CHAPTER 8

Magic Formula, Bartlett Correction and Matching Probabilities

8.1. Introduction. The magic formula of Barndorff-Nielsen (1983) is a beautiful formula in higher order asymptotics for calculating the density of the mle, or its conditional density given an ancillary statistic, which was called a magic formula because it seemed magical when it first appeared. It still retains some of its magical quality, though it is much better understood now than when it first appeared. For example it is still unclear why it is so good an approximation even for very small values of n like $n = 3$ or 4. Barndorff-Nielsen and Cox (1989) provide a nice exposition as well as many interesting applications. The magic formula is a special case of saddle point formulas. For an excellent introduction as well as exposition of recent results, see Reid (1988) and Field and Ronchetti (1990). In Section 8.2 we present the magic formula in the general case, along with a proof in a simple case, and pose a number of open problems.

Bartlett's correction is another formula in higher order asymptotics with a magical quality. We provide a Bayesian argument in Section 8.3 which is natural, general and rigorous. The proof makes clear that the correction is natural in the Bayesian context and the frequentist correction can be derived from this. In Section 8.4 we use the Bayesian argument to generate confidence sets which have the correct coverage probability of $1 - \alpha$ up to $O(n^{-2})$ (uniformly on compact θ sets). These sets have the attractive property of also having posterior probability of $1 - \alpha$ to $O(n^{-2})$ of covering the true value of θ. We show also how the frequentist Bartlett correction can be calculated through a Bayesian route.

8.2. The magic formula. Consider i.i.d. continuous r.v.'s $X_1, X_2, \dots, X_n$ with (linear) exponential density (with respect to Lebesgue measure)

$$p(x|\theta) = c(\theta)e^{\theta x}A(x).$$

Let $L(\theta) = L = \Pi_{i=1}^{n} p(x_i|\theta)$. Then the likelihood equation is

$$0 = \left.\frac{d \log L}{d\theta}\right|_{\hat{\theta}} = \frac{nc'(\hat{\theta})}{c(\hat{\theta})} + \sum_{1}^{n} X_i, \tag{8.1}$$

which is equivalent to

$$\mu(\hat{\theta}) = \bar{X}, \tag{8.2}$$

where $\mu(\theta) = E(X_1|\theta) = -c'(\theta)/c(\theta)$.

From (8.2),

$$b\, d\hat{\theta} = d\bar{X}, \tag{8.3}$$

where

$$b = -\left.\frac{d^2 \log c(\theta)}{d\theta^2}\right|_{\hat{\theta}}. \tag{8.4}$$

By sufficiency of $\bar{X}$,

$$\frac{p(\bar{x}|\theta)}{p(\bar{x}|\hat{\theta})} = \frac{p(x_1, x_2, \dots, x_n|\theta)}{p(x_1, x_2, \dots, x_n|\hat{\theta})}, \tag{8.5}$$

where in (8.5), $p(\bar{x}|\theta)$ stands for the density of $\bar{x}$ under θ. Hence

$$p(\bar{x}|\theta) = \frac{p(x_1, x_2, \dots, x_n|\theta)}{p(x_1, x_2, \dots, x_n|\hat{\theta})} p(\bar{x}|\hat{\theta}). \tag{8.6}$$

If we use the one-term formal Edgeworth expansion to evaluate the density of $\sqrt{n}(\bar{X} - \mu(\theta))$ at $\bar{X} = \bar{x}$ and put $\theta = \hat{\theta}(\bar{x})$, we get

$$p(\bar{x}|\hat{\theta}) = \frac{1}{\sqrt{2\pi nb}}(1 + O(n^{-1})). \tag{8.7}$$

The validity of (8.7) does not follow from Theorem 2.1 or results like that which provide valid expansions for probabilities only. Technically, we need a valid expansion of the density in the sup-norm rather than the L_1-norm, and some uniformity in θ to justify the substitution of $\theta = \hat{\theta}$. Necessary assumptions are stronger than those of Theorem 2.1.

If we assume (8.7) as valid and substitute in (8.6), we get

$$p(\bar{x}|\theta) = \frac{p(x_1, x_2, \dots, x_n|\theta)}{p(x_1, x_2, \dots, x_n|\hat{\theta})} \frac{1}{\sqrt{2\pi nb}}\{1 + O(n^{-1})\}. \tag{8.8}$$

If we now use (8.3) in (8.8) to switch from $\bar{x}$ to $\hat{\theta}$, we get a special case of the magic formula:

$$p(\hat{\theta}|\theta) = \frac{p(x_1, x_2, \dots, x_n|\theta)}{p(x_1, x_2, \dots, x_n|\hat{\theta})} \frac{\sqrt{b}}{\sqrt{2\pi n}}\{1 + O(n^{-1})\}. \tag{8.9a}$$

If we do not assume exponential density, $\hat{\theta}$ is not sufficient and it will not be sufficient even up to $O(n^{-1})$, so formula (8.9a) cannot hold for nonexponential densities without a suitable modification. It is clear we should have on the left-hand side of (8.9a), in addition to $\hat{\theta}$, additional statistics, which together with $\hat{\theta}$ ensure sufficiency up to $O(n^{-1})$. It turns out that if one chooses ancillary statistics suitably, this can be done and the magic formula in the general case will look like

$$(8.9\text{b})\quad p(\hat{\theta}|\text{ancillary}, \theta) = \frac{p(x_1, x_2, \ldots, x_n|\theta)}{p(x_1, x_2, \ldots, x_n|\hat{\theta})} \frac{\sqrt{b}}{\sqrt{2\pi n}}(1 + O(n^{-1})).$$

If one believes, as recommended by Fisher, on working in the framework of conditional inference, where an ancillary statistic is held fixed, then (8.9b) seems to be the right tool to use. Note that (8.9b) is very easy to use since $p(x_1, x_2, \ldots, x_n|\theta)$ is easy to calculate for i.i.d. r.v.'s.

If $p(x|\theta)$ is a location family $g(x_1 - \theta)$, and the ancillary is what Fisher recommended here, that is, $(x_2 - x_1, x_3 - x_1, \ldots, x_n - x_1)$, then (8.9b) is exact, and Fisher knew this formula. Barndorff-Nielsen has shown his formula is exact when a (locally compact) group of transformations leaves the family $\{p(x|\theta), \theta \in \Theta\}$ invariant and $\hat{\theta}$ and the maximal invariant are together a one-to-one function of $x = (x_1, x_2, \ldots, x_n)$.

In many examples, by the right choice of an ancillary statistic and renormalization of the right-hand side of (8.9b) to make it integrate to 1, the error can be made $O(n^{-3/2})$.

We conclude with a few open questions. Only the second question is new.

1. Characterize multiparameter exponentials for which the formula is exact. The one-parameter exponentials for which the formula is exact were characterized by Daniels.
2. Are solutions to question 1 always infinitely divisible? At least for infinitely divisible multiparameter exponentials, solve question 1.
3. Why are the formulas so good even for small n (sometimes as small as 3 or 4)?

8.3. Bayesian and frequentist Bartlett correction. Consider two nested null hypotheses on a two-parameter density $p(x_1, x_2, \ldots, x_n|\theta_1, \theta_2)$, the case of k parameters being exactly similar. Let $L(\theta) = p(X_1, X_2, \ldots, X_n|\theta)$. Consider

$$(8.10)\qquad H_1\colon \theta_1 = \theta_{10}, \qquad H_2\colon \theta_1 = \theta_{10}, \qquad \theta_2 = \theta_{20}.$$

Let $\hat{\theta}$ be the (unrestricted) mle of θ and $\hat{\theta}(\theta_{10})$ the mle of θ assuming H_1 is true. Let

$$(8.11)\qquad \begin{aligned} \lambda_1(\theta_{10}) &= \text{(log) likelihood ratio statistic for testing } H_1 \\ &= 2\left[\log L(\hat{\theta}) - \log L(\hat{\theta}(\theta_{10}))\right], \end{aligned}$$

$$(8.12)\qquad \begin{aligned} \lambda_2(\theta_{10}, \theta_{20}) &= \text{(log) likelihood ratio statistic for testing } H_2 \\ &= 2\left[\log L(\hat{\theta}) - \log L(\theta_0)\right], \end{aligned}$$

where $\theta_0 = (\theta_{10}, \theta_{20})$. In the context of H_1, θ_2 is a nuisance parameter.

Here are some basic facts.

1. Under standard regularity conditions, that is, Assumption A_1 of Chapter 1, λ_i converges in distribution to χ_i^2 under H_i, where χ_i^2 is a χ^2 with i d.f.
2. Under stronger regularity conditions, suitable moment assumptions and Condition D (or C) of Chapter 2 [see Chandra and Ghosh (1979)] there is a valid asymptotic expansion for the probability that $P(\lambda_i \in B|\theta)$ (where θ satisfies H_i). There is an expansion of the density, which is to be integrated over the Borel set B, the Borel set B satisfying suitable conditions on its boundary if Condition C, rather than Condition D, holds. The expansion of the density is of the form

$$f_i(\chi_i^2)\left[1 + \frac{(\;)}{n} + \frac{(\;)}{n^2} + o(n^{-2})\right], \tag{8.13}$$

where f_i is the density of χ^2 with i d.f. and the coefficient ( ) of n^{-r} is a polynomial in χ_i^2 whose coefficients do not depend on n.
3. Let $E(\lambda_i|H_i) = i + b_i/n + O(n^{-2})$ as calculated from (8.13); see the first interpretation of such expansions in Section 2.7. The random variable

$$\lambda'_i \underset{\text{def}}{=} i\lambda_i/(i + b_i/n) = \lambda_i/(1 + b_i/in)$$

is χ_i^2 up to $o(n^{-1})$ in the sense that

$$P(\lambda'_i \in B|H_i) = P\{\chi_i^2 \in B\} + O(n^{-2}) \tag{8.14}$$

under the same assumptions as in the previous paragraph. The transformation $\lambda_i \to \lambda'_i$ is the (frequentist) Bartlett correction and was first proposed by Bartlett (1937) more than 50 years ago. For some of the history of this, see Bickel and Ghosh (1990). Some of the early references are Bartlett (1937), Box (1949) and Lawley (1956). That a valid expansion of the probability in the following sense is true follows immediately from (8.13) and Lemma 2.1:

$$P\{\lambda'_i \in B|H_i\} = \int_B f_i(\chi_i^2)\left(1 + \frac{(\;)}{n} + \frac{(\;)}{n^2}\right) d\chi_i^2 + O(n^{-2}), \tag{8.15}$$

where the coefficient of $n^{-r}(\;)$ is a polynomial in χ_i^2, with coefficients free of n, and the polynomials are different from those in (8.13) and, in principle, can be calculated from (8.13) through tedious algebra. Through long and tricky cumulant calculations, Lawley (1956) showed the coefficient of n^{-1} in (8.15) is zero. This and (8.15) show the truth of (8.14).

It is surprising that merely adjusting the bias produces such dramatic improvement in the χ_i^2 approximation, from an error of $O(n^{-1})$ according to (8.13) to an error of $O(n^{-2})$ according to (8.14). Such a correction does not exist for the two most popular competitors of the likelihood ratio statistic, namely, Rao's score statistic, which is a quadratic form in $\partial \log L/\partial\theta_i|_{\theta_0}$,

$i = 1, 2$, for H_2, and Wald's statistic, which is a quadratic form in $\hat{\theta}$. On the other hand, facts 1 and 2 hold for them; see Chandra and Ghosh (1979).

Lawley's proof is tricky and computational, but does not throw light on why one should expect a relation like (8.14) to hold for the likelihood ratio statistic rather than Rao's or Wald's statistic. In this context, the following lemma is of interest. We omit the simple proof.

LEMMA 8.1. *Equation* (8.15) *is true if and only if the coefficient of* n^{-1} *in* (8.13) *is linear in* χ_i^2.

More recently Barndorff-Nielsen and Cox (1984) prove (8.14) using Barndorff-Nielsen's magic formula. However, the magic formula has not been proved rigorously in all cases and a rigorous proof of a saddle point formula seems to require rather strong assumptions, as indicated even for the simplest case [see (8.6) through (8.8)].

We reproduce below the Bayesian argument of Bickel and Ghosh (1990), which is both rigorous and seems to make clear intuitively why the correction works. We will explain only the main ideas and in the process also develop a Bayesian Bartlett correction based on the posterior distribution of λ_1, λ_2 given $X_1, X_2, \dots, X_n$. It turns out that it is relatively easy to see why the posterior distribution of λ should have the structure in Lemma 8.1.

STEP 1. Use signed square roots of the likelihood ratio statistics with 1 d.f.:

$$T_1 = \{\lambda_1(\theta_{10})\}^{1/2}\operatorname{sign}(\hat{\theta}_1 - \theta_{10}), \tag{8.16}$$

$$T_2 = \{\lambda_2 - \lambda_1(\theta_{10})\}^{1/2}\operatorname{sign}(\hat{\theta}_2(\theta_{10}) - \theta_{20}). \tag{8.17}$$

Note that $T_2^2 = (\lambda_2 - \lambda_1(\theta_{10}))$ is the likelihood ratio statistic for

$$H_3\colon \theta_2 = \theta_{20}$$

assuming $H_1\colon \theta_1 = \theta_{10}$ is true.

The signed square roots, being asymptotically normal, are convenient to work with in many problems involving r.v.'s which have asymptotically χ^2 distribution. Their use goes back to Lawley (1956); see also Chandra and Joshi (1983). For the justification of Bartlett correction, T_i's are convenient, but not essential. One reason for considering T_i's in Bickel and Ghosh (1990) was to prove an extension of a result of Efron (1985) and possibly others (folk theorem?): The normal approximation to T_i's is correct to a higher order than the normal approximation to $\sqrt{n}(\hat{\theta} - \theta)$. Various people, including Efron, have recommended the use of T_i's or λ_i's, rather than the mle, for setting up confidence intervals.

STEP 2. Note that T_i's are functions of θ_{10}, θ_{20} as well as $X_1, X_2, \dots, X_n$. We now treat θ_{10}, θ_{20} as r.v.'s with a prior distribution π and write θ for θ_0. Let $\pi \otimes P_\theta$ stand, as before, for the joint distribution of θ and X's and $P = P_\pi$ for the marginal distribution of the X's.

STEP 3. The posterior density of θ_1, θ_2 is proportional to

$$(8.18) \quad \pi(\theta)\exp\{\log L(\theta) - \log L(\hat{\theta})\} = \pi(\theta)\exp\{-\tfrac{1}{2}T_1^2 - \tfrac{1}{2}T_2^2\}.$$

The posterior density of $\sqrt{n}(\theta - \hat{\theta})$ is also proportional to (8.18) and the posterior density of T_1, T_2 (given $X_1, X_2, \ldots, X_n$) is proportional to

$$(8.19) \qquad \pi(\theta)\exp\{-\tfrac{1}{2}T_1^2 - \tfrac{1}{2}T_2^2\} J,$$

where J is the Jacobian of transformation to T_1, T_2 from θ or $\sqrt{n}(\theta - \hat{\theta})$.

It is easy, though tedious, to show by a Taylor expansion, that is, the delta method, that $T = (T_1, T_2)$ is a nonsingular linear transformation of $\sqrt{n}(\theta - \hat{\theta}) = \sqrt{n}(\theta_1 - \hat{\theta}_1, \theta_2 - \hat{\theta}_2)$ plus terms of order $(\sqrt{n})^{-1}$. Since this is obtained by Taylor expansion, it is clear that the coefficient of $(\sqrt{n})^{-r}$ is a polynomial in $\sqrt{n}(\theta_i - \hat{\theta}_i)$ of one degree more than r. Since calculation of the Jacobian involves a differentiation, the Jacobian J is a constant plus expansion in powers of $(\sqrt{n})^{-1}$ with coefficient $(\sqrt{n})^{-r}$ a homogeneous polynomial in $\sqrt{n}(\theta_i - \hat{\theta}_i)$ of the same degree r. Reexpressing this in terms of T, it can be seen that the Jacobian J is a constant plus an expansion in powers of $(\sqrt{n})^{-1}$, with the coefficient of $(\sqrt{n})^{-r}$ a homogeneous polynomial in T_1, T_2 of degree r.

If one expands $\pi(\theta)$ around $\hat{\theta}$, then again one gets $\pi(\hat{\theta})$ plus an expansion in powers of $(\theta_i - \hat{\theta}_i)$. Hence, if one writes this in terms of $(\sqrt{n}(\theta_i - \hat{\theta}_i)/\sqrt{n})$, one gets $\pi(\hat{\theta})$ plus an expansion in powers of $(\sqrt{n})^{-1}$ with coefficients that are polynomials in $\sqrt{n}(\theta_i - \hat{\theta}_i)$ of the same degree. Finally, once again switching to T, one sees, as in the case of the Jacobian, that the degree of a coefficient polynomial matches the power of $(\sqrt{n})^{-1}$.

From the above considerations it follows that a formal expansion up to $O(n^{-2})$ of the numerator in the posterior of (T_1, T_2) has the form

$$\begin{aligned}
&\pi_3'(t, x_1, x_2, \ldots, x_n) \\
&\quad = \exp(-\tfrac{1}{2}t_1^2 - \tfrac{1}{2}t_2^2)\{1 + \text{a polynomial of degree 3 in } t_1/\sqrt{n}, t_2/\sqrt{n}\} \\
&\qquad + O(n^{-2}).
\end{aligned}$$

If one integrates out t and divides, then one gets a formal expansion of the posterior as

$$\begin{aligned}
&\pi_3(t, x_1, x_2, \ldots, x_n) \\
(8.20) \qquad &= \phi(t)\{1 + P_3(n^{-1/2}, x_1, x_2, \ldots, x_n) \\
&\qquad + Q_3(n^{-1/2}t_1, n^{-1/2}t_2, x_1, x_2, \ldots, x_n)\} + O(n^{-2}),
\end{aligned}$$

where P_3 is a cubic in $(\sqrt{n})^{-1}$ and Q_3 is a cubic in $t_1/\sqrt{n}$, $t_2/\sqrt{n}$.

The expansion (8.20) is justified in the following theorem in Bickel and Ghosh (1990).

THEOREM 8.1. *Under regularity conditions on the density and conditions on π as in Chapter* 5,

$$(8.21) \quad \int |\pi(t|x_1, x_2, \ldots, x_n) - \pi_3(t|x_1, x_2, \ldots, x_n)|\, dt = O(n^{-2}) \quad a.s.\ P_\pi$$

(i.e., the L_1-distance between the posterior and its formal expansion goes to zero like n^{-2}).

STEP 4. Using Theorem 8.1, one can show that the posterior density of $\lambda_2 = (T_1^2 + T_2^2)$ is of the form

$$f_2(\chi_2^2)\left[1 + \frac{(\ )}{n} + O(n^{-2})\right], \tag{8.22}$$

where ( ) is linear in χ_2^2 and similarly for λ_1. It follows from Lemma 8.1 that the posterior density of

$$\lambda_i^* = \frac{i\lambda_i}{1 + b_i^*/n} \tag{8.23}$$

can be approximated by a χ_i^2 density up to $O(n^{-2})$; here

$$E(\lambda_i | x_1, x_2, \ldots, x_n) = i + b_i^*(x_1, x_2, \ldots, x_n)/n + O(n^{-2})$$

as calculated from (8.22). This is what we call the Bayesian Bartlett correction.

STEP 5. Making π converge weakly to the measure putting all the mass at θ_0 (as in the proof of third order efficiency in Chapter 6), and using (8.22), it can be shown that the density of λ_i^2 under θ_0 satisfies the condition of Lemma 8.1 and, hence, the frequentist Bartlett correction works also.

One may ask why such an argument would fail for Rao's or Wald's statistic. To get the posterior density for λ_1, λ_2 (via T_1, T_2), we had to expand only $\pi(\theta)$ and the Jacobian, but not

$$\exp\{\log L(\theta) - \log L(\hat{\theta})\} = \exp\{-\tfrac{1}{2}(T_1^2 + T_2^2)\} = \exp\{-\lambda_2/2\}.$$

However, for Rao's or Wald's statistic, one has also to expand the exponential term; hence, the posterior of those statistics will not have the structure required by Lemma 8.1.

One can derive an elegant formula for the quantity $b_i = b_i(\theta_0)$ appearing in the frequentist Bartlett correction by making π converge weakly to the probability measure δ_{θ_0} sitting on θ_0; see Ghosh and Mukerjee (1991) and the next section for details. A similar argument appears in Dawid (1991).

8.4. Matching probabilities and confidence sets. We indicate how the Bayesian and frequentist Bartlett correction can be used to choose a prior π such that the confidence set obtained from λ_1 or λ_2 has the same frequentist and posterior probability of covering the true value up to $O(n^{-2})$. This topic will be discussed in more detail in the next chapter.

We will need the following analogue of Theorem 8.1, which can be obtained in a similar way.

THEOREM 8.2. *Under regularity conditions on $p(x|\theta)$, and the assumption that π is positive and four times continuously differentiable,*

$$\int |\pi(t|x_1, x_2, \dots, x_n) - \pi_3(t|x_1, x_2, \dots, x_n)\, dt = O(n^{-2}) \quad a.s.\,(P_\theta). \tag{8.24}$$

Compare with Johnson (1970) and Ghosh, Sinha and Joshi (1982) and note that one can have versions of this which are uniform on compact sets of θ. The difference between Theorems 8.1 and 8.2 consists in replacing $\pi \otimes P_\theta$ by P_θ.

We illustrate only with λ_1, that is, we seek a confidence set (or interval) only for the parameter of interest θ_1, treating θ_2 as a nuisance parameter.

Fix a prior π and consider

$$\begin{aligned} A_{1-\alpha}(x_1, x_2, \dots, x_n) = \Bigg\{\theta_1^0; \lambda_1(x_1, x_2, \dots, x_n, \theta_1^0) \\ \le \chi^2_{1,\alpha}\left(1 + \frac{b_1^*(x_1, x_2, \dots, x_n)}{n}\right)\Bigg\}, \end{aligned} \tag{8.25}$$

where $\chi^2_{1,\alpha}$ is the $100(1-\alpha)\%$ point of a χ^2 with 1 d.f. and b_1^* is the quantity appearing in the Bayesian Bartlett correction in Step 4 in the previous section. Then, by Theorem 8.2,

$$\text{posterior probability of } \theta_1 \in A_{1-\alpha} = 1 - \alpha + O(n^{-2}). \tag{8.26}$$

Now choose π and, hence, $A_{1-\alpha}$, such that the frequentist probability

$$P_{\theta_1, \theta_2}\{\theta \in A_{1-\alpha}(x_1, x_2, \dots, x_n)\} = 1 - \alpha + O(n^{-2}) \tag{8.27}$$

(uniformly on compact sets of θ).

The point of doing this is that a Bayesian using such a prior will be in close agreement with the frequentist about the coverage probability. In some sense such a prior may be called noninformative. The idea of matching frequentist and posterior probability, which goes back to Welch and Peers (1963), is discussed in more detail in the next chapter.

We will need the marginal posterior of T_1. The proof in Bickel and Ghosh (1990) is not constructive and does not lead to explicit formulas. One has, therefore, to use expansions of posteriors of $\sqrt{n}(\theta - \hat{\theta})$ (see Chapter 5) and switch to T_1, $\sqrt{n}(\theta_2 - \hat{\theta}_2)$ and integrate out $\sqrt{n}(\theta_2 - \hat{\theta}_2)$. For this purpose it is convenient to have orthogonality of θ_1 and θ_2. We will need the concept of orthogonality also in the next two chapters. The idea of orthogonal parameters seems to be owing to Huzurbazar [see Huzurbazar (1992)]. The following treatment is based on Cox and Reid (1987).

The parameters θ_1 and θ_2 are orthogonal if the Fisher information matrix is diagonal for all θ. Given a scalar parameter of interest, θ_1, and a nuisance parameter of arbitrary dimension, in general, one can reparametrize θ_2 to have θ_1 and $\eta(\theta_1, \theta_2)$, where θ_1 and η are orthogonal. Let $\theta_2 = \theta_2(\theta_1, \eta)$.

Then the function θ_2 must satisfy

$$I_{22}(\theta)\frac{\partial\theta_2}{\partial\theta_1} = -I_{12}(\theta). \tag{8.28}$$

Here is an example from Cox and Reid (1987).

EXAMPLE 8.1. $X_1 = (Y_1, Y_2), Y_1, Y_2$ are independent exponential with means θ_2 and $\theta_1\theta_2$. The ratio of means is the parameter of interest. The equation (8.28) reduces to

$$\frac{2}{\theta_2^2}\frac{\partial\theta_2}{\partial\theta_1} = -\frac{1}{\theta_1\theta_2}, \tag{8.29}$$

which is equivalent to

$$\frac{\partial\theta_2}{\theta_2} = -\frac{\partial\theta_1}{2\theta_1}, \tag{8.30}$$

and the solution is

$$\log\theta_2 + \tfrac{1}{2}\log\theta_1 = \eta \quad (\text{say}) \tag{8.31}$$

so that

$$\eta = \theta_2(\theta_1)^{1/2}. \tag{8.32}$$

If we assume, as we may without loss of generality, that θ_1, θ_2 are orthogonal, then, after some algebra, we get the marginal of T_1 as

$$\begin{aligned}\pi(t_1|x_1, x_2, \dots, x_n) \\ = \phi(t_1)\big[1 + n^{-1/2}(G_1)t_1 + n^{-1}(G_2)(t_1^2 - 1) \\ +n^{-3/2}\{\text{a cubic in } t_1\} + O(n^{-2})\big],\end{aligned} \tag{8.33}$$

where G_1, G_2 depend on π and its derivatives at $\hat{\theta}$. [The corresponding marginal posterior for $h_1 = \sqrt{n}(\theta_1 - \hat{\theta}_1)$ involves a cubic in t_1 as coefficient of $n^{-1/2}$ and a polynomial of degree 6 in h_1 as coefficient of n^{-1}.] Explicit formulas are given in Ghosh and Mukerjee (1992a).

Using (8.33), one can calculate b_1^* from

$$E(T_1^2|x_1, x_2, \dots, x_n) = 1 + b_1^*/n + O(n^{-2}). \tag{8.34}$$

One then chooses π such that

$$b_1^* = b_1 + o_p(1), \tag{8.35}$$

where b_1 is the frequentist Bartlett correction.

We provide some details of the calculations, illustrating in the process how the frequentist Bartlett correction b_1 can be calculated in a Bayesian way.

We first note that integrating t_1^2 with respect to the right-hand side, we get

$$b_1^* = 2G_2.$$

We write down G_2 explicitly below, after introducing some notations. By orthogonality,

$$(8.36) \qquad I \equiv \text{Fisher information matrix} = \begin{bmatrix} I_{20} & I_{11} = 0 \\ 0 & I_{02} \end{bmatrix}.$$

Let $L = \log p(x_1, x_2, \ldots, x_n|\theta)$ and let

$$(8.37) \qquad l_{ij}(\hat{\theta}) = \frac{1}{n}\left(\frac{\partial^{i+j} \log L}{\partial\theta_1^i\, \partial\theta_2^j}\right)\bigg|_{\hat{\theta}},$$

$$(8.38) \qquad \pi_{ij}(\theta) = \frac{\partial^{i+j}\pi(\theta)}{\partial\theta_1^i\, \partial\theta_2^j},$$

$$(8.39) \qquad D = -l_{20}(\hat{\theta}) + \left(l_{02}(\hat{\theta})\right)^{-1}\left(l_{11}(\hat{\theta})\right)^2,$$

$$(8.40) \qquad K_{ij} = E\left\{\frac{\partial^{i+j} \log p(x_1|\theta)}{\partial\theta_1^i\, \partial\theta_2^j}\bigg|\theta\right\},$$

$$(8.41) \qquad K_{ij\cdot i'j'} = E\left[\left\{\frac{\partial^{i+j} \log p(x_1|\theta)}{\partial\theta_1^i\, \partial\theta_2^j}\right\}\left\{\frac{\partial^{i'+j'} \log p(x_1|\theta)}{\partial\theta_1^{i'}\, \partial\theta_2^{j'}}\right\}\bigg|\theta\right].$$

$K_{ij\cdot i'j'\cdot i''j''}$, and so forth, are similarly defined. We will need the following relation (in proving which one uses differentiation under the integral sign):

$$(8.41a) \qquad \frac{\partial}{\partial\theta_1}K_{ij} = K_{ij\cdot 10} + K_{i+1j}$$

$$(8.41b) \qquad \frac{\partial}{\partial\theta_1}I_{20}^{-1} = I_{20}^{-2}(K_{10\cdot 20} + K_{30})$$

Then

$$(8.42) \qquad \begin{aligned} G_2 = {} & \tfrac{1}{8}D^{-2}\Big\{l_{40}(\hat{\theta}) + \tfrac{5}{3}D^{-1}l_{30}^2(\hat{\theta}) - 3\left(l_{02}(\hat{\theta})\right)^{-1}\left(l_{21}(\hat{\theta})\right)^2 \\ & \qquad -2\left(l_{02}(\hat{\theta})\right)^{-1}l_{12}(\hat{\theta})l_{30}(\hat{\theta})\Big\} \\ & +\tfrac{1}{8}\Big\{D\left(l_{02}(\hat{\theta})\right)^2\Big\}^{-1}\left\{2l_{03}(\hat{\theta})l_{21}(\hat{\theta})\right\} \\ & +3\Big\{l_{12}(\hat{\theta})\big)^2\Big\} - \tfrac{1}{4}\left(Dl_{02}(\hat{\theta})\right)^{-1}l_{22}(\hat{\theta}) + \tfrac{1}{2}\left(D\pi(\hat{\theta})\right)^{-1} \\ & \times\Big\{\pi_{20}(\hat{\theta}) - \left(l_{02}(\hat{\theta})\right)^{-1}\left(\pi_{10}(\hat{\theta})l_{12}(\hat{\theta}) + \pi_{01}(\hat{\theta})l_{21}(\hat{\theta})\right) \\ & \qquad +D^{-1}\pi_{10}(\hat{\theta})l_{30}(\hat{\theta})\big)\Big\} + o_p(1) \end{aligned}$$

$$(8.43) \qquad \begin{aligned} & = \hat{\psi}_1 + \tfrac{1}{2}(D\hat{\pi})^{-1}\Big\{\hat{\pi}_{20} - \hat{l}_{02}^{-1}\left(\hat{\pi}_{10}\hat{l}_{12} + \hat{\pi}_{01}\hat{l}_{21}\right) + D^{-1}\hat{\pi}_{10}\hat{l}_{30}\Big\} \\ & \quad + o_p(1), \end{aligned}$$

where $\hat{\psi}_1$ is free from π (and its derivatives) and $\hat{\pi}_{ij} = \pi_{ij}(\hat{\theta}), \hat{l}_{ij} = l_{ij}(\hat{\theta})$.

Under $\theta = \theta_0$,

$$\begin{aligned} b_1^* &= 2\psi_1(\theta_0) + (\pi(\theta_0))^{-1} \\ &\quad \times \left\{ I_{20}^{-1}\pi_{20}(\theta_0) + (I_{20}I_{02})^{-1}(\pi_{10}(\theta_0)K_{12} + \pi_{01}(\theta_0)K_{21}) \right. \\ &\qquad \left. + I_{20}^{-2}K_{30}\pi_{10}(\theta_0) \right\} + o_p(1) \qquad (8.44) \\ &= 2\psi_1(\theta_0) + \psi_2(\theta_o, \pi) + o_p(1) \qquad (8.45) \\ &\underset{\text{def}}{=} \tilde{b}_1(\theta_0) + o_p(1). \end{aligned}$$

Here $\psi_1(\theta_0)$ is the limit of $\hat{\psi}_1$ in probability, and the limit is easy to write down using $l_{ij}(\hat{\theta}) \to_p K_{ij}$, $\pi_{ij}(\hat{\theta}) \to_p \pi_{ij}(\theta_0)$, and so forth. The part depending on π is ψ_2.

To calculate the frequentist Bartlett correction $b_1(\theta_0)$ from $\tilde{b}$, we proceed as follows. For the time being only, we regard π as an auxiliary prior satisfying the conditions of Bickel and Ghosh (1990) (or analogous to those of Chapter 5 at $a = \theta_0 - \delta$, $b = \theta_0 + \delta$) and concentrate on a rectangle with vertices at $\theta_{10} \pm \delta$, where $\delta > 0$ will eventually tend to zero. (For example, the product of two marginal priors of θ_1, θ_2 satisfying the conditions of Chapter 5 will do.) To make the dependence of π on δ clear, let us denote π by π_δ and $\tilde{b}_1$ by $\tilde{b}_{1\delta}$ at this point. Then, using a double integral over the rectangular support of π_δ,

$$b_1(\theta_0) = \lim_{\delta \downarrow 0} \int_{\theta_0 - \delta}^{\theta_0 + \delta} \tilde{b}_{1\delta}(\theta)\pi_\delta(\theta)\, d\theta. \tag{8.46}$$

The idea is to take the expectation of b_1^* with respect to $\pi_\delta \otimes P_\theta$ and then make $\delta \downarrow 0$. The expectation is taken by first taking expectation given θ, getting $\tilde{b}_1$ and then integrating out θ. We indicate how one calculates the limit on the right-hand side of (8.46). Note ψ_1 does not involve π_δ and is a continuous function of θ. Hence,

$$\lim_{\delta \downarrow 0} \int \psi_1(\theta)\pi_\delta(\theta)\, d\theta = \psi_1(\theta_0). \tag{8.47}$$

For the terms ψ_2 in $\tilde{b}_\delta$ which involve π_δ, we have to take recourse to integration by parts before taking the limit. Thus,

$$\lim_{\delta \downarrow 0} \int I_{20}^{-1} \frac{\pi_{20,\delta}}{\pi_\delta} \cdot \pi_\delta\, d\theta = \lim_{\delta \downarrow 0} \left[\int \frac{\partial^2}{\partial\theta_1^2}(I_{20}^{-1})\pi_\delta\, d\theta \right] = \left. \frac{\partial^2 I_{20}^{-1}(\theta)}{\partial\theta_1^2} \right|_{\theta_0}. \tag{8.48}$$

We get finally, after using the results (8.41a) and (8.41b),

$$b_1(\theta_0) = 2\psi_1(\theta_0) + \psi_3(\theta_0), \tag{8.49}$$

where

$$\text{(8.49a)}\qquad \begin{aligned}\psi_3(\theta_0) = {}& \frac{\partial^2}{\partial\theta_1^2} I_{20}^{-1} - \frac{\partial}{\partial\theta_1}\left\{(I_{20}I_{02})^{-1}K_{12} + I_{20}^{-1}K_{30}\right\}\\ & - \frac{\partial}{\partial\theta_2}\left\{(I_{20}I_{02})^{-1}K_{21}\right\}.\end{aligned}$$

It is a remarkable fact that ψ_3 is often zero, which corresponds to the fact that the uniform prior satisfies (8.50) below.

We now determine π by matching $b_1(\theta_0)$ and $\tilde{b}_1(\theta_0)$ for all θ_0, that is, from the differential equation

$$\text{(8.50)}\qquad \psi_2(\theta, \pi) = \psi_3(\theta),$$

which may be written in the following form, using the fact $(\partial/\partial\theta)I_{20}^{-1} = I_{20}^{-2}(K_{10\cdot 20} + K_{30})$ [see (8.41b)]:

$$\text{(8.51)}\qquad \frac{\partial}{\partial\theta_1}\left\{\frac{\pi_{10}(\theta)}{I_{20}(\theta)} - \frac{K_{10\cdot 20}\pi(\theta)}{I_{20}^2} + \frac{K_{12}\pi(\theta)}{I_{20}I_{02}}\right\} + \frac{\partial}{\partial\theta_2}\left\{\frac{K_{21}\pi(\theta)}{I_{20}I_{02}}\right\} = 0.$$

The assumptions needed for deriving such equations as well as the interpretation to be attached are discussed in Section 9.4. In the present case, we need the conclusion of Theorem 8.2 to be true and (8.13) to be valid. For Theorem 8.2 to be true, we need conditions of Johnson (1970) for posterior expansion up to $o(n^{-1})$. For (8.13), the Edgeworth assumptions in Section 2.6 suffice; see Chandra and Ghosh (1979).

EXAMPLE 8.2. Let x_i's be i.i.d. normal with mean θ_2 and variance θ_1, that is, θ_1 is the parameter of interest. Then the solution of (8.51) is

$$\text{(8.52)}\qquad \pi(\theta) = d_1(\theta_2)\theta_1^{-1} + d_2(\theta_2)\theta_1^{-3},$$

where d_1, d_2 are arbitrary functions. In particular, $\pi = d(\theta_2)\theta_1^{-1}$ satisfies (8.52). The common prior for this, namely [the prior induced by the right Haar measure for $(\theta_1^{1/2}, \theta_2)$ under the affine group of transformation],

$$\pi(\theta) = \theta_1^{-1},$$

satisfies this but not the Jeffreys measure (see the next chapter for its definition)

$$\pi(\theta) = \theta_1^{-3/2}$$

[which is induced by the left Haar measure for $(\theta_1^{1/2}, \theta_2)$].