Exercises from ‘Bayesian Data Analysis’ (Ch. 2-3)

STK9021 Applied Bayesian Analysis and Numerical Methods

Jeriek Van den Abeele

October 6, 2016

University of Oslo

Ex. 2.11: Computing with a non-conjugate single-parameter model

Suppose $y_1, \dots, y_5$ are independent samples from a Cauchy distribution with unknown center θ and known scale 1: $p(y_i|\theta) \propto 1/(1 + (y_i - \theta)^2)$. Assume, for simplicity, that the prior distribution for θ is uniform on $[0, 100]$. Given the observations $(y_1, \dots, y_5) = (43, 44, 45, 46.5, 47.5)$:

(a) Compute the unnormalised posterior density function, $p(\theta)p(y|\theta)$, on a grid of points $\theta = 0, \frac{1}{m}, \frac{2}{m}, \dots, 100$ for some large integer m. Using the grid approximation, compute and plot the normalized posterior density function, $p(\theta|y)$, as a function of θ.

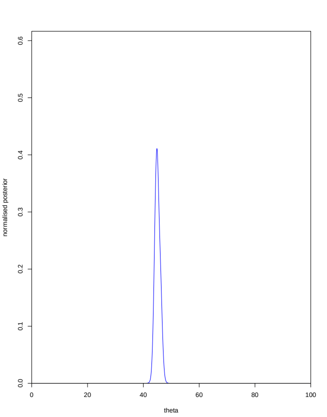
Ex. 2.11: Computing with a non-conjugate single-parameter model

(a) Compute the unnormalised posterior density function, $p(\theta)p(y|\theta)$, on a grid of points $\theta = 0, \frac{1}{m}, \frac{2}{m}, \ldots, 100$ for some large integer m. Using the grid approximation, compute and plot the normalized posterior density function, $p(\theta|y)$, as a function of θ.

```r
## Exercise 2.11a
# Sampling distribution
dist <- function (y, th){
  dist0 <- NULL
  for (i in 1:length(th))
    dist0 <- c(dist0, prod( dcauchy(y, th[i])))
  dist0}
# Data
y <- c(43,44,45,46.5,47.5)
# Parameter grid
step <- .01
theta <- seq(0, 100, step)
# Unnormalised density
dist.unnorm <- dist(y,theta)
# Normalised density
dist.norm <- dist.unnorm/(step*sum(dist.unnorm))
# Plot
plot(theta, dist.norm, ylim=c(0,1.5*max(dist.norm)),
      type="l", xlab="theta", ylab="normalised posterior",
      xaxs="i", yaxs="i", col="blue")
```

Ex. 2.11: Computing with a non-conjugate single-parameter model

(a) Compute the unnormalised posterior density function, $p(\theta)p(y|\theta)$, on a grid of points $\theta = 0, \frac{1}{m}, \frac{2}{m}, \ldots, 100$ for some large integer m. Using the grid approximation, compute and plot the normalized posterior density function, $p(\theta|y)$, as a function of θ.



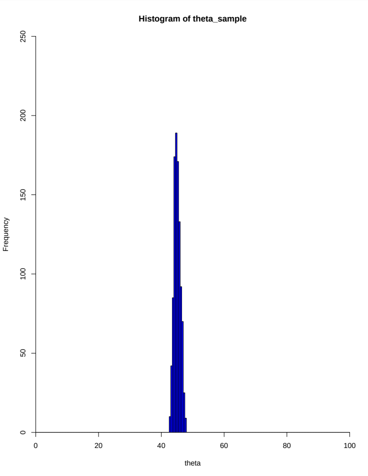
Ex. 2.11: Computing with a non-conjugate single-parameter model

(b) Sample 1000 draws of θ from the posterior density and plot a histogram of the draws.

```
## Exercise 2.11b
# Sampling
theta_sample <- sample(theta, 1000, step*dist.norm,
                       replace=TRUE)
# Histogram
hist(theta_sample, xlab="theta", breaks=seq(0,100,0.5),
     xaxs="i", yaxs="i", ylim=c(0,250), col="blue")
hist(theta_sample, xlab="theta", breaks=seq(40,50,0.5),
     xaxs="i", yaxs="i", ylim=c(0,250), col="blue")
```

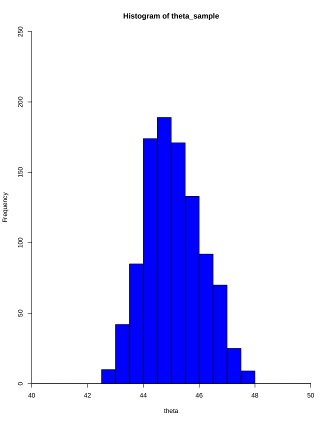
Ex. 2.11: Computing with a non-conjugate single-parameter model

(b) Sample 1000 draws of θ from the posterior density and plot a histogram of the draws.



Ex. 2.11: Computing with a non-conjugate single-parameter model

(b) Sample 1000 draws of θ from the posterior density and plot a histogram of the draws.



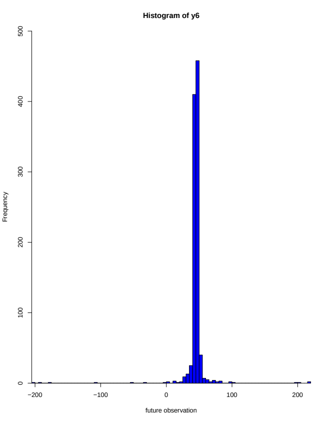
Ex. 2.11: Computing with a non-conjugate single-parameter model

(c) Use the 1000 samples of θ to obtain 1000 samples from the predictive distribution of a future observation, y_6, and plot a histogram of the predictive draws.

```
## Exercise 2.11c
#
y6 <- rcauchy(length(theta_sample), theta_sample, 1)
hist(y6, xlab="future observation", nclass=100,
     xaxs="i", yaxs="i", ylim=c(0,500), col="blue")
# Cauchy distribution has fat tails
```

Ex. 2.11: Computing with a non-conjugate single-parameter model

(c) Use the 1000 samples of θ to obtain 1000 samples from the predictive distribution of a future observation, y_6, and plot a histogram of the predictive draws.



Ex. 2.12: Jeffrey's prior distributions

Suppose $y|\theta \sim \text{Poisson}(\theta)$. Find Jeffreys' prior density for θ, and then find α and β for which the Gamma(α, β) density is a close match to Jeffreys' density.

Poisson distribution: $p(y|\theta) = \frac{\theta^y e^{-\theta}}{y!}$

$\rightarrow$ $J(\theta) = E\left[-\frac{d^2 \log p(y|\theta)}{d\theta^2}|\theta\right] = E\left[\frac{y}{\theta^2}\right] = \frac{1}{\theta}$

Jeffreys' prior: $p(\theta) \propto [J(\theta)]^{1/2} = \theta^{-1/2}$

Compare to Gamma distribution: $p(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$

$\rightarrow$ $\alpha = 1/2$ and $\beta = 0$

Ex. 3.3: Estimation from 2 independent experiments

Effects of magnetic fields on calcium flow in chicken brains

1. Control group of 32 untreated chickens
2. Exposed group of 36 treated chickens

   One measurement is taken on each chicken.

   Goal: measure average flows μ_c and μ_t in the two groups.

   1. Control group: 32 measurements; sample mean 1.013 and std 0.24
   2. Exposed group: 36 measurements; sample mean 1.173 and std 0.20

Ex. 3.3: Estimation from 2 independent experiments

(a) Assuming control measurements taken at random from a normal distribution with mean μ_c and variance σ_c^2, what is the posterior distribution of μ_c? Similarly, use the treatment group measurements to determine the marginal posterior distribution of μ_t. Assume a uniform prior distribution on $(\mu_c, \mu_t, \log \sigma_c, \log \sigma_t)$.

Distribution of data: $\prod_{i=1}^{32} N(y_{ci}|\mu_c, \sigma_c^2) \prod_{j=1}^{36} N(y_{tj}|\mu_t, \sigma_t^2)$

Joint posterior distribution:

$$
\begin{aligned}
p(\mu_c, \mu_t, \log \sigma_c, \log \sigma_t | y) &= p(\mu_c, \mu_t, \log \sigma_c, \log \sigma_t) p(y|\mu_c, \mu_t, \log \sigma_c, \log \sigma_t) \\
&= \prod_{i=1}^{32} N(y_{ci}|\mu_c, \sigma_c^2) \prod_{j=1}^{36} N(y_{tj}|\mu_t, \sigma_t^2) \\
&= p(\mu_c, \log \sigma_c | y) p(\mu_t, \log \sigma_t | y)
\end{aligned}
$$

→ Joint posterior factorises, hence (μ_c, σ_c^2) and (μ_t, σ_t^2) can be treated independently in the posterior.

Ex. 3.3: Estimation from 2 independent experiments

(a) Assuming control measurements taken at random from a normal distribution with mean μ_c and variance σ_c^2, what is the posterior distribution of μ_c? Similarly, use the treatment group measurements to determine the marginal posterior distribution of μ_t. Assume a uniform prior distribution on $(\mu_c, \mu_t, \log\sigma_c, \log\sigma_t)$.

Results from §3.2 on normal data with non-informative prior:

$$\frac{\mu - \bar{y}}{s/\sqrt{n}} \sim t_{n-1}(0,1)$$

Hence, the marginal posteriors of μ_c and μ_t are t densities:

$$\mu_c|y \sim t_{31}(1.013, 0.24^2/32) = t_{31}(1.013, 0.042^2)$$

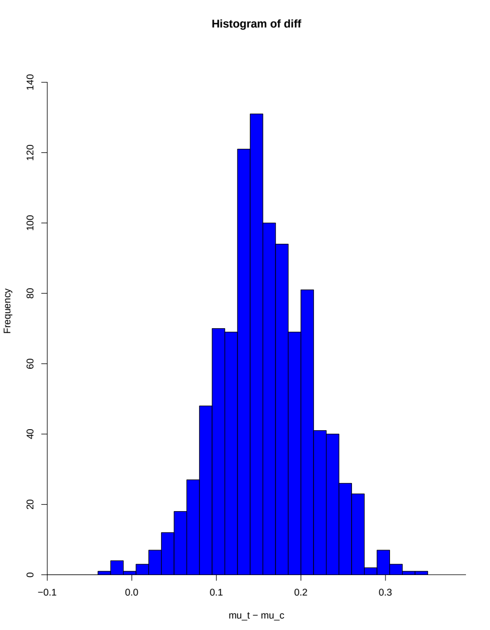
$$\mu_t|y \sim t_{35}(1.173, 0.20^2/36) = t_{35}(1.173, 0.033^2)$$

Ex. 3.3: Estimation from 2 independent experiments

(b) What is the posterior distribution for the difference, $\mu_t - \mu_c$? To get this, you may sample from the independent Student-t distributions you obtained in part (a) above. Plot a histogram of your samples and give an approximate 95% posterior interval for $\mu_t - \mu_c$.

```
## Exercise 3.3b
# Average flow samples in control/untreated group from normal dist
mu.c <- 1.013+(0.24/sqrt(32))*rt(1000,31)
# Average flow samples in exposed/treated group from normal dist
mu.t <- 1.173+(0.20/sqrt(36))*rt(1000,35)
# Histogram of difference
diff <- mu.t - mu.c
hist (diff, xlab="mu_t - mu_c", xaxs="i", yaxs="i", ylim=c(0,150),
      breaks=seq(-.1,.4,.015), col="blue")
# 95% posterior interval of difference
print( sort(diff)[c(25,976)] )
#[0.04539115, 0.26858123]
```

Ex. 3.3: Estimation from 2 independent experiments



Ex. 3.9: Conjugate normal model

Suppose y is an independent and identically distributed sample of size n from the distribution $N(\mu, \sigma^2)$, where (μ, σ^2) have the N-Inv-$\chi^2(\mu_0, \sigma_0^2/\kappa_0; nu_0, \sigma_0^2)$ prior distribution.

- Hence, $\sigma^2 \sim$ Inv-$\chi^2(\nu_0, \sigma_0^2)$ and $\mu|\sigma^2 \sim$ N$(\mu_0, \sigma^2/\kappa_0)$.

The posterior distribution, $p(\mu, \sigma^2|y)$, is also normal-inverse-χ^2; derive explicitly its parameters in terms of the prior parameters and the sufficient statistics of the data. Starting from (3.2) and (3.7):

$$
\begin{aligned}
p(\mu, \sigma^2|y) &\propto p(\mu, \sigma^2) p(y|\mu, \sigma^2) \\
&\propto (\sigma^2)^{-n/2} \exp\left(-\frac{(n-1)s^2 + n(\mu - \bar{y})^2}{2\sigma^2}\right) \times \\
&\sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{\nu_0\sigma_0^2 + \kappa_0(\mu - \mu_0)^2}{2\sigma^2}\right)
\end{aligned}
$$

Ex. 3.9: Conjugate normal model

$$
\begin{aligned}
p(\mu, \sigma^2|y) &\propto p(\mu, \sigma^2)p(y|\mu, \sigma^2) \\
&\propto (\sigma^2)^{-n/2} \exp\left(-\frac{(n-1)s^2 + n(\mu - \bar{y})^2}{2\sigma^2}\right) \times \\
&\sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{\nu_0\sigma_0^2 + \kappa_0(\mu - \mu_0)^2}{2\sigma^2}\right) \\
&\propto \sigma^{-1}(\sigma^2)^{-((\nu_0+n)/2+1)} \exp\left(-\frac{(*)}{2\sigma^2}\right)
\end{aligned}
$$

where

$$
(*) = \nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0(\bar{y} - \mu_0)^2}{n + \kappa_0} + (n + \kappa_0)\left(\mu - \frac{n\bar{y} + \mu_0\kappa_0}{n + \kappa_0}\right)^2
$$

Ex. 3.9: Conjugate normal model

$$p(\mu, \sigma^2 | y) \propto \sigma^{-1} (\sigma^2)^{-((\nu_0+n)/2+1)} \exp\left(-\frac{(*)}{2\sigma^2}\right)$$

where

$$(*) = \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0(\bar{y} - \mu_0)^2}{n+\kappa_0} + (n+\kappa_0)\left(\mu - \frac{n\bar{y} + \mu_0\kappa_0}{n+\kappa_0}\right)^2.$$

Comparing to (3.6), this can be brought expressed as

$$\mu, \sigma^2 | y \sim \text{N-Inv-}\chi^2\left(\frac{\mu_0\kappa_0 + n\bar{y}}{n+\kappa_0}, \frac{\sigma_n^2}{n+\kappa_0}; n+\nu_0, \sigma_n^2\right)$$

with

$$\sigma_n^2 := \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0(\bar{y}-\mu_0)^2}{n+\kappa_0}}{n+\nu_0}.$$