, 1$. Of course for any prior supported on a bounded interval, (5.3) cannot hold for θ_0 equal to the endpoints of the support. It is these facts which lead to technical difficulties if we try to get a P_π result in a simple minded way from (5.4). It turns out that a P_π version exists if $\pi(\theta)$ is supported on a bounded interval and has smooth contacts at both endpoints, that is, $\pi(\theta)$ and $(d^i\pi(\theta))/d\theta^i$, $i = 1, 2, \dots, k+1$, are zero at both endpoints.

The proof of this is nontrivial, because, under these conditions on the prior, the constant M in (5.4) blows up as θ_0 tends to the endpoints. The detailed treatment in Ghosh, Sinha and Joshi (1982) is both very technical and tedious. We shall only reproduce later some of the conclusions we will need in dealing with third order efficiency in the general regular case.

We make one final remark about (5.4). So far we have assumed $\pi(\theta)$ is proper, that is, $\int \pi(\theta)\, d\theta = 1$. However, (5.3) and (5.4) continue to hold even if π is improper, provided the other assumptions hold and there is an n_0 such that the posterior given $(X_1, X_2, \dots, X_{n_0})$ is proper for all $(X_1, X_2, \dots, X_{n_0})$. If π is improper, P_π versions are not true. Throughout this chapter π is proper. Improper priors will be used in Chapters 8 and 9.

We now introduce classes of priors for which P_π versions are available.

DEFINITION 5.1. D_{k+2} is the class of priors $\pi(\theta)$ with support $[a, b]$, which are at least $(k-1)$ times continuously differentiable on $[a, b]$ and positive on (a, b) with

$$\begin{aligned}\pi^{(i)}(\theta) \equiv \frac{d^i\pi(\theta)}{d\theta^i} &= (\theta - a)^{k-i}(c_i + o(1)) \quad \text{as } \theta \downarrow a \\ &= (b - \theta)^{k-i}(c_i' + o(1)) \quad \text{as } \theta \uparrow b,\end{aligned}$$

$c_i, c_i' > 0$, $k \ge 2$. D_∞ is the class of infinitely differentiable priors which are positive on (a, b) and zero on $[a, b]^c$, monotone in a neighborhood of each endpoint and $\pi^{(i)}(\theta)\{\pi(\theta)\}^{V-1} \to 0$ as $\theta \to a, b$ $\forall\, 0 < V < 1$.

Take $a = 0$, $b = 1$. Then the following $\pi \in D_{K+2}$:

$$\begin{aligned}\pi(\theta) &= c\theta^{K-1}(1-\theta)^{K-1} \quad \text{on } (0,1) \\ &= 0 \quad \text{outside}(0,1).\end{aligned} \tag{5.5}$$

The following $\pi \in D_\infty$:

$$\begin{aligned}\pi(\theta) &= c\exp\left\{-\frac{1}{\theta(1-\theta)}\right\} \quad \text{on}(0,1) \\ &= 0 \quad \text{outside}(0,1).\end{aligned} \tag{5.6}$$

Even with such priors, the following is, in general, false. There is a finite, positive A such that

$$P_\pi\left\{\sup_h |F_n(h) - \Phi(h)| \le An^{-1/2}\right\} = 1 - O(n^{-r}) \qquad \forall\, r > 0 \tag{5.7}$$

[even when the regularity conditions needed for (5.4) hold].

Specifically assume the X's are i.i.d. $N(\theta, 1)$ and $\pi(\theta)$ satisfies (5.5) with $K = 6$. It is shown in Ghosh, Sinha and Joshi (1982) that (5.7) is false here.

To get a flavor of P_π versions that are true, consider a (linear) exponential

$$p(x|\theta) = c(\theta)\exp\{\theta f(x)\}A(x). \tag{5.8}$$

This satisfies all the regularity conditions on the family of densities for (5.4) to be true for all r, k. Then, writing $\Phi_{n,k}$ for the expansion appearing in (5.4),

$$\begin{aligned} p_\pi\left\{\sup_n |F_n(h) - \Phi_{n,k}(h)| < An^{-(K+1)/2 - \varepsilon}\right\} &= 1 - O(n^{-r}) \qquad \forall\, r \text{ if } \pi \in D_\infty \\ &= 1 - O(n^{-t_1}) - O(n^{-t_2}) \quad \text{if } \pi \in D_s,\ s > k+2, \end{aligned} \tag{5.9}$$

where

$$t_1 = \frac{(s+1)}{2}\left(\frac{s-k-3}{2} + \varepsilon\right), \qquad t_2 = \frac{(s+1)\varepsilon}{k+1}.$$

This is taken from Ghosh, Sinha and Joshi (1982), where references are given to similar work of Burnašev (1979) for a location parameter.

5.2. Expansion of the Bayes estimate and Bayes risk. We will need expansions for the (integrated) Bayes risk for squared error loss in the form

$$R_n(\pi) = a_1 n^{-1} + a_2 n^{-2} + o(n^{-2}), \tag{5.10}$$

where $R_n(\pi) = \pi \times P_\theta$-expectation of $(B - \theta)^2$, B = posterior mean $E(\theta|X_1, X_2, \dots, X_n)$ and a_1, a_2 do not depend on n.

We begin by noting that without smooth contact at the endpoints of the support, such an expansion need not exist. Take X_i's to be i.i.d. $N(\theta, 1)$, $\pi(\theta)$ the uniform density on (0, 1). It is plausible from the P_π version of (5.3), and is in fact proved in Ghosh, Sinha and Joshi (1982), that in this case

$$R_n(\pi) = \frac{1}{n} + o(n^{-1}). \tag{5.11}$$

Assuming (5.11), we now verify that (5.10) is false.

Consider an estimate

$$T_n = \bar{X} - c\frac{(\bar{X})^r}{n},$$

where c is a positive constant and r is a positive integer,

$$\underset{\text{def}}{=} \bar{X} + \frac{d(\bar{X})}{n},$$

$$E\left\{\left(\bar{X} + \frac{d(\bar{X})}{n} - \theta\right)^2 \middle| \theta\right\} = E\left[\left\{(\bar{X} - \theta) + \frac{d(\bar{X}) - d(\theta)}{n} + \frac{d(\theta)}{n}\right\}^2 \middle| \theta\right]$$

$$= \frac{1}{n} + \frac{d^2(\theta)}{n^2} + \frac{2d'(\theta)}{n^2} + o(n^{-2}),$$

which is easy to get by the delta method and justify rigorously; since $d(\bar{X})$ is a polynomial, one can write down the exact value of the left-hand side.

Now,

$$\int_0^1 (d^2(\theta) + 2d'(\theta))\, d\theta = c^2/(2r + 1) - 2c \to -\infty$$

if $c, r \to \infty$ such that c^2/r is bounded.

Suppose (5.10) is true [with $a_1 = 1$, by (5.11)]. Then

$$R_n(\pi) \le R(\pi, T_n)$$

implies, for each fixed c, r

$$a_2 \le c^2/(2r + 1) - 2c \to -\infty,$$

which is a contradiction.

THEOREM 5.1a. *Under regularity conditions stated in Ghosh, Sinha and Joshi* (1982) *and for* $\pi \in D_s$, $11 < s \le \infty$,

$$R_n(\pi) = \left\{\int_a^b \frac{1}{I(\theta)} \pi(\theta)\, d\theta\right\} n^{-1} + a_2 n^{-2} + o(n^{-2}), \tag{5.12}$$

where

$$a_2 = \int_a^b a_2(\theta)\pi(\theta)\, d\theta, \tag{5.13}$$

$$a_2(\theta) = \frac{d}{d\theta}\left(\frac{1}{I(\theta)} \frac{d}{d\theta} \log \frac{\pi(\theta)}{I(\theta)}\right)$$

$$+ \frac{I(\theta)}{\pi(\theta)} \frac{d}{d\theta}\left(\frac{1}{I(\theta)} \frac{d}{d\theta} \frac{\pi(\theta)}{I(\theta)}\right) \tag{5.14}$$

$+$ *a function* $\psi(\theta)$ *which does not depend on* π.

(The expression for a_2 given here agrees with that in Ghosh, Sinha and Joshi [(1982), page 427] after we add to the latter the term $2I^{-2}(\theta)\pi''(\theta)/\pi(\theta)$, which was dropped by mistake.)

THEOREM 5.1b. *Under the same conditions as in Theorem* 5.1a *and for* $\pi \in D_s$, $11 < s \leq \infty$,

$$P_\pi\{|E(\theta|X_1,\dots,X_n) - B_n| \leq Mn^{-(3/2+\varepsilon)}, \quad \text{and } |n^{-1}\pi'(\hat\theta)\pi(\hat\theta)| \leq Mn^{-\delta}\} = 1 - o(n^{-2}), \tag{5.15}$$

where B_n *is the expansion for the posterior mean due to Lindley* (1961) *and Johnson* (1970), *namely,*

$$B_n = \hat\theta - \lambda_n n^{-1}, \tag{5.16}$$

$$\lambda_n = \frac{a_{3,n} b^{-2}}{2} + \frac{b^{-1}\pi'(\hat\theta)}{\pi(\hat\theta)}, \tag{5.17}$$

$$a_{3,n} = \frac{1}{n}\frac{d^3 \log p(X_1, X_2, \dots, X_n|\theta)}{d\theta^3}\bigg|_{\hat\theta} \tag{5.18}$$

and b is defined in (5.1).

[See Ghosh, Sinha and Joshi (GSJ) (1982), page 425; note our a_{3n} is 6 times their $a_{3,n}$ and they miss a factor of 1/2; see (5.20), which agrees with GSJ, page 434.]

Note that the first term in the expression for λ_n can be shown to be bounded with probability $1 - o(n^{-2})$, but the second term is unbounded. The following result takes care of this, by truncating B_n to $\overline{B}_n = a$ or b according as B_n is less than a or exceeds b.

THEOREM 5.1c. *Under the same conditions as in Theorem* 5.1a, *a truncated version of B, namely* $\overline{B}_n$, *attains the Bayes risk up to* $o(n^{-2})$, *that is,*

$$R(\pi, \overline{B}_n) = R_n(\pi) + o(n^{-2}) \quad \textit{for } \pi \in D_2,\ 11 < s \leq \infty. \tag{5.19}$$

[See Ghosh, Sinha and Joshi (1982), pages 435 and 436.]

Finally we replace B_n and $\overline{B}_n$ by estimates depending on $\hat\theta$ only. To do this, we replace the terms in λ_n by functions of $\hat\theta$ which are close. Let

$$B'_n = \hat\theta + n^{-1}\left\{l_3\frac{(\hat\theta)}{2}I^{-2}(\hat\theta) + I^{-1}(\hat\theta)\frac{\pi'(\hat\theta)}{\pi(\hat\theta)}\right\}, \tag{5.20}$$

where

$$l_3(\theta) = E\left(\frac{d^3 \log p(X_1|\theta)}{d\theta^3}\bigg|\theta\right).$$

Let

$$\begin{aligned} \overline{B}'_n &= B'_n \quad \text{if } B'_n \in (a,b) \\ &= a \quad \text{if } B'_n < a \\ &= b \quad \text{if } B'_n > b. \end{aligned} \tag{5.21}$$

THEOREM 5.1d [Ghosh, Sinha and Joshi (1982), pages 434 and 435]. *Under the same conditions as in Theorem* 5.1a, $\overline{B}'_n$ *attains the Bayes risk up to*

$o(n^{-2})$, *that is*,

$$R(\pi, \bar{B}'_n) = R_n(\pi) + o(n^{-2}) \tag{5.22}$$

for $\pi \in D_s$, $11 < s \leq \infty$.

Theorem 5.1d is a somewhat surprising fact since $\hat{\theta}$ is not asymptotically sufficient up to the third order. Even from the point of view of zero third order information loss (Section 4.4), one needs $\hat{\theta}$ and b for some sort of third order sufficiency. An intuitive argument making Theorem 5.1d plausible appears in Ghosh and Subramanyan (1974), where Theorem 5.1d is conjectured. As indicated there, this result is at the heart of third order efficiency of $\hat{\theta}$.

We recall briefly the plausibility argument in favor of Theorem 5.1d. Note that $\bar{B}_n$ has been chosen so as to have the same bias up to $o(n^{-1})$. Suppose, by the delta method, both B_n and B'_n satisfy

$$E(B_n|\theta) = \theta + d(\theta)/n + o(n^{-1}), \tag{5.23}$$

$$E_\theta(B'_n|\theta) = \theta + d(\theta)/n + o(n^{-1}). \tag{5.24}$$

Then, again by the delta method, the mean squares are

$$\begin{aligned} E\{(B_n - \theta)^2|\theta\} &= E\{(\hat{\theta} - \theta)^2|\theta\} + \frac{d^2(\theta)}{n^2} + \frac{2d(\theta)b_0(\theta)}{n^2} \\ &\quad + \frac{2d'(\theta)}{n^2 I(\theta)} + o(n^{-2}) \\ &= E\{(B'_n - \theta)^2|\theta\}. \end{aligned} \tag{5.25}$$

To make these calculations rigorous expansions of the mean square (rather than the second moment of an Edgeworth expansion), one needs to truncate B_n, B'_n.

Note the following interesting fact. B'_n is a perturbation of $\hat{\theta}$ with two components, one of which is free of π. If we ignore the contribution from π (e.g., if we assume π is a constant over R), then the remaining part of $\bar{B}_n$ has expectation, from (5.20),

$$\begin{aligned} n^{-1}\left\{b_0(\theta) - \frac{J(\theta)}{2I^2(\theta)}\right\} &= n^{-1}\left\{\frac{\mu_{11}(\theta)}{I^2} + \frac{J(\theta)}{2I^2} + \frac{J(\theta)}{2I^2}\right\} \\ &= n^{-1}\left\{\frac{\mu_{11}(\theta)}{I^2(\theta)} + \frac{J(\theta)}{I^2(\theta)}\right\} \\ &= n^{-1}\left\{\frac{I'(\theta)}{I^2(\theta)}\right\}, \end{aligned} \tag{5.26}$$

which is zero for a location family. The expressions μ_{11} and J are defined in Theorem 3.1.

The expansions obtained here are identical to those of Kadane and Tierney (1986) up to $o(n^{-2})$. Their form is more convenient for numerical computations, whereas the present version seems more suitable for theoretical applications or algebraic manipulations.