Stat 591 Notes – Bayes factor for testing in the normal model

Ryan Martin (rgmartin@uic.edu)

September 20, 2013

Problem setup

Consider an iid sample $X_1, \ldots, X_n$ from a normal distribution $\mathsf{N}(\mu, \sigma^2)$, where both μ and σ^2 are unknown. The goal is to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. Without loss of generality, take $\mu_0 = 0$.[1] Note that, under H_0 there is only one parameter, namely, σ, but under H_1 there are two parameters. Under both H_0 and H_1, we shall follow Jeffreys and consider a prior $\pi_\sigma(\sigma) = 1/\sigma$ for σ, an improper prior.[2] Under H_1, the conditional prior for $\mu|\sigma$ is taken to be a location-scale transformation of a given PDF $g(\cdot)$, i.e.,

$$\pi_{\mu|\sigma}(\mu) = (1/\sigma)g(\mu/\sigma).$$

Note that the use of σ in the prior for μ makes the conditional variance for μ, given σ, scaled appropriately with the data variance. We shall consider two particular PDFs g. Our goal is to construct the Bayes factor $B = B_{01}$ for comparing the models H_0 and H_1. We shall write this as

$$B = B(x) = \frac{m_0(x)}{m_1(x)},$$

where $m_0(x)$ and $m_1(x)$ are marginal likelihoods for the observed data $x = (x_1, \ldots, x_n)$ under models H_0 and H_1, respectively. Below we work through the necessary (and tedious) marginal likelihood calculations.

Marginal likelihood under the null

Under H_0, the only parameter is σ, and we have the prior π_σ. So, the marginal likelihood is obtained by doing some integration. Write the likelihood function as

$$L(\mu, \sigma) = \Bigl(\frac{1}{\sigma}\Bigr)^n e^{-\frac{n}{2\sigma^2}[\hat\sigma^2 + (\hat\mu - \mu)^2]},$$

where $\hat\mu = \bar x$ and $\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar x)^2$, the maximum likelihood estimators. We omitted the $(2\pi)^{-n/2}$ term because it's a constant that will appear in both the null and alternative marginal likelihoods, hence it will cancel out of the Bayes factor. Now, the marginal likelihood is found by integrating $L(0, \sigma)$ with respect to π_σ. That is,

$$\begin{aligned}
m_0(x) &= \int_0^\infty \Bigl(\frac{1}{\sigma}\Bigr)^n e^{-\frac{n(\hat\sigma^2 + \hat\mu^2)}{2\sigma^2}} \frac{1}{\sigma}\, d\sigma \\
&= \int_0^\infty \Bigl(\frac{1}{\sigma}\Bigr)^{n+1} e^{-\frac{D_0}{2\sigma^2}}\, d\sigma \qquad [D_0 = n(\hat\sigma^2 + \hat\mu^2)] \\
&= \frac{1}{2}\int_0^\infty \Bigl(\frac{1}{y}\Bigr)^{\frac{n}{2}+1} e^{-D_0/2y}\, dy \qquad [y = \sigma^2] \\
&= \frac{1}{2}\frac{\Gamma(n/2)}{(D_0/2)^{n/2}},
\end{aligned}$$

[1] Why is there no loss of generality?

[2] It's OK to use an improper prior for σ because (a) the posterior is proper and (b) σ appears in both H_0 and H_1 so the arbitrary scale of the prior cancels out in the Bayes factor ratio.

where the last equality follows from the definition of the inverse gamma distribution.[3,4] In particular, the part in red is the form of the $\mathsf{InvGam}(\text{shape} = \frac{n}{2}, \text{rate} = \frac{D_0}{2})$ distribution. Therefore, the integral must equal the normalizing constant.

Marginal likelihood under the alternative

The general form for the marginal likelihood under H_1 is

$$m_1(x) = \int_0^\infty \int_{-\infty}^\infty \Big(\frac{1}{\sigma}\Big)^n e^{-\frac{n}{2\sigma^2}[\hat{\sigma}^2 + (\mu - \hat{\mu})^2]} \pi_{\mu|\sigma}(\mu) \frac{1}{\sigma} \, d\mu \, d\sigma.$$

Here, like on page 47 in [GDS], we shall consider two prior distributions for $\mu|\sigma$ under H_1. The first is normal and the second is Cauchy.

1. Consider g as a standard normal density, so that $\pi_{\mu|\sigma}(\mu) = \mathsf{N}(\mu \mid 0, \sigma^2)$, a $\mathsf{N}(0, \sigma^2)$ density. Then the marginal likelihood looks like

$$m_1(x) = \int_0^\infty \Big[\int_{-\infty}^\infty e^{-\frac{n}{2\sigma^2}(\mu - \hat{\mu})^2} \mathsf{N}(\mu \mid 0, \sigma^2) \, d\mu\Big] \Big(\frac{1}{\sigma}\Big)^{n+1} e^{-\frac{n\hat{\sigma}^2}{2\sigma^2}} \, d\sigma.$$

   Note that if $\mu \sim \mathsf{N}(0, \sigma^2)$, then $(\mu - \hat{\mu})^2/\sigma^2$ is a non-central chi-square random variable, with non-centrality parameter $\lambda = \hat{\mu}^2/\sigma^2$ and one degree of freedom. Now, the inner-most integral—in brackets $[\cdots]$—is just the moment-generating function of this non-central chi-square, evaluated at $t = -n/2$. Using the formula for this moment-generating function,[5] the inner-most integral is just

$$\int_{-\infty}^\infty e^{-\frac{n}{2\sigma^2}(\mu - \hat{\mu})^2} \mathsf{N}(\mu \mid 0, \sigma^2) \, d\mu = \frac{1}{(n+1)^{1/2}} e^{-\frac{n\hat{\mu}^2}{2(n+1)\sigma^2}}.$$

   The first term above is free of σ and (thankfully) the second term is of a form that will conveniently combine with the remaining integrand. So, it remains to work out

$$m_1(x) = \frac{1}{(n+1)^{1/2}} \int_0^\infty \Big(\frac{1}{\sigma}\Big)^{n+1} e^{-\frac{D_{11}}{2\sigma^2}} \, d\sigma.$$

   where $D_{11} = n\hat{\sigma}^2 + n\hat{\mu}^2/(n+1)$. We can now finish this, as above, by observing that the integral is the normalizing constant for the inverse gamma density, i.e.,

$$m_1(x) = \frac{1}{2} \frac{1}{(n+1)^{1/2}} \frac{\Gamma(n/2)}{(D_{11}/2)^{n/2}}.$$

2. Now consider g as a standard Cauchy density, so that

$$\pi_{\mu|\sigma}(\mu) = \frac{1}{\sigma} \frac{1}{\pi(1 + \mu^2/\sigma^2)}$$

[3] This is the distribution of $Y = 1/X$ when X is gamma; see `http://en.wikipedia.org/wiki/Inverse-gamma_distribution`

[4] The expression for $m_0(x)$ is equivalent to that in Eq. (26), Jeffreys' *Theory of Probability*, §5.5.2.

[5] See `http://en.wikipedia.org/wiki/Noncentral_chi-squared_distribution`; this can also be found by direct integration, without using this fancy mgf formula.

is a Cauchy density with location 0 and scale σ. As pointed out in [GDS], the Cauchy distribution can be expressed as a scale mixture of normals, in particular,

$$\pi_{\mu|\sigma}(\mu) = \int_0^\infty \mathsf{N}(\mu \mid 0, v)\mathsf{InvGam}(v \mid \tfrac{1}{2}, \tfrac{\sigma^2}{2})\, dv.$$

The idea is the insert this integral in the definition of the marginal likelihood, then interchange the order of integration. That is,

$$m_1(x) = \int_0^\infty H(v)\, dv,$$

where $H(v)$ is given by

$$\int_0^\infty \Bigl(\frac{1}{\sigma}\Bigr)^{n+1} e^{-\frac{n\hat\sigma^2}{2\sigma^2}} \Bigl[\int_{-\infty}^\infty e^{-\frac{n(\mu-\hat\mu)^2}{2\sigma^2}} \mathsf{N}(\mu \mid 0, v)\, d\mu\Bigr] \mathsf{InvGam}(v \mid \tfrac{1}{2}, \tfrac{\sigma^2}{2})\, d\sigma.$$

It is possible to find a (basically) closed-form expression for $H(v)$, but it's a mess. So, rather than trying to put down all these difficult calculations here, I'll just give the answer with a reference. After considerable work, the marginal likelihood $m_1(x)$ in the Cauchy prior case is given by

$$m_1(x) = m_0(x) \cdot \frac{2}{\pi} \int_0^\infty M\Bigl(\frac{n}{2}, \frac{1}{2}, \frac{n\hat\mu^2 v^2}{2(\hat\sigma^2 + \hat\mu^2)}\Bigr) e^{-nv^2/2} \frac{1}{1+v^2}\, dv,$$

where $M \equiv {}_1F_1$ denotes Kummer's confluent hypergeometric function (first kind).[6] This formula, along with some more details concerning its derivation are given in Jeffreys, *Theory of Probability*, §5.5.2.

Bayes factors

For the normal prior, the Bayes factor looks like

$$B = \frac{m_0(x)}{m_1(x)} = (n+1)^{1/2} \Bigl(\frac{D_{11}}{D_0}\Bigr)^{n/2} = (n+1)^{1/2} \Bigl\{\frac{\hat\sigma^2 + \hat\mu^2/(n+1)}{\hat\sigma^2 + \hat\mu^2}\Bigr\}^{n/2}.$$

For the Cauchy prior, the Bayes factor is more complicated, but can be written as

$$\frac{1}{B} = \frac{2}{\pi} \int_0^\infty M\Bigl(\frac{n}{2}, \frac{1}{2}, \frac{n\hat\mu^2 v^2}{2(\hat\sigma^2 + \hat\mu^2)}\Bigr) e^{-nv^2/2} \frac{1}{1+v^2}\, dv.$$

It's one thing to write down a formula for the Bayes factor—actually evaluating it is another thing. Some R code is given below to evaluate the Cauchy prior Bayes factor. It's based on an idea involving a Poisson expectation. This comes up in the next section, so the explanation of this strategy will be postponed till after the next section.

[6] See `http://en.wikipedia.org/wiki/Confluent_hypergeometric_function`

Jeffreys' claim about Bayes factor behavior

The middle of page 47 in [GDS] discusses a claim by Jeffreys that the Bayes factor, as a function of $(\hat{\mu}, \hat{\sigma}^2)$, should vanish if $\hat{\mu} \to \infty$ and $\hat{\sigma}^2$ remains bounded. This is not a statement about large samples, the limit in question is for fixed n with data changing so that $\hat{\mu} \to \infty$ and $\hat{\sigma}^2$ stays bounded. This property does not hold for the Bayes factor based on the normal prior; this is easy to see from the formula for B above. However, the property does hold for the Cauchy prior Bayes factor. To check this, we need the series definition of M:

$$M(a, b, z) = \sum_{k=0}^{\infty} \frac{a^{(k)}}{b^{(k)}} \frac{z^k}{k!},$$

where $a^{(0)} = 1$ and $a^{(k)} = a(a+1)\cdots(a+k-1)$ is the rising factorial. If we interchange the limit, as $\hat{\mu} \to \infty$, with the integral over v, then we want to consider

$$\lim_{\hat{\mu}\to\infty} \frac{1}{B} \propto \int_0^{\infty} \Bigl\{ \sum_{k=0}^{\infty} \frac{(n/2)^{(k)}}{(1/2)^{(k)}} \Bigl(\frac{nv^2}{2}\Bigr)^k \frac{1}{k!} \Bigr\} e^{-nv^2/2} \frac{1}{1+v^2}\, dv.$$

Look at the ratio of rising factorials and define $r_n(k)$ as follows:

$$\frac{(n/2)^{(k)}}{(1/2)^{(k)}} = \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{n}{2})} \frac{\Gamma(\frac{n}{2}+k)}{\Gamma(\frac{1}{2}+k)} \propto \frac{\Gamma(\frac{n}{2}+k)}{\Gamma(\frac{1}{2}+k)} =: r_n(k).$$

It is not too difficult to show, using induction, that $r_n(k) \geq (k + \frac{n}{2})^{-(n-1)/2}$ for all $k \geq 0$. Then the messy integral above is proportional to

$$\int_0^{\infty} \mathsf{E}_{v^2/2}\{r_n(K)\} \frac{1}{1+v^2}\, dv \geq \int_0^{\infty} \mathsf{E}_{v^2/2}\{(K + \tfrac{n}{2})^{(n-1)/2}\} \frac{1}{1+v^2}\, dv,$$

where K is a Poisson random variable with parameter $\lambda = v^2/2$. Since Poisson moments are polynomials in its parameter, and integrands like $v^p/(1+v^2)$ for $p \geq 1$ are not integrable over $v \in (0, \infty)$, it follows that the messy integral diverges. Therefore,

$$\lim_{\hat{\mu}\to\infty} \frac{1}{B} = \infty \implies \lim_{\hat{\mu}\to\infty} B = 0,$$

as was to be proved.

Example 2.12 in [GDS]

This examples concerns a test of Einstein's theory of relativity. There are $n = 4$ observations and [GDS] quotes the value of the Cauchy prior Bayes factor as $B = 2.98$, which lends some support to Einstein's claim, but not too much. [GDS] does not discuss how they found this value—they leave it as an exercise. A computational strategy, along with R code, is presented below; it involves both Monte Carlo and numerical integration. When I ran the code, I got $B = 2.974$ for the Bayes factor, which is on target with the value reported in [GDS]; there will be some slight variation from run to run.

Computing the Cauchy prior Bayes factor

In the proof of the behavior of the Cauchy prior Bayes factor, as $\hat{\mu} \to \infty$, we saw that a Poisson expectation came into the picture. In particular, we evaluate the confluent hypergeometric function M by Monte Carlo by simulating from a Poisson distribution with parameter $\lambda(v)$, depending on both data and the particular value v, i.e.,

$$\lambda(v) = \frac{nv^2}{2} \frac{\hat{\mu}^2}{\hat{\sigma}^2 + \hat{\mu}^2}.$$

Do a Monte Carlo simulation for each candidate v to evaluate the M function, and use the function `integrate` to evaluate the outermost integral numerically. Then it's easy to recover B.

```
# Example 2.12 (and Problem 2.7) in [GDS]
x <- c(1.98, 1.61, 1.18, 2.24)
x <- x - 1.75
n <- length(x)
mu.hat <- mean(x)
sig2.hat <- (n - 1) * var(x) / n
f <- function(k) gamma(0.5) * gamma(n / 2 + k) / gamma(n / 2) / gamma(0.5 + k)
N <- 5000
G <- function(v) {

  o <- 0 * v
  for(i in seq_along(v)) {

    lambda <- n * v[i]**2 * (mu.hat**2 / (sig2.hat + mu.hat**2)) / 2
    set.seed(7)
    K <- rpois(N, lambda)
    o[i] <- mean(f(K))

  }
  return(o)

}
H <- function(v) {

  lambda.v <- n * v**2 * (sig2.hat / (sig2.hat + mu.hat**2)) / 2
  return(G(v) * exp(-lambda.v) / (1 + v**2))

}
BF <- pi / integrate(H, 0, Inf)$value / 2
print(BF)
```
