Predictive probabilities for normal outcomes

John Cook

September 15, 2011

Suppose $Y \sim \text{normal}(\theta, \sigma^2)$ and *a priori* $\theta \sim \text{normal}(\mu, \tau)$. After observing $y_1, y_2, \ldots, y_n$ the posterior distribution on θ is normal with

$$\mathrm{E}(\theta \mid y_1 \ldots y_n) = \frac{\sigma^2/n}{\tau^2 + \sigma^2/n}\mu + \frac{\tau^2}{\tau^2 + \sigma^2/n}\bar{y}_{1:n}$$

and

$$\mathrm{Var}(\theta \mid y_1 \ldots y_n) = \frac{\tau^2\sigma^2/n}{\tau^2 + \sigma^2/n}$$

If we were to observe future data $y_{n+1}, y_{n+2}, \ldots, y_{n+m}$ the posterior distribution would again be normal with some mean α and some variance β^2.

We can write α as $\alpha_0 + \alpha_1\bar{y}_{n+1:n+m}$ where

$$\alpha_0 = \frac{\sigma^2/(n+m)}{\tau^2 + \sigma^2/(n+m)}\mu + \frac{\tau^2}{\tau^2 + \sigma^2/(n+m)}\frac{n\bar{y}_{1:n}}{n+m}$$

and

$$\alpha_1 = \frac{\tau^2}{\tau^2 + \sigma^2/(n+m)}\frac{m}{n+m}$$

Also,

$$\beta = \frac{\tau^2\sigma^2/(n+m)}{\tau^2 + \sigma^2/(n+m)}$$

Now suppose we're interested in whether a stopping rule will apply where a trial stops if

$$P(\theta > c) < d$$

where the distribution on θ is the posterior distribution after observing $y_1, \ldots, y_{n+m}$. Now

$$P(\theta > c) = 1 - P(\theta < c) = 1 - = \Phi\left(\frac{\theta - \alpha}{\beta}\right)$$

so the trial will stop if and only if

$$\Phi^{-1}(1-d) < (c - \alpha)/d$$

or equivalently

$$\bar{y}_{n+1:n+m} < k = \frac{c + \beta\Phi^{-1}(1-d) - \alpha_0}{\alpha_1}$$

Note that everything on the right side above is known. That is, the trial will stop if and only if the sample average of the future observations is below a cutoff that does not depend on the data.

Therefore the predictive probability of the trial stopping is the predictive probability that $\bar{y}_{n+1:n+m}$ is below a constant. The predictive distribution on $\bar{y}_{n+1:n+m}$ is normal, and so the entire calculation reduces to computing the probability of a normal random variable being below a cutoff.

The mean of the predictive distribution on $\bar{y}_{n+1:n+m}$ is the same as the posterior mean of θ after observing y_1 through y_n, i.e.

$$\mu_p = \frac{\sigma^2/n}{\tau^2 + \sigma^2/n}\mu + \frac{\tau^2}{\tau^2 + \sigma^2/n}\bar{y}_{1:n}.$$

The variance on the predictive distribution on $\bar{y}_{n+1:n+m}$ is σ^2/m plus the variance on θ given y_1 through y_n, i.e.

$$\sigma_p^2 = \frac{\sigma^2}{m} + \frac{\tau^2\sigma^2/n}{\tau^2 + \sigma^2/n}.$$

Putting everything together, the probability of the trial stopping after m future observations is $P(W < k)$ where W is normal with mean μ_p and variance σ_p^2. This equals or

$$\Phi^{-1}\left(\frac{c + \beta\Phi^{-1}(1-d) - \alpha_0 - \alpha_1\mu_p}{\alpha_1\sigma_p}\right).$$

The following R code implements the method described above.

```
# xbar = current sample mean
# n = current number of observations
# m = future observations
# Trial stops if P(\theta | data > c) < d
# mu = prior mean on theta
# tau = prior variance on theta
# sigma = sampling variance
normalpp <- function(xbar, n, m, c, d, mu, tau, sigma)
{
    temp <- (n+m)tau^2 + sigma^2
    alpha0 <- (mu*sigma^2 + m*xbar*tau^2) / temp
    alpha1 <- m*tau^2 / temp
    beta <- (tau*sigma)^2 / temp

    temp <- n*tau^2 + sigma^2
    mup <- mu*sigma^2 / temp + n*tau^2*xbar
    sigmap <- (sigma*tau)^2 / temp + sigma^2/m
    arg < (c + beta*qnorm(1-d) - alpha0 - alpha1*mup)/(alpha1*sigmap)
    return( qnorm(arg) )
}
```