

A Two Sample Normal Problem

Suppose we observe two data sets

$$y_{11}, \dots, y_{1n_1}$$

and

$$y_{21}, \dots, y_{2n_2}$$

and we want to assess whether the characteristics of the populations from which we believe the data are drawn are different.

A Two Sample Normal Problem

The classical testing route proceeds as follows:

- ▶ Assume $Y_{1i} \sim \text{Normal}(\mu_1, \sigma_1^2)$, $Y_{2i} \sim \text{Normal}(\mu_2, \sigma_2^2)$
- ▶ Carry out the hypothesis test

$$H_0 : \mu_1 = \mu_2$$

against some stated alternative, say

$$H_a : \mu_1 \neq \mu_2$$

- ▶ Might assume $\sigma_1 = \sigma_2$
- ▶ Complete the test using the Rejection Region or p-value approach.

A Two Sample Normal Problem

A common test statistic is

$$T_{obs} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where

$$\bar{y}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} y_{1i} \quad \bar{y}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} y_{2i}$$

and

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 \quad s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2$$

and a large value of $|T_{obs}|$ supports the conclusion that $\mu_1 \neq \mu_2$.

The Bayesian analogue

We may reasonably assume that the data arise from two *independent* infinitely exchangeable systems.

That is, we assume that for any finite collections

$$\mathbf{Y}_1 = (Y_{11}, \dots, Y_{1n_1})$$

and

$$\mathbf{Y}_2 = (Y_{21}, \dots, Y_{1n_2})$$

we assume that

$$f_{\mathbf{Y}_1, \mathbf{Y}_2}(\mathbf{y}_1, \mathbf{y}_2) = f_{\mathbf{Y}_1}(\mathbf{y}_1)f_{\mathbf{Y}_2}(\mathbf{y}_2)$$

The Bayesian analogue

First assume that σ_1^2 and σ_2^2 are *known*; we may reasonably assume independent Normal *priors* for μ_1 and μ_2

$$\pi_0(\mu_1) \equiv \text{Normal}(\eta_{10}, \sigma_{10}^2/\lambda_{10}) \quad \pi_0(\mu_2) \equiv \text{Normal}(\eta_{20}, \sigma_{20}^2/\lambda_{20})$$

so that by previous results the *posteriors* are also Normal

$$\pi_n(\mu_1) \equiv \text{Normal}(\eta_{1n}, \sigma_{10}^2/\lambda_{1n}) \quad \pi_n(\mu_2) \equiv \text{Normal}(\eta_{2n}, \sigma_{20}^2/\lambda_{2n})$$

with the updated parameters given by the previously established formulae, and with μ_1 and μ_2 being *independent*.

Also, note that the two parameters are *independent*, that is

$$\pi_n(\mu_1, \mu_2) = \pi_n(\mu_1)\pi_n(\mu_2)$$

which follows as

- ▶ the parameters are presumed to have independent priors,
- ▶ the two observable sequences are presumed independent of each other.

Consider the difference parameter

$$\theta = \mu_1 - \mu_2.$$

In light of the model specification and the data, and properties of the Normal distribution, we may deduce that the posterior distribution on θ is also Normal

$$\pi_n(\theta) \equiv \text{Normal}(\eta_{1n} - \eta_{2n}, \sigma_{10}^2/\lambda_{1n} + \sigma_{20}^2/\lambda_{2n})$$

The posterior parameters are given by

$$\eta_{1n} = \frac{n_1 \bar{y}_1 + \lambda_{10} \eta_{10}}{n_1 + \lambda_{10}} \qquad \lambda_{1n} = n_1 + \lambda_{10}$$

and

$$\eta_{2n} = \frac{n_2 \bar{y}_2 + \lambda_{20} \eta_{20}}{n_2 + \lambda_{20}} \qquad \lambda_{2n} = n_2 + \lambda_{20}.$$

The posterior uncertainty in θ is reflected by $\pi_n(\theta)$.

The Bayesian analogue

Now assume that σ_1^2 and σ_2^2 are *unknown*. We might assume

$$\sigma_1^2 = \sigma_2^2$$

for simplicity, but more generally would have to assume that $\sigma_1^2 \neq \sigma_2^2$.

In the latter case, we might specify an *Inverse Gamma* prior

- If $Z \sim \text{Gamma}(\alpha, \beta)$, then $X = 1/Z \sim \text{InvGamma}(\alpha, \beta)$: by standard transformation results

$$f_X(x) = \frac{1}{x^2} f_Z(1/x)$$

- We make this choice for analytical convenience.
- Here we assume that

$$\pi_0(\sigma_j^2) \equiv \text{InvGamma}(\alpha_{j0}, \beta_{j0}) \quad j = 1, 2$$

The Bayesian analogue

After some calculations, we discover that the marginal posterior distribution for $\mu_j, j = 1, 2$ can be computed analytically as

$$\pi_n(\mu_j) = \frac{\Gamma((a_{jn} + 1)/2)}{\Gamma(a_{jn}/2)\sqrt{\pi}} \left(\frac{\lambda_{jn}}{b_{jn}} \right)^{1/2} \left\{ 1 + \frac{\lambda_{jn}}{b_{jn}} (\mu_j - \eta_{jn})^2 \right\}^{-(a_{jn}+1)/2}$$

where

$$a_{jn} = n_j + \alpha_{j0}$$

$$b_{jn} = \frac{n_j \lambda_{j0}}{n_j + \lambda_{j0}} (\bar{y}_j - \eta_{j0})^2 + \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j)^2 + \beta_{j0}$$

Often, this density is rewritten by using a rescaled version of $b_{jn}, j = 1, 2$:

$$\phi_{jn} = \frac{b_{jn}}{a_{jn} \lambda_{jn}}.$$

Then we have

$$\pi_n(\mu_j) = \frac{\Gamma((a_{jn} + 1)/2)}{\Gamma(a_{jn}/2)\sqrt{\pi}} \frac{1}{a_{jn}^{1/2}} \left(\frac{1}{\phi_{jn}} \right)^{1/2} \left\{ 1 + \frac{1}{a_{jn}} \frac{(\mu_j - \eta_{jn})^2}{\phi_{jn}} \right\}^{-(a_{jn}+1)/2}$$

This is the general *Student-t* distribution with a_{jn} degrees of freedom, centered at η_{jn} , with scale parameter $\phi_{jn}^{1/2}$.

In this case, however, the posterior distribution of

$$\theta = \mu_1 - \mu_2$$

is hard to compute, as the difference of two Student-t random variables does not have a standard distribution.

Numerical example

Suppose we have data:

Sample 1: 1.12, 1.25, 3.38, 2.24, 2.11, 3.20

Sample 2: 4.08, 6.28, 5.83, 6.55

These data yield the following statistics:

```
y1<-c(1.12, 1.25, 3.38, 2.24, 2.11, 3.20)
y2<-c(4.08, 6.28, 5.83, 6.55)
n1<-length(y1)
n2<-length(y2)
c(n1,n2)

## [1] 6 4

y1bar<-mean(y1);y2bar<-mean(y2)
c(y1bar,y2bar) #sample variances

## [1] 2.216667 5.685000

c(var(y1),var(y2)) #sample variances

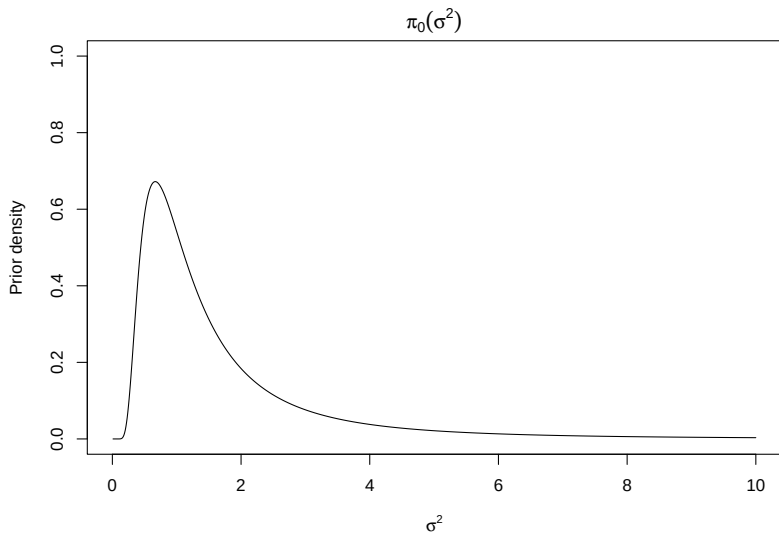
## [1] 0.8938667 1.2331000
```

Numerical example

Prior specification: for simplicity, we might choose identical priors for the unknown parameters:

```
eta10<-eta20<-0
lambda10<-lambda20<-0.1
a110<-2.0;be10<-2
a120<-2.0;be20<-2
yv<-seq(0.01,10,by=0.01)
pi0.sig<-dgamma(1/yv,a110,be10)/yv^2
par(mar=c(4,4,2,2))
plot(yv,pi0.sig,type='l',ylim=range(0,1),xlab=expression(sigma^2),
      ylab='Prior density')
title(expression(pi[0](sigma^2)))
```

Numerical example



Numerical example

Posterior parameters:

```
eta1n<-(n1*y1bar+lambda10*eta10)/(n1+lambda10)
lambda1n<-lambda10+n1
eta2n<-(n2*y2bar+lambda20*eta20)/(n2+lambda20)
lambda2n<-lambda20+n2

c(eta1n, eta2n)

## [1] 2.180328 5.546341

a1n<-a110+n1
b1n<-((n1*lambda10)/(n1+lambda10))*(y1bar-eta10)^2 + sum(y1-y1bar)^2 + be10
a2n<-a120+n2
b2n<-((n2*lambda20)/(n2+lambda20))*(y2bar-eta20)^2 + sum(y2-y2bar)^2 + be20

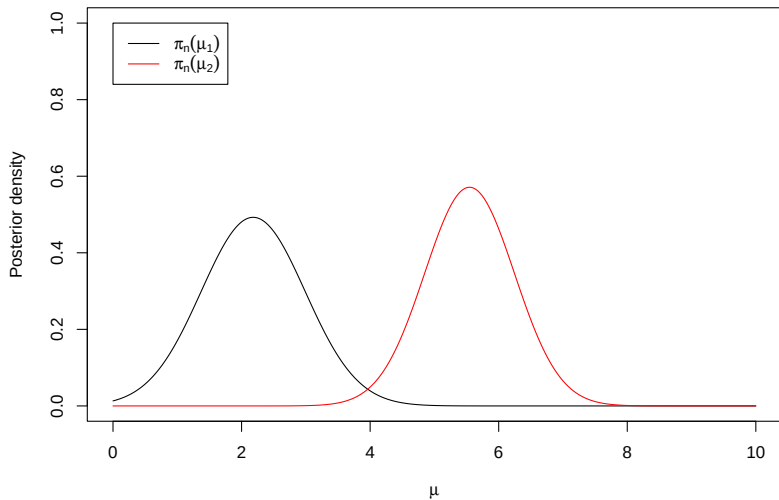
phi1n<-b1n/(a1n*lambda1n)
phi2n<-b2n/(a2n*lambda2n)
```

Numerical example

Known variances: Suppose we assume $\sigma_1^2 = 4$, $\sigma_2^2 = 2$. Using the calculations above, we may compute and plot the posteriors for μ_1 and μ_2 , and for $\theta = \mu_1 - \mu_2$:

```
sig10<-2
sig20<-sqrt(2)
yv<-seq(0,10,by=0.01)
post1<-dnorm(yv,eta1n,sig10/sqrt(lambda1n))
post2<-dnorm(yv,eta2n,sig20/sqrt(lambda2n))
par(mar=c(4,4,2,2))
plot(yv,post1,type='l',ylim=range(0,1),xlab=expression(mu),
     ylab='Posterior density')
lines(yv,post2,type='l',col='red')
legend(0,1,c(expression(pi[n](mu[1])),expression(pi[n](mu[2]))),
     col=c('black','red'),lty=1)
```

Numerical example

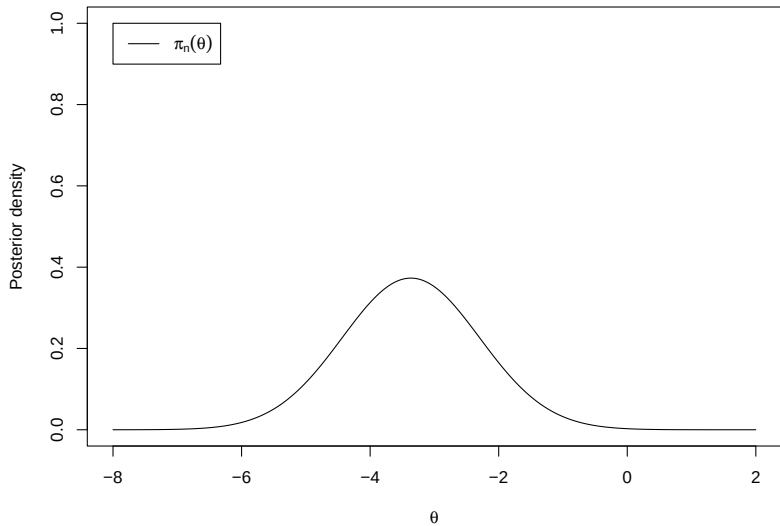


Numerical example

Posterior for θ :

```
yv<-seq(-8,2,by=0.01)
post.th<-dnorm(yv,eta1n-eta2n,sqrt(sig10^2/lambda1n+sig20^2/lambda2n))
par(mar=c(4,4,2,2))
plot(yv,post.th,type='l',ylim=range(0,1),xlab=expression(theta),
      ylab='Posterior density')
legend(-8,1,c(expression(pi[n](theta))),lty=1)
```

Numerical example



Numerical example

Unknown variances: Need to use the Student-t distributions.

The `dt` function computes the standard Student-t density, and we need the *location-scale* transformed version: if

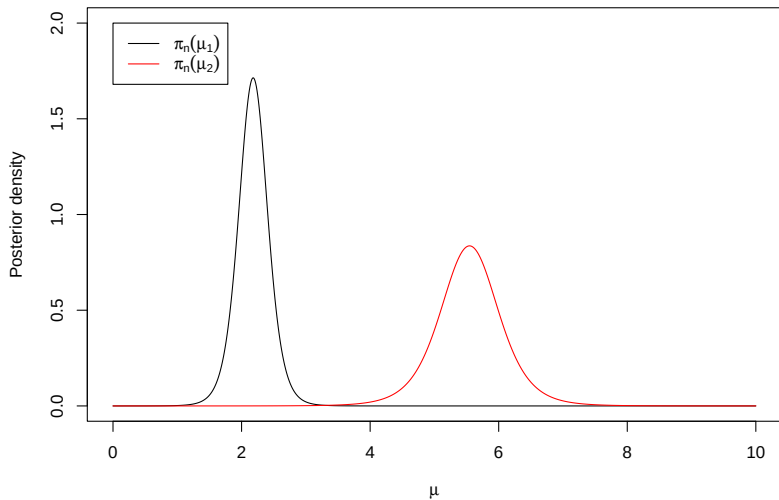
$$X = cZ + m$$

for constants $c > 0$ and m , then

$$f_X(x) = \frac{1}{c} f_Z((x - m)/c)$$

```
yv<-seq(0,10,by=0.01)
post1<-dt((yv-eta1n)/sqrt(phi1n),df=a1n)/sqrt(phi1n)
post2<-dt((yv-eta2n)/sqrt(phi2n),df=a2n)/sqrt(phi2n)
par(mar=c(4,4,2,2))
plot(yv,post1,type='l',ylim=range(0,2),xlab=expression(mu),ylab='Posterior density')
lines(yv,post2,type='l',col='red')
legend(0,2,c(expression(pi[n](mu[1])),expression(pi[n](mu[2]))),
      col=c('black','red'),lty=1)
```

Numerical example

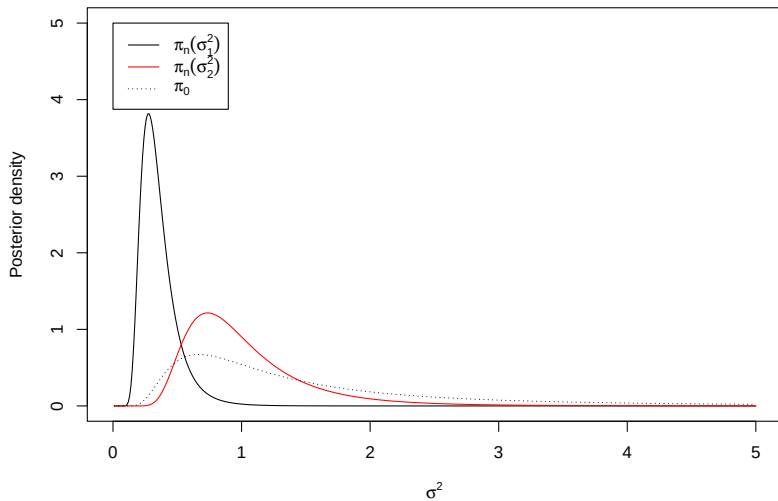


Numerical example

For the posteriors for σ_1^2 and σ_2^2 :

```
yv<-seq(0,5,by=0.01)
post1.sig<-dgamma(1/yv,a1n,b1n)/yv^2
post2.sig<-dgamma(1/yv,a2n,b2n)/yv^2
par(mar=c(4,4,2,2))
plot(yv,post1.sig,type='l',ylim=range(0,5),
      xlab=expression(sigma^2),ylab='Posterior density')
lines(yv,post2.sig,type='l',col='red')
lines(yv,dgamma(1/yv,a110,b110)/yv^2,lty=3)
legend(0,5,c(expression(pi[n](sigma[1]^2)),
              expression(pi[n](sigma[2]^2)),
              expression(pi[0])),col=c('black','red'),lty=c(1,1,3))
```

Numerical example



Numerical example

For θ , it is not easy to compute the posterior distribution as the difference of two Student-t random variables does not have a standard distribution.

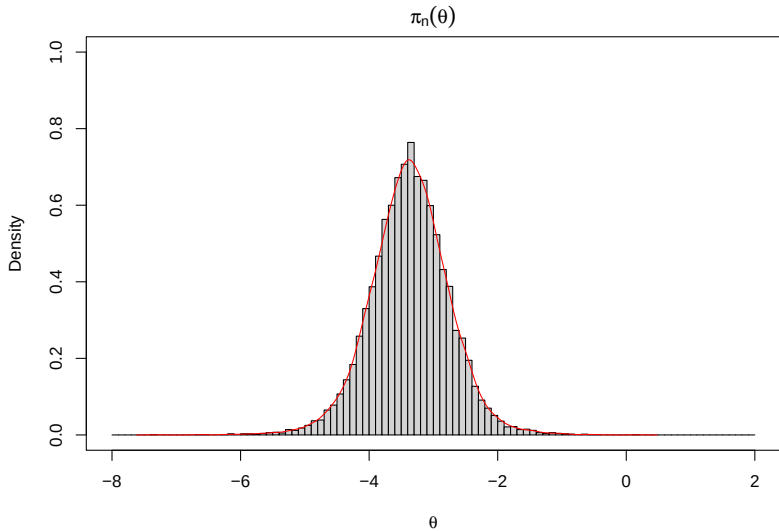
Instead of using analytical calculations, we use *sampling*:

- ▶ draw large samples from each of the two posteriors
- ▶ compute the differences
- ▶ plot a histogram/density estimate of the differenced samples.

Numerical example

```
N<-10000
post1.samp<-rt(N,df=a1n)*sqrt(phi1n)+eta1n
post2.samp<-rt(N,df=a2n)*sqrt(phi2n)+eta2n
theta.samp<-post1.samp-post2.samp
dest<-density(theta.samp,bw=0.1)
par(mar=c(4,4,2,2))
hist(theta.samp,breaks=seq(-8,2,by=0.1),ylim=range(0,1),
      main=expression(pi[n](theta)),xlab=expression(theta),freq=FALSE)
lines(dest,col='red')
box()
```

Numerical example



Numerical example

To complete the analysis, we examine $\pi_n(\theta)$:

- ▶ does the value 0 lie in the *high probability* region ?
- ▶ what is the *variance* of the posterior distribution ?

All reporting and decisions are performed via the posterior distribution.