Example 4.17. Suppose that $X \sim N(\theta, 1)$ given $\Theta = \theta$ and $\Omega_H = \{\theta_0\}$, $\Omega_A = (-\infty, \theta_0) \cup (\theta_0, \infty)$. Let the prior probability of the hypothesis be $\Pr(\Theta = \theta_0) = p_0 > 0$. Suppose that the conditional prior distribution of Θ given $\Theta \neq \theta_0$ is a measure λ. It is not difficult to show that

$$\frac{p_1}{1-p_1} = \frac{p_0}{1-p_0} \exp\left(-\frac{\theta_0^2}{2}\right) \left[\int \exp\left(x[\theta-\theta_0] - \frac{\theta^2}{2}\right) d\lambda(\theta)\right]^{-1}.$$

If λ puts positive mass on both sides of θ_0, then it is easily verified that $\int \exp(x[\theta-\theta_0] - \theta^2/2) d\lambda(\theta)$ is convex as a function of x and goes to ∞ as $x \to \pm\infty$. So all formal Bayes rules will be of the form "reject H if x is outside of some bounded interval."

When testing hypotheses of the form $H : \Theta = \theta_0$, the formal Bayes rule can be written in the form "reject H if the Bayes factor is less than something." It is possible to bound the Bayes factor from below when the likelihood function is bounded above. That is, we might be able to find a distribution λ that would lead to the smallest possible Bayes factor.[5] This lower bound would give a bound on how much the probability of the hypothesis could decrease from the prior to the posterior.

Example 4.18 (Continuation of Example 4.17). The Bayes factor in this example is

$$\exp\left(-\frac{\theta_0^2}{2}\right) \left[\int \exp\left(x[\theta-\theta_0] - \frac{\theta^2}{2}\right) d\lambda(\theta)\right]^{-1},$$

which is minimized (over λ) by the distribution that puts probability 1 on that value of θ which maximizes the integrand, which is the likelihood function. In this case, that would be $\theta = x$, and the lower bound on the Bayes factor is $\exp(-[x-\theta_0]^2/2)$. For example, if $x = \theta_0 + 1.96$, which is the critical value for the usual two-sided level 0.05 test of H, we get a lower bound of 0.1465. This says that a data value that would just barely lead to rejecting H at level 0.05 could not possibly change one's odds against the hypothesis by more than a factor of 7, and then only in the extremely unlikely case that one believed before seeing the data that Θ was sure to equal $\theta_0 + 1.96$ if it was not θ_0. Put another way, in order for the posterior probability of H to be as low as 0.05, the prior probability p_0 would have to be lower than 0.2643, and much lower if a more reasonable prior on the alternative were used.

A more realistic expression of prior opinion might be that the prior, given $\Theta \neq \theta_0$, is a normal distribution with mean θ_0 and some variance τ^2. In this case, the Bayes factor is

$$\sqrt{1+\tau^2} \exp\left(-\frac{(x-\theta_0)^2\tau^2}{2(1+\tau^2)}\right). \tag{4.19}$$

The prior in this class that leads to the smallest Bayes factor can easily be shown (see Problem 7 on page 286) to be the one with

$$\tau^2 = \begin{cases} (x-\theta_0)^2 - 1 & \text{if } |x-\theta_0| > 1, \\ 0 & \text{otherwise.} \end{cases}$$

[5]For more discussion of this technique, see Edwards, Lindman, and Savage (1963).

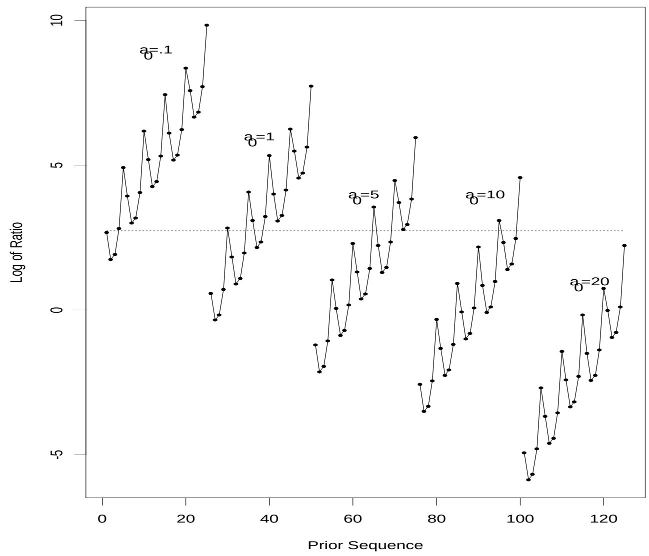

FIGURE 4.30. Logarithms of Ratios of Prior Densities

then make detailed comparisons with Bayes factors. At this point, we only mention that the results can often be in stark contrast. In particular, data with very small significance probability (supposedly suggesting that the data do not support the hypothesis) can have relatively large values for the lower bounds on the Bayes factor (suggesting that the data would not change one's mind drastically about the probability of the hypothesis.) This conflict is sometimes called "Lindley's paradox" [see Lindley (1957) and Jeffreys (1961)].

If one believes that the probability is 0 that the parameter lies in a low-dimensional subset of the parameter space, then it is not appropriate to test the types of hypotheses we have considered in this section. As an alternative, one can calculate a measure of how far the parameter of interest is from the hypothesized low-dimensional subset. For example, if the parameter is $\Theta = (\mathrm{M}, \Sigma)$ and the hypothesis is that M is near 0, then the posterior distribution of $|\mathrm{M}|$ contains a great deal of information about how far M is from 0. Also, the posterior distribution of $|\mathrm{M}|/\Sigma$ contains information of a similar sort.

Example 4.31. Suppose that $X \sim N(\theta, 1)$, given $\Theta = \theta$, and that our hypothesis is that Θ is near θ_0. If we consider all prior distributions of the form $\Theta \sim N(0, \tau^2)$, then the posterior distribution of Θ is $N([\theta_0 + \tau^2 x]/[1+\tau^2], \tau^2/[1+\tau^2])$. For each $\delta > 0$, we can calculate

$$\Pr(|\Theta - \theta_0| \le \delta) = \Phi\left(\lambda[\theta_0 - x] + \frac{\delta}{\lambda}\right) - \Phi\left(\lambda[\theta_0 - x] - \frac{\delta}{\lambda}\right),$$

where $\lambda = \tau/\sqrt{1+\tau^2}$, and Φ is the standard normal CDF. There is no useful upper bound on this probability, but a lower bound is obtained by letting $\lambda \rightarrow$

$S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1) = s^2$ are observed, the P-value for $H : a \leq \mathrm{M} \leq b$ will be the level of the LR test that rejects H when $\sqrt{n}(\overline{X} - b)/S > \sqrt{n}(\overline{x} - b)/s$ or when $\sqrt{n}(\overline{X} - a)/S < -\sqrt{n}(\overline{x} - b)/s$. Call this P-value $p_H(x)$. It is easy to see that $p_H(x)$ is precisely the same as the P-value for the hypothesis $H_b : \mathrm{M} = b$ versus $A_b : \mathrm{M} \neq b$ relative to the collection of UMPU (two-sided) tests. Also $p_H(x)$ is precisely twice the size of the P-value for the hypothesis $H'_b : \mathrm{M} \leq b$ versus $A'_b : \mathrm{M} > b$ relative to the collection of UMPU (one-sided) tests.

Since the one-sided P-values equal posterior probabilities of the hypotheses when using improper priors (see Problem 62 on page 294), we find that the one-sided P-value for the hypothesis $H'_c : \mathrm{M} \leq c$ is a continuous function of c. It follows that there exists $c > b$ such that the one-sided P-value for H'_c satisfies $p_H(x) > p_{H'_c}(x) > p_H(x)/2$. If we are to interpret the P-values relative to the collections of LR tests as degrees of support for the respective hypotheses, then if $x > b$, the degree of support for every hypothesis of the form $H_a : \mathrm{M} \in [a, b]$ (for varying a but fixed b) is the same number $p_H(x)$ (even if $a = b$ or if a is much less than b). But the degree of support for the hypothesis $H'_c : \mathrm{M} \in (-\infty, c]$ (where $c > b$ is chosen as above) is $p_{H'_c}(x) < p_H(x)$. In words, the data offer more support for every hypothesis of the form $\mathrm{M} \in [a, b]$ than they do for $\mathrm{M} \in (\infty, c]$ even though $[a, b] \subseteq (-\infty, c]$ for every a.

We see that there are cases (usually one-sided testing) in which P-values can correspond to a degree of support for the hypothesis, but there are other cases (e.g., two-sided alternatives) when they cannot. It is possible, for example, with normal data, to express certain P-values as weighted averages over the corresponding hypotheses of P-values for testing point hypotheses. (See Problem 66 on page 294.) To generalize this idea beyond normal distributions (or symmetric location families), one needs to consider tests that may not be UMPU. Spjøtvoll (1983) defines a measure of "acceptability" of point hypotheses, which has the property that for many distributions and certain hypotheses, the weighted average of the acceptability over the hypothesis equals something closely related to the P-value.

In Section 6.3.1, we give some general conditions under which P-values are equal to the posterior probabilities of hypotheses. Casella and Berger (1987) study the problem of testing a one-sided hypothesis–alternative pair and find that in many cases, the P-value is approximately a limit of posterior probabilities. Examples 4.143 and 4.148 point out that the P-value cannot be taken as a method for providing a "degree of support" for general hypotheses, however.

4.6.2 P-Values and Bayes Factors

In Section 4.2.2, we introduced Bayes factors as ways to quantify the extent to which the data affects our beliefs about a hypothesis. There were lower bounds on Bayes factors which indicate the maximum extent to which the odds in favor of the hypothesis could decrease from the prior to the posterior. When the lower bound is not very small, the data cannot make us believe very much less in the hypothesis than we did a priori. Since P-values have been suggested as measures of support for the hypothesis is conjugate as in Example 4.22 on page 224. The Bayes factor for the hypothesis $H : \mathrm{M} = \mu_0$ was given in (4.23). Consider what happens to this expression as $n \to \infty$. First, suppose that the usual t statistic converges to a constant t_0. That is, assume that $\sqrt{n}(\overline{x}_n - \mu_0)/s_n$ converges to t_0, where $s_n^2 = w/(n-1)$. In this case, the formula in (4.23) behaves asymptotically (as $n \to \infty$) like $\sqrt{n}$ times a constant. This means that the Bayes factor goes to ∞ as n increases, hence the posterior probability of the hypothesis goes to 1. What happens to the P-value for this same sequence of data sets? Since the t statistic is converging to a constant t_0, the P-value is converging to $1-2\Phi(t_0)$. For example, if $t_0 = 1.96$, the P-value will go to 0.05, while the posterior probability of the hypothesis goes to 1. This is an extreme example of Lindley's paradox. Once again, the suggestion of Lehmann (1958) to let α decrease as n increases would seem appropriate.

The situation is much the same if one uses the approximate Bayes factor as calculated in (4.29). This formula does not require that the prior be natural conjugate, but merely smooth in some sense. Since both $\hat{\Psi}$ and $\hat{\Theta}$ converge to finite values almost surely given Θ (by the strong law of large numbers 1.62), the expression in (4.29) also behaves asymptotically like $\sqrt{n}$ times a constant if the t statistic converges to t_0.

Of course, the t statistic will converge to a finite value with positive probability only if the hypothesis is true. So, it is comforting that virtually any smooth prior distribution will lead to eventually discovering that the hypothesis is true, if it is indeed true.[44] On the other hand, it is a bit disconcerting that the P-value will stay bounded away from 1 with positive probability no matter how much data we observe. (See Problem 65 on page 294 to see how to prove that the P-value has $U(0,1)$ distribution given that the hypothesis is true.) If the hypothesis is false, it is easy to check that the P-value will go to 0, as will the Bayes factor.

The irreconcilability of P-values and posterior probabilities as illustrated in Examples 4.150 and 4.151 is quite typical of cases in which Ω_H is a lower-dimensional set than Ω. [See Schervish (1996) for some examples with distributions other than normal.] Together with the strange behavior of P-values in Examples 4.148 and 4.149, it becomes difficult to justify their use to measure strength of evidence in favor of the hypothesis for two-sided problems.*

4.7 Problems

Section 4.1:

1. Prove that the loss function in (4.2) is equivalent to a 0–1–c loss function. (By "equivalent" we mean that both the posterior risk and the risk function will rank all decision rules the same way regardless of which loss is used.)
2. Prove that the general form of the hypothesis-testing loss function (in Definition 4.1) can be written as $d(v)$ times a 0–1 loss function for some function $d > 0$ if the loss is 0 whenever a correct decision is made.

---

[44] In Section 7.4, we will prove some results that make more precise this limiting ability of posterior distributions to identify the value of a parameter.

*Michael Lavine has pointed out that Bayes factors also have the property that if H_2 implies H_1, the Bayes factor in favor of H_2 might be larger than the Bayes factor in favor of H_1. This means that Bayes factors do not measure support for hypotheses either. They only measure the degree to which the support changes.