Stat 136 (Bayesian Statistics)

Lesson 3.2 Bayesian Inference About the Variance of a Normal Distribution

Introduction

The spread of any distribution is equally important as the measure of center (e. g. mean). The spread of any distribution is measured by the standard deviation (or the variance). Recall, that in making Bayesian inference for the mean μ of a normal distribution, either the standard deviation σ is known or unknown.

To do inferences on the standard deviation of the normal distribution, we reverse the roles of the parameters. The variance is in squared units, and it is hard to visualize our belief about it. So, we will make the transformation to the corresponding prior and posterior density for the standard deviation.

The inverse chi-squared distribution

A random variable Y has an **inverse chi-squared** distribution with k degrees of freedom if its pdf is given by

$$f_Y(y) = \begin{cases} \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})y^{\frac{k}{2}+1}}e^{-\frac{1}{2y}}, \ 0 < y < \infty \\ 0, \ \text{otherwise} \end{cases}$$

A random variable Y has an **S times an inverse chi-squared** distribution with k degrees of freedom if its pdf is given by

$$f_Y(y) = \begin{cases} \frac{S^{\frac{k}{2}}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})y^{\frac{k}{2}+1}}e^{-\frac{S}{2y}}, \ 0 < y < \infty \\ 0, \ \text{otherwise} \end{cases}$$

Furthermore, when Y has *S times an inverse chi-squared distribution* with k degrees of freedom, then $W = \frac{S}{Y}$ has the chi-squared distribution with k degrees of freedom. This transformation allows us to find probabilities for the inverse chi-squared random variables using the upper tail area of the chi-squared distribution.

When Y has an S *times an inverse chi-squared* distribution with k degrees of freedom, then

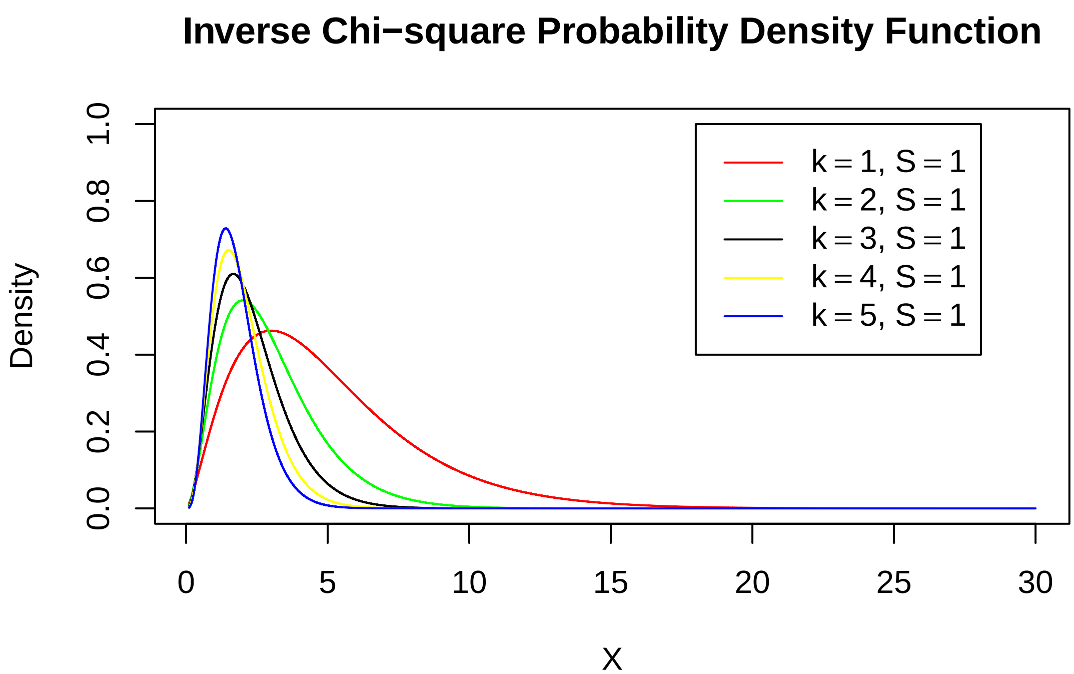
$$E(Y) = \frac{S}{k-1},\ k > 2$$

and

$$V(Y) = \frac{2S^2}{(k-2)^2(k-4)},\ k > 4$$

The following code chunk will generate plots for some inverse chi-squared densities ($S = 1$)

```
Warning: package 'extraDistr' was built under R version 4.4.3
```



Likelihood of variance for a random sample from a normal distribution

Let $Y_1, Y_2, \cdots, Y_n$ be a random sample from $N(\mu, \sigma^2)$, where μ is known. The likelihood of σ^2 is

$$\begin{aligned}
L(\mathbf{y}|\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \\
&\propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2} \\
&= \frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}SST}, \text{ where: } SST = \sum_{i=1}^{n}(y_i-\mu)^2
\end{aligned}$$

Thus, the variance σ^2 has the **SST times an inverse chi-squared** distribution with $k = n-2$ degrees of freedom.

Recall that Y has an S *times an inverse chi-squared* distribution if

$$f_Y(y) \propto \frac{1}{y^{\frac{k}{2}+1}} e^{-\frac{S}{2y}}$$

And,

$$L(\mathbf{y}|\sigma^2) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}SST}$$

Thus, it is easy to see that $y = \sigma^2$ and $S = SST$, thus,

$$\begin{aligned}
\frac{k}{2} + 1 = \frac{n}{2} &\implies \frac{k}{2} + \frac{2}{2} = \frac{n}{2} \\
&\implies k + 2 = n \\
&\implies k = n - 2
\end{aligned}$$

Positive uniform prior density for σ^2

Recall that the variance is in squared units, not the same unit as the mean. This means that they are not directly comparable so that it is much harder to understand a prior density for σ^2. The standard deviation is in the same units as the mean, so it is much easier to understand.

To get the prior density for σ^2 that corresponds to the prior density for σ (and vice versa), we use the following identity:

$$\pi(\sigma^2) = \pi(\sigma) \times \frac{1}{2\sigma}$$

Suppose we decide that we consider all positive values of the variance σ^2 to be equally likely and do not wish to favor any particular value over another. This means we give all positive values of σ^2 equal prior weight.

This gives the positive uniform prior density for the variance, i. e. $\pi(\sigma^2) = 1$

Consequently, the posterior distribution of σ^2 is

$$\pi(\sigma^2|\mathbf{y}) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}SST}$$

Therefore, the posterior density of σ^2 is an *SST times an inverse chi-square* with $n-2$ degrees of freedom.

Suppose $\pi(\sigma) = 1$. This means that $\pi(\sigma^2) = \frac{1}{2\sigma}$ based on the identity (1). Hence, the posterior distribution of σ^2 is

$$\begin{aligned}\pi(\sigma^2|\mathbf{y}) &\propto \frac{1}{\sigma} \times \frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}SST} \\ &= \frac{1}{(\sigma^2)^{\frac{n+1}{2}}} e^{-\frac{1}{2\sigma^2}SST}\end{aligned}$$

Therefore, the posterior density of σ^2 is an *SST times an inverse chi-square* with $n-1$ degrees of freedom.

Jeffrey's prior density for σ^2

Let $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$. The posterior distribution of σ^2 is

$$\begin{aligned}\pi(\sigma^2|\mathbf{y}) &\propto \frac{1}{\sigma^2} \times \frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}SST} \\ &= \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} e^{-\frac{1}{2\sigma^2}SST}\end{aligned}$$

Therefore, the posterior density of σ^2 is an *SST times an inverse chi-square* with n degrees of freedom.

Inverse chi-square prior density for σ^2

Suppose we use an S *times an inverse chi-squared* with k degrees of freedom as the prior for σ^2, i. e.

$$\pi(\sigma^2) \propto \frac{1}{(\sigma^2)^{\frac{k}{2}+1}} e^{-\frac{S}{2\sigma^2}}$$

Hence, the posterior distribution of σ^2 is

$$\begin{aligned} \pi(\sigma^2|\mathbf{y}) &\propto \frac{1}{(\sigma^2)^{\frac{k}{2}+1}} e^{-\frac{S}{2\sigma^2}} \times \frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}SST} \\ &= \frac{1}{(\sigma^2)^{\frac{n+k}{2}+1}} e^{-\frac{1}{2\sigma^2}(S+SST)} \end{aligned}$$

This distribution can be viewed as an $S^{'} = S + SST$ times an inverse chi-squared distribution with $k^{'} = n + k$ degrees of freedom.

Note that the inverse chi-squared distribution is a special case of the inverse gamma distribution. If Y has an inverse gamma distribution then its pdf is

$$f_Y(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} e^{-\frac{\beta}{y}},\ y > 0$$

Now, if we let $\alpha = \frac{k}{2}$ and $\beta = \frac{1}{2}$, then this is the same density as an inverse chi-square distribution with k degrees of freedom.

On the other hand, if we let $\alpha = \frac{k}{2}$ and $\beta = \frac{S}{2}$, then this is the same density as an S times an inverse chi-square distribution with k degrees of freedom.

Let

$$\pi(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} e^{-\frac{\beta}{\sigma^2}}$$

Then, the posterior distribution of σ^2 is

$$\begin{aligned} \pi(\sigma^2|\mathbf{y}) &\propto \frac{1}{(\sigma^2)^{\alpha+1}} e^{-\frac{\beta}{\sigma^2}} \times \frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}SST} \\ &= \frac{1}{(\sigma^2)^{\frac{n}{2}+\alpha+1}} e^{-\frac{2\beta+SST}{2\sigma^2}} \end{aligned}$$

This is proportional to the pdf of an inverse gamma distribution with $\alpha^{'} = \frac{n}{2} + \alpha$ and $\beta^{'} = \frac{2\beta+SST}{2}$.

Remember that if $\sigma^2 \sim InvGamma(\alpha', \beta')$, then $\frac{1}{\sigma^2} \sim Gamma(\alpha', \beta')$.

Also, Gelman et al. (2003) showed that the $InvGamma(\alpha, \beta)$ can be reparameterized as a scaled inverse chi-squared distribution with scale $S = \frac{\beta}{\alpha}$ and $k = 2\alpha$ degrees of freedom.

It is suggested that in the scaled inverse chi-squared parameterization, the prior distribution will have at least $k = 1$ degree of freedom, and hence will be a proper prior.

An example

Aroha, Bernardo, and Carlos are three statisticians employed at a dairy factory who want to do inference on the standard deviation of the content weights of "1 kg" packages of dried milk powder coming to the production line. The three employees consider that the weights of the packages is distributed as $N(1015, \sigma^2)$. They take a random sample of size 10 and measure the content weights in grams. These are: 1011, 1009, 1019, 1012, 1011, 1016, 1018, 1021, 1016, 1012.

For this sample, the summary statistics are:

```r
library(rMR)
mu <- 1015
wt <- c(1011, 1009, 1019, 1012, 1011, 1016, 1018, 1021, 1016, 1012)
print(cbind(n=length(wt), ybar=mean(wt), sd=sd(wt),SST=sumsq(wt-mu)))
```

```
      n   ybar       sd   SST
[1,] 10 1014.5 4.034573 146.5
```

Aroha decides that she will use the positive uniform prior for the standard deviation, $\pi(\sigma) = 1$, for $\sigma > 0$. Bernardo decides he will use Jeffrey's prior $\pi(\sigma) = \frac{1}{\sigma}$. Meanwhile, Carlos decides that his prior belief about the standard deviation is that its median equals 5. Hence,

$$
\begin{aligned}
P(\sigma > 5) = 0.5 &\implies P\left(\frac{\sigma^2}{S} > \frac{5^2}{S}\right) = 0.5 \\
&\implies P\left(\frac{S}{\sigma^2} < \frac{S}{5^2}\right) = 0.5 \\
&\implies P\left(W < \frac{S}{5^2}\right) = 0.5
\end{aligned}
$$

Recall that W has a chi-square distribution with k degrees of freedom. To get the median of the chi-squared distribution with 1 degree of freedom we use the following code:

```r
round(qchisq(0.5,1),4)
```

```
[1] 0.4549
```

Thus,

$$\begin{aligned} P(W < 0.4549) = 0.5 &\implies \frac{S}{5^2} = 0.4549 \\ &\implies S = 0.4549(5^2) \\ &\implies S \approx 11.37 \end{aligned}$$

Therefore, Carlos' prior for σ^2 will be 11.37 times an inverse chi-squared distribution with 1 degree of freedom.

To summarize the priors and posteriors of the three statisticians, we have

Aroha:

- Prior for σ: $\pi(\sigma) = 1$
- Prior for σ^2: $\pi(\sigma^2) = \frac{1}{2\sigma}$
- Posterior for σ^2:

$$\pi(\sigma^2|y) \propto \frac{1}{(\sigma^2)^{\frac{11}{2}}} e^{-\frac{146.5}{2\sigma^2}}$$

- This is the *146.5 times an inverse chi-square* distribution with 9 degrees of freedom

Bernardo:

- Prior for σ: $\pi(\sigma) = \frac{1}{\sigma}$
- Prior for σ^2: $\pi(\sigma^2) = \frac{1}{2\sigma^2}$
- Posterior for σ^2:

$$\pi(\sigma^2|y) \propto \frac{1}{(\sigma^2)^{\frac{12}{2}}} e^{-\frac{146.5}{2\sigma^2}}$$

- This is the *146.5 times an inverse chi-square* distribution with 10 degrees of freedom

Carlos:

- Prior for σ: $\pi(\sigma) = \frac{1}{(\sigma^2)^{\frac{1}{2}}} e^{-\frac{11.37}{2\sigma^2}}$
- Prior for σ^2: $\pi(\sigma^2) = \frac{1}{(\sigma^2)^{\frac{3}{2}}} e^{-\frac{11.37}{2\sigma^2}}$

- Posterior for σ^2:

$$\pi(\sigma^2|y) \propto \frac{1}{(\sigma^2)^{\frac{13}{2}}} e^{-\frac{11.37+146.5}{2\sigma^2}}$$

This is the *157.87 times an inverse chi-square* distribution with 11 degrees of freedom.

Bayesian estimator for σ

We calculate the measure of location from the posterior distribution of the variance and then take the square root for our estimator of the standard deviation. The three possible measures of location are the posterior mean, posterior mode, and posterior median.

Suppose the posterior distribution is *S' times an inverse chi-square* distribution with $k^{'}$ degrees of freedom, where: $S^{'} = S + SST$ and $k^{'} = n + k$

Then - Posterior mean $(m^{'})$: $m^{'} = \frac{S^{'}}{k^{'}-2} \implies \hat{\sigma} = \sqrt{m^{'}}$

- Posterior mode (mo): $mo = \frac{S^{'}}{k^{'}+2} \implies \hat{\sigma} = \sqrt{mo}$
- Posterior median (md): $\int_0^{md} \pi(\sigma^2|y) d\sigma^2 = 0.5 \implies \hat{\sigma} = \sqrt{md}$

Therefore,

For Aroha: $S^{'} = 146.5$ and $k^{'} = 9$, thus,

$$\begin{aligned} m' &= \frac{S^{'}}{k^{'} - 2} \\ &= \frac{146.5}{9 - 2} \\ &= 20.92857 \end{aligned}$$

$$\implies \hat{\sigma} = \sqrt{20.92857} \approx 4.5748$$

$$\begin{aligned} mo &= \frac{S^{'}}{k^{'} + 2} \\ &= \frac{146.5}{9 + 2} \\ &= 13.31818 \end{aligned}$$

$$\implies \hat{\sigma} = \sqrt{13.31818} \approx 3.6494$$

The median is computed using this code:

```
md <- sqrt(146.5*qinvchisq(0.5,9))
print(round(md,4))
```

```
[1] 4.1905
```

YOUR TASK:

Verify the entries (for Bernardo and Carlos) in the following summary table

Statistician	Mean	Median	Mode
Aroha	4.5748	4.1905	3.6494
Bernardo	4.2793	3.9601	3.4940
Carlos	4.1882	3.9072	3.4848

Bayesian credible interval for σ

Recall that the posterior distribution of the variance σ^2 given the sample data is $S^{'}$ times an inverse chi-square with $k^{'}$ degrees of freedom. Thus, $W = \frac{S^{'}}{\sigma^2}$ has a chi-square distribution with $k^{'}$ degrees of freedom.

Let u be the chi-squared value with area to the right of $1 - \frac{\alpha}{2}$ and let ℓ be the chi-square value with area $\frac{\alpha}{2}$ to the right

$$
\begin{aligned}
P(u < W < \ell) = 1 - \alpha \iff & P\left(u < \frac{S^{'}}{\sigma^2} < \ell\right) = 1 - \alpha \\
\iff & P\left(\frac{S^{'}}{\ell} < \sigma^2 < \frac{S^{'}}{u}\right) = 1 - \alpha \\
\iff & P\left(\sqrt{\frac{S^{'}}{\ell}} < \sigma < \sqrt{\frac{S^{'}}{u}}\right) = 1 - \alpha
\end{aligned}
$$

Bayesian credible interval for σ: an example

- The code chunk below illustrates the principle how to construct a credible interval for σ

```
#Based on Aroha's posterior distribution
u <- qchisq(0.975, 9,lower.tail = F)
l <- qchisq(0.025, 9, lower.tail = F)
LL <- round(sqrt(146.5/l),4)
UL <- round(sqrt(146.5/u),4)
print(cbind(LL,UL))
```

```
         LL     UL
[1,] 2.7751 7.3656
```

The 95% credible intervals for σ for the 3 statisticians are summarized in the table below.

Statistician	LL	UL
Aroha	2.7751	7.3656
Bernardo	2.6744	6.7171
Carlos	2.6837	6.4322

Testing a one-sided hypothesis about σ

Suppose we want to test $H_0 : \sigma \leq \sigma_0$ against $H_1 : \sigma > \sigma_0$. We will test this by calculating the posterior probability of the null hypothesis and comparing this to the level of significance α

$$\begin{aligned}
P(H_0|y_1, y_2, \cdots, y_n) &= P(\sigma \leq \sigma_0|y_1, y_2, \cdots, y_n) \\
&= P(\sigma^2 \leq \sigma_0^2|y_1, y_2, \cdots, y_n) \\
&= P\left(\frac{S'}{\sigma^2} > \frac{S'}{\sigma_0^2}|y_1, y_2, \cdots, y_n\right) \\
&= P(W > W_0), \text{where:} W = \frac{S'}{\sigma^2} \text{ and} W_0 = \frac{S'}{\sigma_0^2}
\end{aligned}$$

Again $W \sim \chi^2$ with k' degrees of freedom.

Testing a one-sided hypothesis about σ: an example

Suppose the three statisticians want to determine if the standard deviation is greater than 2. That is, we wish to test $H_0 : \sigma \leq 2$ versus $H_1 : \sigma > 2$.

Statistician	$P(H_0\|data)$	Decision on H_0
Aroha	3.069768e-05	Reject
Bernardo	6.569889e-05	Reject
Carlos	4.410607e-05	Reject

The above probabilities were computed using this code chunk:

```
a <- pchisq(146.5/4,9,lower.tail=F)#Aroha
b <- pchisq(146.5/4,10,lower.tail=F)#Bernardo
c <- pchisq(157.87/4,11,lower.tail=F)#Carlos
```