CLARK UNIVERSITY 1887

A short introduction to Bayesian statistics, part IV
Math 217 Probability and Statistics

Prof. D. Joyce, Fall 2014

6 Normal distributions.

Bayesian statistics is fairly easy to use with normal distributions when the variance is known. In that case it turns out that the conjugate priors to use for the distribution $f(\mu)$ on the mean μ for normal distributions are normal distributions themselves.

It's harder to deal with the case when both the mean and and variance are unknown, but not insurmountable. We'll start out with the easier case.

Normal distributions with known variances. Let's assume that we have a normal population with an unknown mean μ and a known variance σ^2. Since σ^2 is known, it's just a constant, but μ is unknown so it has a prior density $f(\mu)$ and a posterior density $f(\mu|\mathbf{x})$.

If the prior distribution $f(\mu)$ on the mean μ is the normal distribution $\textsc{Normal}(\mu_0, \sigma_0^2)$, then the posterior distribution $f(\mu|\mathbf{x})$ based on the random sample $\mathbf{x} = (x_1, \ldots, x_n)$ turns out also to be a normal distribution. Let $\overline{x}$ be the sample mean and σ^2 the sample variance. We'll leave out the computations, but the posterior distribution is $\textsc{Normal}(\mu^*, \sigma^{*2})$ where

$$\mu^* = p_0\overline{x} + q_0\mu_0 \quad \text{and} \quad \sigma^{*2} = \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1},$$

and where

$$p_0 = \frac{\sigma_0^2}{\sigma^2/n + \sigma_0^2} \quad \text{and} \quad q_0 = 1 - p_0 = \frac{\sigma^2/n}{\sigma^2/n + \sigma_0^2}.$$

Thus, the family of all normal distributions is a conjugate family for μ.

Example. Let's take an example. Suppose we're monitoring the production line of cans of coffee. Suppose that we know from past experience that the mean content of cans is 16.0 oz., and the standard deviation is 0.1 oz. We can take that to say that the prior distribution $f(\mu)$ is $\textsc{Normal}(\mu_0, \sigma_0^2) = \textsc{Normal}(16.0, 0.1^2)$ and that σ^2 is also 0.1^2. If σ_0 is equal to σ, that simplifies the formulas above to $p_0 = \dfrac{n}{n+1}$ and $q_0 = \dfrac{1}{n+1}$. Therefore, $\mu^* = \dfrac{n\overline{x} + \mu_0}{n+1}$, and $\sigma^{*2} = \dfrac{\sigma^2}{n+1}$.

In the coffee can example, that becomes

$$\mu^* = \frac{n\overline{x} + 16.0}{n+1} \quad \text{and} \quad \sigma^{*2} = \frac{0.01}{n+1}.$$

Now let's suppose 9 cans are tested and the resulting sample mean is 16.1. Then

$$\mu^* = (9 \cdot 16.1 + 16.0)/10 = 16.09,$$

and $\sigma^{*2} = 0.001$, so $\sigma^* = 0.03$. Thus, the posterior distribution $f(\mu|\mathbf{x})$ is the normal distribution $\textsc{Normal}(16.09, 0.03^2)$.

We can use this posterior distribution to find probabilities of intervals. For instance, we can ask what the probability is that the mean μ is in the interval $[15.95, 16.05]$, and since we know the distribution of μ, we can answer the question. Indeed, since 15.95 is 4.67 standard deviations below the mean while 16.05 is 1.33 standard deviations below the mean, the probability that μ lies in $[15.95, 16.05]$ is 0.091. And note, this answer is an actual probability, not some confidence level, power, or whatever.

Reparameterization of the conjugate family. It's much easier to see what you need to do to go from the prior distribution to the posterior distribution if we change the parametrization of the family of normal distributions from μ and σ^2 into a different one.

Let's let

$$\alpha_0 = \frac{\sigma^2}{\sigma_0^2} \quad \text{and} \quad \beta_0 = \alpha_0\mu_0$$
$$\alpha^* = \frac{\sigma^2}{\sigma^{*2}} \quad \text{and} \quad \beta^* = \alpha^*\mu^*$$

so that the prior distribution for μ is $\text{Normal}(\mu_0, \sigma_0^2) = \text{Normal}(\frac{\beta_0}{\alpha_0}, \frac{\sigma^2}{\alpha_0})$ and the posterior distribution for μ is $\text{Normal}(\mu^*, \sigma^{*2}) = \text{Normal}(\frac{\beta^*}{\alpha^*}, \frac{\sigma^2}{\alpha^*})$. Then the equations above involving μ^*, μ_0, σ^{*2}, σ_0^2, p_0, and q_0, simplify to

$$\begin{aligned} \alpha^* &= \alpha_0 + n \\ \beta^* &= \beta_0 + \sum x_i \end{aligned}$$

Thus, α keeps track of the number of observations while β keeps track of their sum.

Naturally, a know-nothing prior would take $\alpha_0 = 0$ and $\beta_0 = 0$. Although this know-nothing prior doesn't actually describe a distribution, it will once one observation is made.

Math 217 Home Page at
`http://math.clarku.edu/~djoyce/ma217/`