STAT 8700 Homework 4

Brian Detweiler

Friday, September 23, 2016

1. Consider data from a Normal population with unknown mean μ and variance σ^2. A random sample of 100 observations is taken from this population, and the sample mean and variance were calculated to be 50 and 25 respectively.

(a) If we choose to use a $N - Inv - \chi^2(40, 0.64, 1, 16)$ prior distribution, write down the corresponding posterior distribution.

We are given the following:

$$
\begin{aligned}
n &= 100 \\
\overline{y} &= 50 \\
s^2 &= 25 \\
\mu_0 &= 40 \\
\kappa_0 &= 25 \\
\nu_0 &= 1 \\
\sigma_0^2 &= 16
\end{aligned}
$$

Now we can use these values to calculate the joint posterior distribution, $N - Inv - \chi^2(\mu_n, \sigma_n^2/\kappa_n; \nu_n, \sigma_n^2)$:

$$
\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\overline{y} \\
&= \frac{25}{25 + 100}40 + \frac{100}{25 + 100}50 = 48 \\
\kappa_n &= \kappa_0 + n \\
&= 25 + 100 = 125 \\
\nu_n &= \nu_0 + n \\
&= 1 + 100 = 101 \\
\nu_n \sigma_n^2 &= \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\overline{y}\mu_0)^2 \\
&= 1(16) + (100 - 1)25 + \frac{25(100)}{25 + 100}(50 - 40)^2 = 4491 \\
\sigma_n^2 &\approx 44.4653465347
\end{aligned}
$$

And thus our joint posterior distributin is $N - Inv - \chi^2(48, 0.355722772278; 101, 44.4653465347)$.

■

(b) Either analytically or via simulation, construct 95% credible intervals for σ^2 and μ.

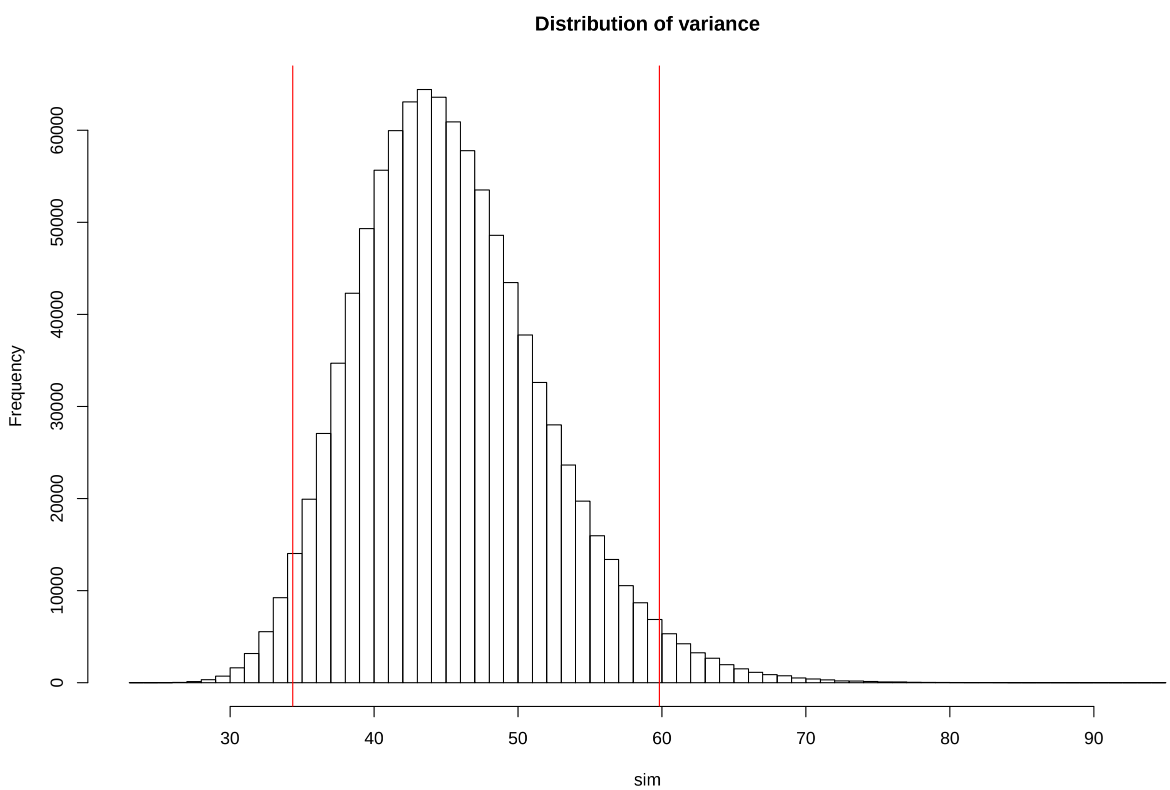
To simulate this, we first draw σ^2 from its marginal posterior distribution, $\sigma^2|y \sim Inv-\chi^2(\nu_n, \sigma_n^2)$

```
library(geoR)

set.seed(124)

x <- seq(0, 100, by = 0.001)
nu_n <- 101
sigma_n_2 <- 44.4653465347

sim <- rinvchisq(n = 1000000, df = nu_n, scale = sigma_n_2)
hist(sim, breaks = 90, main = 'Distribution of variance')
lower <- sort(sim)[25000]
upper <- sort(sim)[975000]
abline(v=lower, col='red')
abline(v=upper, col='red')
```



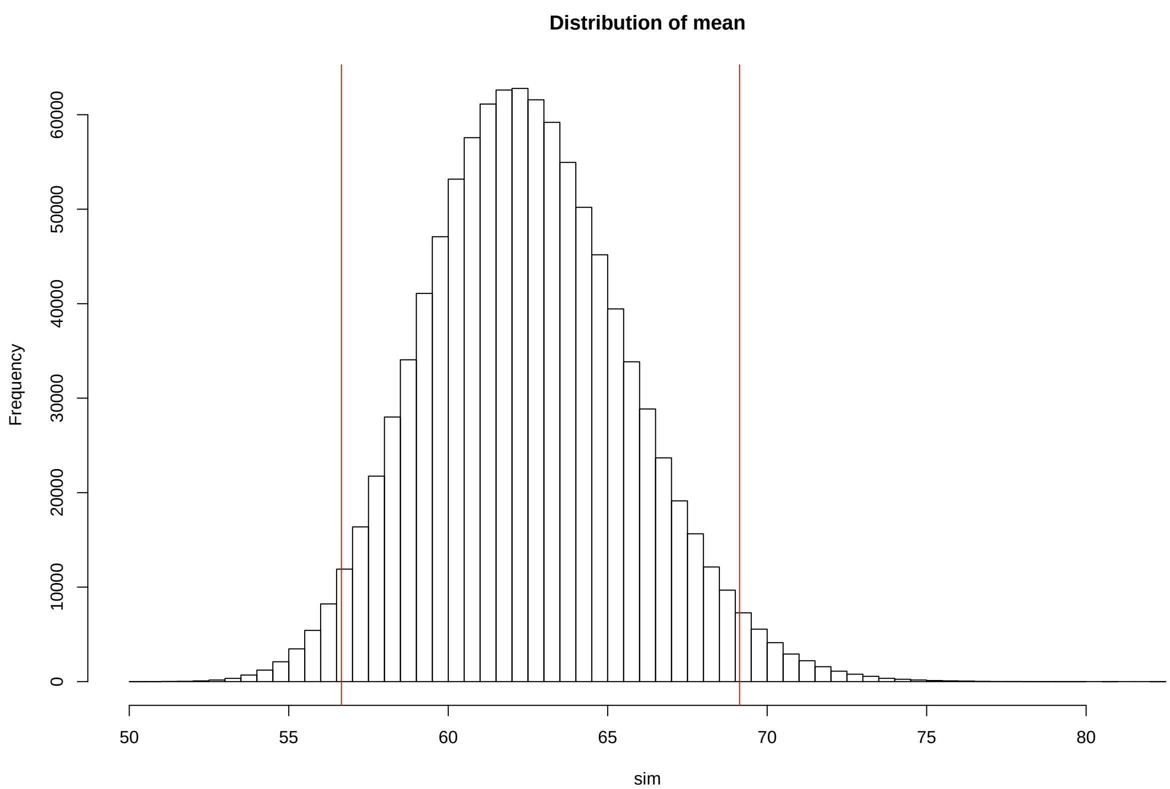
A 95% credible interval for σ^2 is (34.3564188, 59.8085225).

Then we sample from $N\left(\frac{\frac{\kappa_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\overline{y}}{\frac{\kappa_0}{\sigma^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma^2}+\frac{n}{\sigma^2}}\right)$ using the previous values for σ^2.

```
sigma_2 <- sim
kappa_0 <- 25
```

```r
mu_0 <- 40
n <- 100
y_bar <- 50
mu_n <- ((kappa_0 / sigma_2) * mu_0) + ((n / sigma_2) * y_bar) / ((kappa_0 / sigma_2) + (n / sigma_2))
sigma_2_kappa_n <- 1 / ((kappa_0 / sigma_2) + (n / sigma_2))
sim <- rnorm(n = 1000000, mu_n, sigma_2_kappa_n)
hist(sim, breaks = 90, main = 'Distribution of mean')

lower <- sort(sim)[25000]
upper <- sort(sim)[975000]
abline(v=lower, col='red')
abline(v=upper, col='red')
```



A 95% credible interval for μ is (56.6518857, 69.1340067).

■

2. Two random variables are said to have a bivariate normal distribution with parameters, $\mu_U, \mu_V, \sigma_U^2, \sigma_V^2$, and ρ if they have the following density function:

$$f(u,v) = \frac{1}{2\pi\sigma_U\sigma_V\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(u-\mu_U)^2}{\sigma_U^2} + \frac{(v-\mu_V)^2}{\sigma_V^2} - \frac{2\rho(u-\mu_U)(v-\mu_V)}{\sigma_U\sigma_V}\right]}$$

where μ_U and σ_U^2 are the mean and variance of U, μ_V and σ_V^2 are the mean and variance of V, and ρ is the correlation between U and V.

Replace the uniform prior on α and β in the analysis of the bioassay by a bivariate normal prior with $\alpha \sim Normal(0,4), \beta \sim Normal(10,100)$, and $corr(\alpha,\beta) = 0.5$. Repeat all the computations and plots discussed in section 3.7 and in class.

```r
x <- c(-0.86, -0.3, -0.05, 0.73)
n <- c(5, 5, 5, 5)
y <- c(0, 1, 3, 5)
data <- cbind(y, n - y)
fit <- glm(data ~ x, family = binomial)
# This is the MLE:
coef(fit)
```

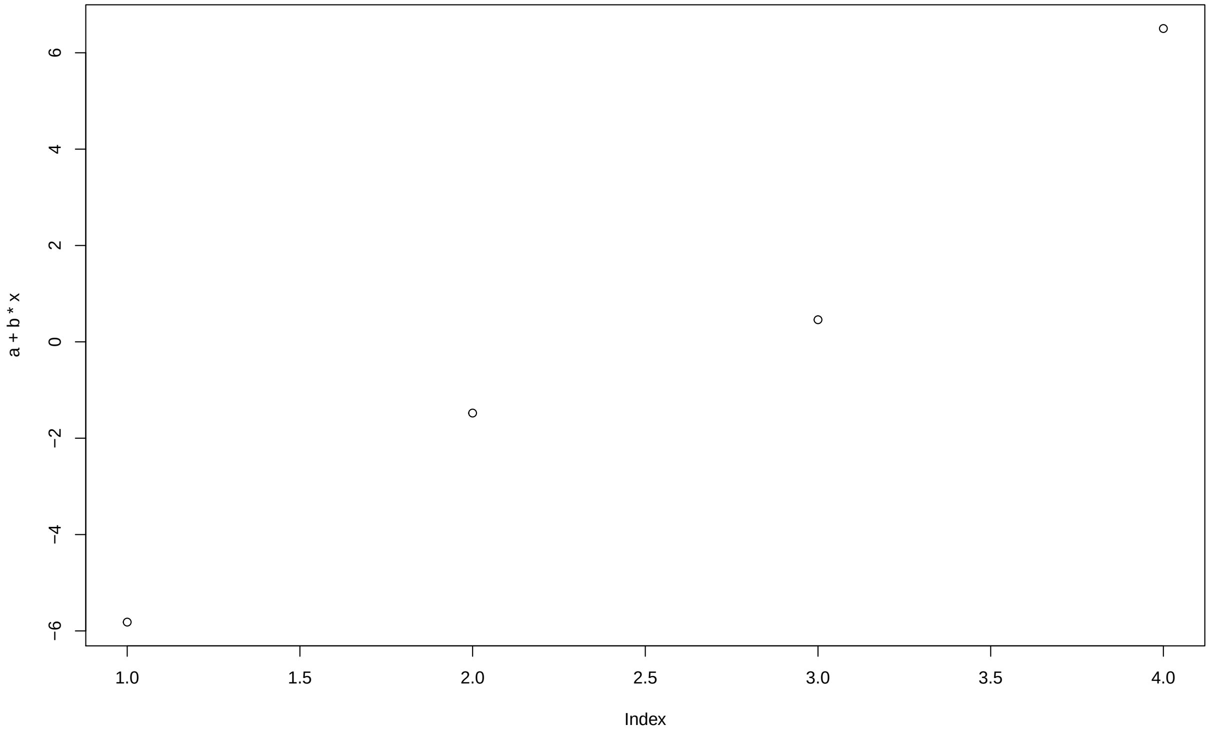
```
## (Intercept)           x
##   0.8465802   7.7488172
```

```r
# This is the approximate covariance matrix:
summary(fit)$cov.unscaled
```

```
##             (Intercept)         x
## (Intercept)    1.038535  3.545987
## x              3.545987 23.743865
```

```r
a <- coef(fit)[[1]]
b <- coef(fit)[[2]]

plot(a + b*x)
```




```r
mu <- 0
sigma_2 <- 4

a <- rnorm(n = 1000, mu, sqrt(sigma_2))
#hist(a, breaks = 90)

mu <- 10
sigma_2 <- 100

b <- rnorm(n = 1000, mu, sqrt(sigma_2))
#hist(b, breaks = 90)
#plot(a + b*5)

f <- function(u, v) {
  mu_U <- 0
  mu_V <- 10

  sigma_2_U <- 4
  sigma_2_V <- 100
  sigma_U <- sqrt(sigma_2_U)
  sigma_V <- sqrt(sigma_2_V)

  rho <- 0.5

  first <- 1 / (2 * pi * sigma_2_U * sigma_2_V * sqrt(1 - rho^2))
  second <- exp(-(1/2*(1 - rho^2)) * (((u - mu_U)^2 / sigma_2_U)
              + ((v - mu_V)^2 / sigma_2_V)
```

```
                - (2 * rho * (u - mu_U) * (v - mu_V)) / sqrt(sigma_U) * sqrt(sigma_V)))
  rval <- first * second
  return(rval)
}

x <- c()

for (i in 0:100) {
  for (j in 0:100) {
    x <- c(x, f(i, j))
  }
}
#hist(x, breaks=100)
```

3. Consider the airline fatalities data discussed in the previous exercise. Let us suppose that we now assume that the number of fatal accidents in year t follows a Poisson distribution with mean $\alpha + \beta t$.

(a) If we let y_t represent the number of fatal accidents in year t, write down $p(y_t|\alpha, \beta)$ the likelihood for year t in terms of the parameters α, and β.

(b) If we assume uniform priors on α and β, write the posterior density for (α, β).

(c) Following the same idea as the boassay example (and the previous question) create a grid of possible α and β values on which to evaluate the joint posterior and plot the contours. Start with large ranges for α and β and refine based on the countour plot. Include all your iterations in your answer, not just you final grid and contour plot.

(d) Simulate 100,000 values of α and β from the joint posterior and plot the histogram of the posterior density of the expected number of fatal accidents in 1986, $\alpha + 1986\beta$.

(e) Use your simulated values of α and β to simulate the number of fatal accidents in 1986. Use your simulations to construct a 95% predictive (credible) interval.

(f) Return to your simulated values of β, calculate (well, estimate) $P(\beta < 0)$, that is, the probability that the number of fatal accidents per year is decreasing.