
Exercise 1

Check the S-Plus code on the course homepage.

Exercise 2

(a) The basic details here are really just a special case of the "mixture priors" handout on the course homepage. In the notation there, the required Bayes' factor is $1/L(\bar{y}) = p_0(\bar{y})/p_1(\bar{y})$ where $\bar{y}$ stands for the data in this case. Under the null hypothesis $\mu = 0$ we have

$$p_0(\bar{y}) = A\exp(-n\bar{y}^2/2v)$$

with $A = 1/\sqrt{2\pi v/n}$. Under the alternative $\mu > 0$ we need to compute the (prior predictive) density

$$\begin{aligned} p_1(\bar{y}) &= \int_{-\infty}^{\infty} p(\bar{y}|\mu)N(\mu|0,w)d\mu \\ &= \exp(-\bar{y}^2/(2(w+v/n)))/\sqrt{(2\pi(w+v/n))} \end{aligned}$$

using the identity given in the question.

Note that this result is "immediate" anyhow: $p_1(\bar{y})$ is the marginal distribution under the joint specified for $(\bar{y},\mu)$ by the product $N(\bar{y}|\mu,v/n)N(\mu|0,w)$. Equivalently, $\bar{y} = \mu+\epsilon$ where $\epsilon \sim N(\epsilon|0,v/n)$ independently of μ. Thus $\bar{y}$ is the sum of two independent normals and the margin is easily $N(\bar{y}|0,w+v/n)$, i.e., the distribution with the required density $p_1(\bar{y})$.

It follows that the Bayes' factor is

$$\begin{aligned} p_0(\bar{y})/p_1(\bar{y}) &= \sqrt{((w+v/n)n/v)}\exp(-(n\bar{y}^2/2v)\{1-1/(1+wn/v)\}) \\ &= \sqrt{(1+wn/v)}\exp(-z^2/2a) \end{aligned}$$

with $a = 1+v/nw$, as required.

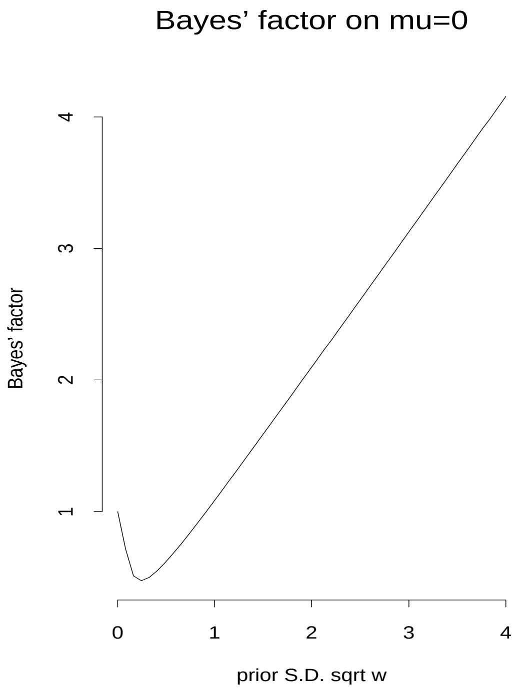
(b) With $n = 50$, $v = 1$ and $z = 1.96$, we get Bayes' factor

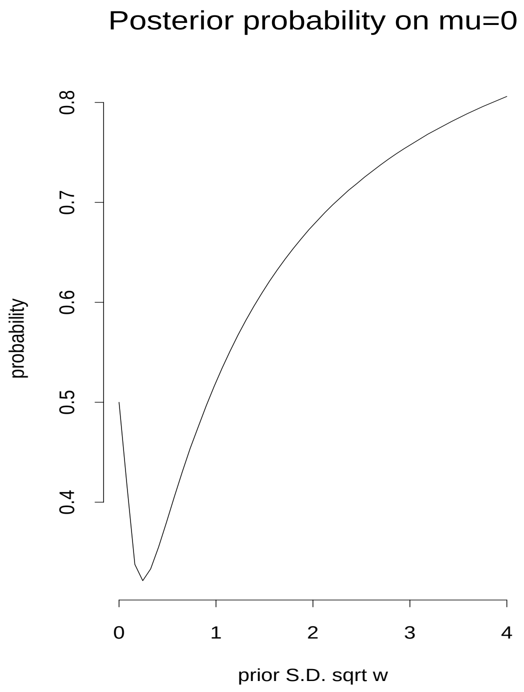
$$B = p_0(\bar{y})/p_1(\bar{y}) = \sqrt{(1+50w)}\exp(-1.9208/(1+0.02/w)).$$

By definition, the posterior probability $1-\pi^*$ on $\mu = 0$ is implictly defined by

$$\frac{(1-\pi^*)}{\pi^*} = \frac{(1-\pi)}{\pi}B$$

so that, at $\pi = 0.5$, we have $1-\pi^* = B/(1+B)$. The figures below plot both B and $1-\pi^*$ for varying values of the standard deviation $\sqrt{w}$ of the prior on the alternative $\mu \neq 0$.





This shows several things. First, an investigator with a prior with $w > 1$ (roughly) will have a posterior probability greater than the prior of 0.5 on the null hypothesis. As conservative or "vague" alternative priors will have larger variances, this indicates that such investigators will conclude quite the opposite result than implied by the standard testing paradigm: i.e., there is no strong evidence to reject the null hypothesis $\mu = 0$ based on the observed value $z = 1.96$ of the usual test statistic. A second point is that, for small values of w, this z value does provide modest evidence against the null hypothesis, i.e., the posterior probability on the null decreases from its prior value. The third and perhaps most important point is that the posterior probability is *bounded below* by 0.32 – i.e., for *any* $N(\mu|0, w)$ prior, an investigator will have a posterior probability no less than about one third on the null. This is relative to the prior probability of 0.5 – so look back to the Bayes' factor graph which does not depend on π. Here we see that the Bayes' factor exceeds 0.47 (roughly) whatever the value of $\sqrt{w}$. This is a result of *Bayesian robustness,* as follows: For *any* investigator having a $N(\mu|0, w)$ prior, whatever the value of w, on observing the "extreme" test statistics $z = 1.96$, his/her posterior odds on the null hypothesis are always greater than 0.47 times the prior odds. So the most the $z = 1.96$ can do it halve the odds on the null hypothesis. Most statisticians view a Bayes' factor (likelihood ratio) of about one half as really very modest evidence against the hypothesis; only when we get down to values of 0.1 or less does the evidence begin to appear meaningful. So this result shows how conservative Bayesians are in comparison to non-Bayesians; for the latter, $z = 1.96$ is "significant at the 5% level", whereas the "normal prior" Bayesians simply are not persuaded. Note finally that this line of thought can be developed to explain the problems of proliferating "significant" results in medical and social-science research literatures, many of which turn out later to be insignificant! This is a special example of much more general and far-reaching "robustness" results in standard hypothesis testing. Similar results exist for other priors on $\mu \neq 0$, i.e., a wide range of non-normal priors, and for other models. For example, see the article on signal detection in astronomy in the coursepack handout for another. I can provide more examples and references.