BDA3 Solutions

Adam Bartonicek

Last updated: 2021-09-03

Contents

Chapter 4

4.1 Normal approximation: Cauchy

We observe 5 independent observations from Cauchy distribution with an unknown parameter θ: $(y_1, \ldots, y_5) = (-2, -1, 0, 1.5, 2.5)$.

(a)

Determine the first and second derivative of log-posterior density:

$$p(\mathbf{y}|\theta) = \prod_{i=1}^{5} \frac{1}{1 + (y_i - \theta)^2}$$

$$log(p(\mathbf{y}|\theta)) = \sum_{i=1}^{5} log\left(\frac{1}{1 + (y_i - \theta)^2}\right) = -\sum_{i=1}^{5} log(1 + (y_i - \theta)^2)$$

$$\frac{\partial}{\partial\theta} log(p(\mathbf{y}|\theta)) = -\sum_{i=1}^{5} \frac{\partial}{\partial\theta} log(1 + (y_i - \theta)^2)$$

$$= -\sum_{i=1}^{5} \frac{1}{1 + (y_i - \theta)^2} \cdot 2(y_i - \theta) \cdot (-1)$$

$$= 2\sum_{i=1}^{5} \frac{y_i - \theta}{1 + (y_i - \theta)^2}$$

$$\frac{\partial^2}{\partial\theta^2}log(p(\mathbf{y}|\theta)) = 2\sum_{i=1}^{5}\frac{\partial}{\partial\theta}\frac{y_i-\theta}{1+(y_i-\theta)^2}$$

$$= 2\sum_{i=1}^{5}\frac{(-1)[1+(y_i-\theta)^2]-2\cdot(y_i-\theta)\cdot(-1)\cdot(y_i-\theta)}{[1+(y_i-\theta)^2]^2}$$

$$= 2\sum_{i=1}^{5}\frac{(y_i-\theta)^2-1}{[1+(y_i-\theta)^2]^2}$$

(b)

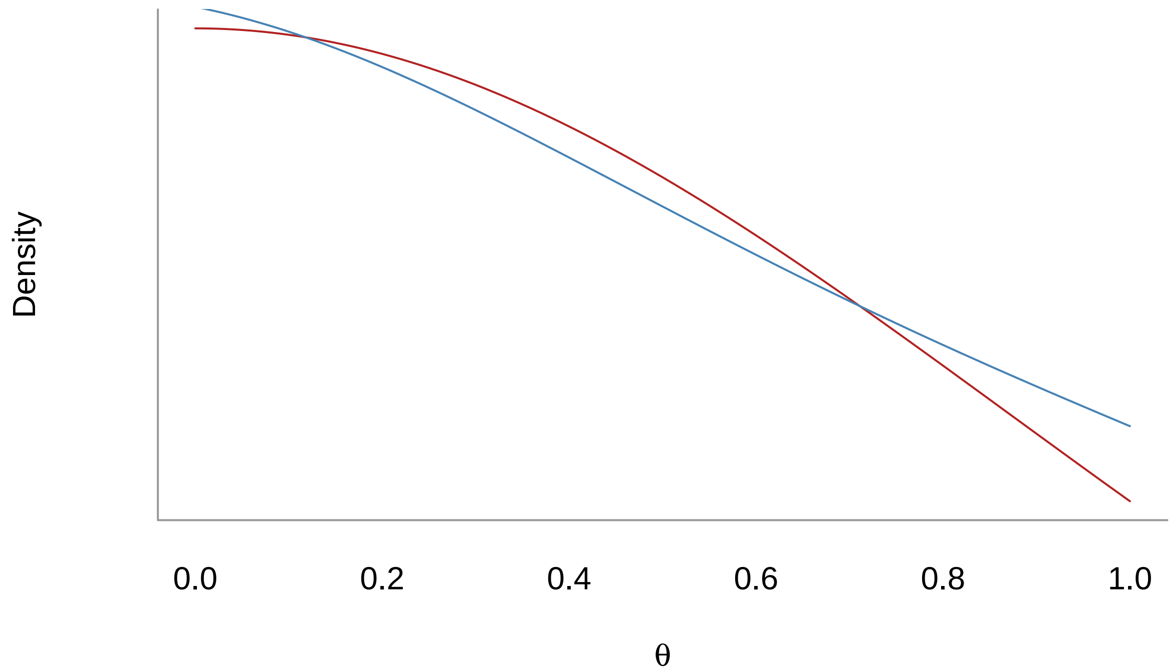
To find the posterior mode, we can use numerical optimization:

```
y <- c(-2, -1, 0, 1.5, 2.5)

scorefun <- function(theta) {
  if (theta < 0 || theta > 1) return(Inf)
  2 * sum((y - theta) / (1 + (y - theta)^2)^2)
}

mode <- uniroot(scorefun, c(0, 1))$f.root
```

(c)

Calculate the normal approximation:

```
I <- -2 * sum(((y - mode)^2 - 1) / (1 + (y - mode)^2)^2)
var_theta <- 1 / I

theta <- seq(0, 1, 0.01)

post_true <- sapply(theta, function(x) exp(-sum(log(1 + (y - x)^2))))
post_true <- post_true / sum(post_true)
post_approx <- dnorm(theta, mode, sqrt(var_theta))
post_approx <- post_approx / sum(post_approx)

plot(theta, post_approx, type = 'l',
     axes = FALSE, col = 'firebrick',
     xlab = expression(theta), ylab = 'Density')
lines(theta, post_true, col = 'steelblue', type = 'l')
axis(1, tick = FALSE)
box(bty = 'L', col = 'grey60')
```



4.2 Normal approximation: Bioassay

We have four observations from four independent experiments:

$$\mathbf{y} = (0, 1, 3, 5)$$
$$\mathbf{n} = (5, 5, 5, 5)$$
$$\mathbf{x} = (-0.86, -0.3, -0.05, 0.73)$$

$$y_i|\theta_i \sim \text{Binomial}(n_i, \theta_i)$$
$$log\left(\frac{\theta_i}{1-\theta_i}\right) = \alpha + \beta x_i \implies \theta_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$$

The likelihood is:

$$p(\mathbf{y}|\alpha, \beta) = \prod_{i=1}^{4} \binom{5}{y_i} \left(\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\right)^{y_i} \left(1 - \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\right)^{5-y_i}$$
$$= \prod_{i=1}^{4} \binom{5}{y_i} \left(\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\right)^{y_i} \left(\frac{1}{1+e^{\alpha+\beta x_i}}\right)^{5-y_i}$$

$$log(p(\mathbf{y}|\alpha,\beta)) \propto \sum_{i=1}^{4} y_i \cdot (\alpha + \beta x_i) - y_i \cdot log(1+e^{\alpha+\beta x_i}) - 5 \cdot log(1+e^{\alpha+\beta x_i}) + y_i \cdot log(1+e^{\alpha+\beta x_i})$$
$$= \sum_{i=1}^{4} y_i \cdot (\alpha+\beta x_i) - 5 \cdot log(1+e^{\alpha+\beta x_i})$$

$$\frac{\partial}{\partial\alpha} log(p(y_i|\alpha,\beta)) = y_i - 5 \cdot \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$$
$$\frac{\partial}{\partial\beta} log(p(y_i|\alpha,\beta)) = y_i x_i - 5 \cdot \frac{x_i \cdot e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$$

$$\frac{\partial^2}{\partial\alpha^2} log(p(y_i|\alpha,\beta)) = -5 \cdot \frac{e^{\alpha+\beta x_i}(1+e^{\alpha+\beta x_i}) - e^{\alpha+\beta x_i} \cdot e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2}$$
$$= -\frac{5e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2}$$

$$\frac{\partial^2}{\partial\alpha\partial\beta} log(p(y_i|\alpha,\beta)) = -5 \cdot \frac{x_i e^{\alpha+\beta x_i}(1+e^{\alpha+\beta x_i}) - x_i e^{\alpha+\beta x_i} e^{\alpha+betax_i}}{(1+e^{\alpha+\beta x_i})^2}$$
$$= -\frac{5x_i e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2}$$

$$\frac{\partial^2}{\partial\beta^2} log(p(y_i|\alpha,\beta)) = -5 \cdot \frac{x_i^2 e^{\alpha+\beta x_i}(1+e^{\alpha+\beta x_i}) - x_i e^{\alpha+\beta x_i} \cdot x_i e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2}$$
$$= -\frac{5x_i^2 e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2}$$

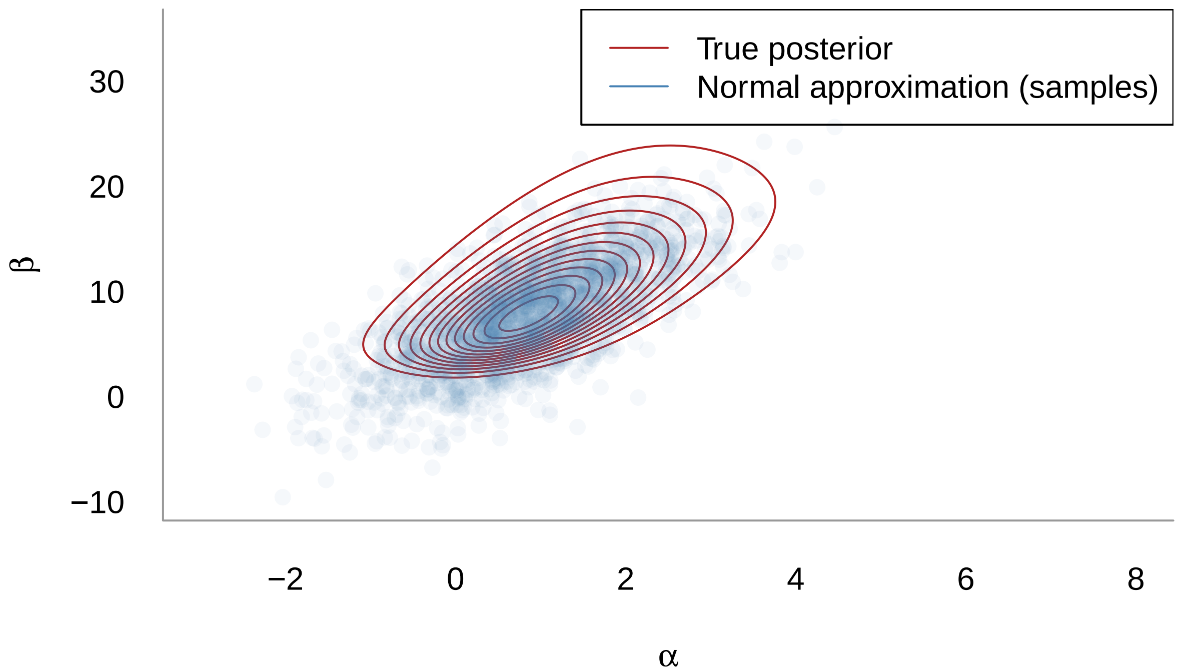
Therefore:

$$I(\hat{\theta}) = \begin{bmatrix} -\sum_{i=1}^{n} \frac{5e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} & -\sum_{i=1}^{n} \frac{5x_i e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} \\ -\sum_{i=1}^{n} \frac{5x_i e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} & -\sum_{i=1}^{n} \frac{5x_i^2 e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} \end{bmatrix}$$

To find the posterior mode, we can use numerical optimization:

```r
y <- c(0, 1, 3, 5)
x <- c(-0.86, -0.3, -0.05, 0.73)

llfun <- function(theta) {
  if(any(theta < 0)) return(Inf)
  a <- theta[1]; b <- theta[2]

  -sum(y * (a + b * x) - 5 * log(1 + exp(a + b * x)))

}

mode <- optim(c(1, 1), llfun)$par
```

```r
a <- mode[1]
b <- mode[2]
# Second derivaties evaluated at mode
plpa2 <- -sum( (5 * exp(a + b * x)) / (1 + exp(a + b * x))^2)
plpab <- -sum( (5 * x * exp(a + b * x)) / (1 + exp(a + b * x))^2)
plpb2 <- -sum( (5 * x^2 * exp(a + b * x)) / (1 + exp(a + b * x))^2)

I <- matrix(c(plpa2, plpab, plpab, plpb2), ncol = 2)
var_theta <- -solve(I)

# Draw samples from the approximate posterior
draws_approx <- MASS::mvrnorm(2e3, mode, var_theta)

# Compute the true posterior across a grid of values
anew <- seq(-3, 8, 0.025)
bnew <- seq(-10, 35, 0.1)

post_true <- matrix(nrow = length(anew),
                    ncol = length(bnew))

for (i in seq_along(anew)) {
  for (j in seq_along(bnew)) {
    ps <- exp(anew[i] + bnew[j] * x) / (1 + exp(anew[i] + bnew[j] * x))
    post_true[i, j] <- prod(dbinom(y, 5, ps))

  }
}

contour(anew, bnew, post_true, col = 'firebrick',
        axes = FALSE, drawlabels = FALSE,
        xlab = expression(alpha), ylab = expression(beta))
points(draws_approx[, 1], draws_approx[, 2],
     pch = 19, col = colt('steelblue', 0.05))
axis(1, tick = FALSE)
axis(2, tick = FALSE, las = 1)
box(bty = 'L', col = 'grey60')
legend('topright', col = c('firebrick', 'steelblue'),
       legend = c('True posterior', 'Normal approximation (samples)'),
       lwd = 1, border = 'grey60')
```



4.3 Delta method: Bioassay

From the previous question, we have:

```
# Posterior mode
mode
```

```
## [1] 0.846716 7.750138
```

```
# Posterior variance
var_theta
```

```
##           [,1]      [,2]
## [1,] 1.038716  3.547155
## [2,] 3.547155 23.753307
```

Now the function we want to approximate is LD50:

$$\text{LD50} = g(\alpha, \beta) = -\frac{\alpha}{\beta}$$