STA 114: Statistics

Notes 16. Normal means: Bayesian approach

The two means problem with equal variance

As in the previous lecture, we consider two samples $X = (X_1, \cdots, X_n)$, $Y = (Y_1, \cdots, Y_m)$ modeled as $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_1, \sigma^2)$, $Y_1, \cdots, Y_m \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_2, \sigma^2)$, X_i's and Y_j's are independent, with model parameters $\mu_1 \in (-\infty, \infty)$, $\mu_2 \in (-\infty, \infty)$, $\sigma^2 \in (0, \infty)$.

For a any one group, say the first group, we could assign its parameter (μ_1, σ^2) a normal-inverse-chi-square prior. However, we need to assign a prior on the triplet (μ_1, μ_2, σ^2). Here we consider a trivariate extension of the $\mathsf{N}\chi^{-2}$ pdf.

For $-\infty < m_1, m_2 < \infty$, $k_1, k_2, r, s > 0$, we denote by $\mathsf{NN}\chi^{-2}(m_1, k_1, m_2, k_2, r, s)$ the trivariate pdf:

$$f(w_1, w_2, v) = \text{const.} \times v^{-(r+4)/2} \exp\left\{-\frac{rs + k_1(w_1 - m_1)^2 + k_2(w_2 - m_2)^2}{2v}\right\}$$

defined over $w_1 \in (-\infty, \infty)$, $w_2 \in (-\infty, \infty)$ and $v > 0$. A couple of results would be handy.

Result 1. $(W_1, W_2, V) \sim \mathsf{NN}\chi^{-2}(m_1, k_1, m_2, k_2, r, s)$ if and only if

1. $\frac{rs}{V} \sim \chi^2(r)$
2. Conditionally on $V = v$ we have, $W_1 \sim \mathsf{Normal}(m_1, v/k_1)$, $W_2 \sim \mathsf{Normal}(m_2, v/k_2)$ and W_1 and W_2 are independent.

Result 2. If $(W_1, W_2, V) \sim \mathsf{NN}\chi^{-2}(m_1, k_1, m_2, k_2, r, s)$ then for any constants a_1, a_2, we must have $(a_1 W_1 + a_2 W_2, V) \sim \mathsf{N}\chi^{-2}(a_1 m_1 + a_2 m_2, (a_1^2/k_1 + a_2^2/k_2)^{-1}, r, s)$.

Some special cases of Result 2 are very useful:

- With $a_1 = 1$ and $a_2 = 0$ we get $(W_1, V) \sim \mathsf{N}\chi^{-2}(m_1, k_1, r, s)$. Similarly, $(W_2, V) \sim \mathsf{N}\chi^{-2}(m_2, k_2, r, s)$.
- With $a_1 = 1$ and $a_2 = -1$ we get $(W_1 - W_2, V) \sim \mathsf{N}\chi^{-2}(m_1 - m_2, (\frac{1}{k_1} + \frac{1}{k_2})^{-1}, r, s)$.

Conjugacy

The trivariate $\mathsf{NN}\chi^{-2}$ pdfs are particularly useful for our two groups normal model because they forma conjugate family. If we choose $\xi(\mu_1, \mu_2, \sigma^2) = \mathsf{NN}\chi^{-2}(m_1, k_1, m_2, k_2, r, s)$ then $\xi(\mu_2, \mu_2, \sigma^2|x, y) = \mathsf{NN}\chi^{-2}(m_1', k_1', m_2', k_2', r', s')$ where

$$\begin{aligned}
m_1' &= \frac{k_1 m_1 + n\bar{x}}{k_1 + n} \\
k_1' &= k_1 + n \\
m_2' &= \frac{k_2 m_2 + m\bar{y}}{k_2 + m} \\
k_2' &= k_2 + m \\
r' &= r + n + m \\
s' &= \frac{rs + (n-1)s_x^2 + (m-1)s_y^2 + \frac{k_1 n}{k_1+n}(\bar{x} - m_1)^2 + \frac{k_2 m}{k_2+m}(\bar{y} - m_2)^2}{r + n + m}.
\end{aligned}$$

This is fairly straightforward once we note

$$\begin{aligned}
\ell_{x,y}(\mu_1, \mu_2, \sigma^2) &= \text{const} - \frac{n+m}{2}\log\sigma^2 - \frac{(n-1)s_x^2 + (m-1)s_y^2 + n(\bar{x} - \mu_1)^2 + m(\bar{y} - \mu_2)^2}{2\sigma^2} \\
\log\xi(\mu_1, \mu_2, \sigma^2) &= \text{const} - \frac{r+4}{2}\log\sigma^2 - \frac{rs + k_1(\mu_1 - m_1)^2 + k_2(\mu_2 - m_2)^2}{2\sigma^2}
\end{aligned}$$

and so,

$$\begin{aligned}
\log\xi(\mu_1, \mu_2, \sigma^2) = \text{const} &- \frac{r+n+m+r}{2}\log\sigma^2 \\
&- \frac{rs + (n-1)s_x^2 + (m-1)s_y^2}{2\sigma^2} \\
&- \frac{n(\bar{x} - \mu_1)^2 + k_1(\mu_1 - m_1)^2}{2\sigma^2} \\
&- \frac{m(\bar{y} - \mu_2)^2 + k_2(\mu_2 - m_2)^2}{2\sigma^2}
\end{aligned}$$

and use our old identities (handout 10/07 on conjugate models)

$$n(\bar{x} - \mu_1)^2 + k_1(\mu_1 - m_1)^2 = (k_1 + n)\left(\mu_1 - \frac{k_1 m_1 + n\bar{x}}{k_1 + n}\right)^2 + \frac{k_1 n}{k_1 + n}(\bar{x} - m_1)^2$$

$$\text{and,}\quad m(\bar{y} - \mu_2)^2 + k_2(\mu_2 - m_2)^2 = (k_2 + m)\left(\mu_2 - \frac{k_2 m_2 + m\bar{y}}{k_2 + m}\right)^2 + \frac{k_2 m}{k_2 + m}(\bar{y} - m_2)^2$$

to recognize

$$\log\xi(\mu_1, \mu_2, \sigma^2) = \text{const} - \frac{r'+4}{2}\log\sigma^2 - \frac{r's' + k_1'(\mu_1 - m_1')^2 + k_2'(\mu_2 - m_2')^2}{2\sigma^2}$$

with $m_1', k_1', m_2', k_2', r', s'$ given as before.

Reference prior

Another convenient choice of the prior distribution is $\xi(\mu_1, \mu_2, \sigma^2) = \text{const}/\sigma^2$, the reference prior for this model. Under this prior, the posterior $\xi(\mu_1, \mu_2, \sigma^2)$ is indeed a $\mathsf{NN}\chi^{-2}(\bar{x}, n, \bar{y}, m, n+m-2, (n-1)s_x^2 + (m-1)s_y^2)$.

Inference on $\eta = \mu_1 - \mu_2$

With $\xi(\mu_1, \mu_2, \sigma^2 | x, y) = \mathsf{NN}\chi^{-2}(m_1', k_1', m_2', k_2', r', s')$, Result 2 gives us the joint posterior pdf of $\eta = \mu_1 - \mu_2$ and σ^2 to be $\mathsf{N}\chi^{-2}(m_1' - m_2', (1/k_1' + 1/k_2')^{-1}, r', s')$. And therefore under the posterior,

$$\frac{\eta - (m_1' - m_2')}{\sqrt{s'(1/k_1' + 1/k_2')}} \sim t(r').$$

This readily leads to $100(1-\alpha)\%$ central, posterior credible intervals for η of the form

$$(m_1' - m_2') \mp z_{r'}(\alpha)\sqrt{s'(1/k_1' + 1/k_2')}.$$

An interesting thing happens when we use the reference prior $\xi(\mu_1, \mu_2, \sigma^2) = \text{const}/\sigma^2$. The above interval then equals

$$(\bar{x} - \bar{y}) \mp z_{n+m-2}(\alpha)\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2)}}$$

which is same as the 95% ML confidence interval for η (see notes 10/26).

Prediction of $D^* = X^* - Y^*$

Now consider future variables $X^* = X_{n+1}$ and $Y^* = Y_{n+1}$ with our model suitably extended to: $X_1, \cdots, X_n, X_{n+1} \stackrel{\text{IID}}{\sim} \mathsf{Normal}(\mu_1, \sigma^2)$, $Y_1, \cdots, Y_m, Y_{m+1} \stackrel{\text{IID}}{\sim} \mathsf{Normal}(\mu_2, \sigma^2)$, X_i's and Y_j's are independent. Clearly, conditional on (μ_1, μ_2, σ^2), $D^* = X^* - Y^* \sim \mathsf{Normal}(\mu_1 - \mu_2, 2\sigma^2) = \mathsf{Normal}(\eta, 2\sigma^2)$. This coupled with the posterior pdf of (η, σ^2) as described above gives the posterior predictive distribution of D^* to be

$$\frac{D^* - (m_1' - m_2')}{\sqrt{s'(2 + 1/k_1' + 1/1k_2')}} \sim t(r')$$

(see Result 2 from handout 10/19 on Prediction).