ASQ Workshop on Bayesian Statistics for Industry

March 8, 2006

Prof. Stephen Vardeman

Statistics and IMSE Departments

Iowa State University

vardeman@iastate.edu
Module 2: A (Relatively) Simple Continuous Example of the “Bayes Paradigm”

The purpose of this module is to show precisely the operation of Bayes paradigm in a simple continuous context, where the calculus works out nicely enough to produce clean formulas. Suppose for sake of example that a variable $X$ is thought to be normal with standard deviation 1, but somehow the mean, $\theta$, is unknown. That is, we suppose that $X$ has a probability density function given by

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} (x - \theta)^2 \right)$$

and pictured in Figure 1.
Figure 1: The normal probability density with mean $\theta$ and standard deviation $\frac{1}{\sqrt{3}}$. 
This model assumption is commonly written in "short-hand" notation as

\[ X \sim N(\theta, 1) \]

Suppose that one wants to combine some "prior" belief about \( \theta \) with the observed value of \( X \) to arrive at an inference for the value of this parameter that reflects both (the "prior" and the data \( X = x \)). A mathematically convenient form for specifying such a belief is to assume that \textit{a priori}

\[ \theta \sim N(m, \gamma^2) \]

that is, that the parameter is itself normal with some mean, \( m \), and some variance \( \gamma^2 \), i.e. \( \theta \) has probability density

\[ g(\theta) = \frac{1}{\sqrt{2\pi\gamma^2}} \exp\left(-\frac{1}{2\gamma^2}(\theta - m)^2\right) \]

(An analyst is here specifying not only a form for the distribution of \( X \), but a complete description of his or her pre-data beliefs about \( \theta \) in terms of "normal," "roughly \( m \)," "with uncertainty characterized by a standard deviation \( \gamma \)."
Intermediate level probability calculations then imply that $X$ and $\theta$ have jointly continuous distribution with conditionals for $\theta | X = x$ (posterior densities for $\theta$ given $X = x$)

$$g(\theta|x) \propto f(x|\theta) g(\theta)$$

i.e.

posterior $\propto$ likelihood · prior

where one thinks of $x$ as fixed/observed and treats the expression above as a function of $\theta$. This is formally exactly as in the simple discrete example of Module 1. The posterior density is proportional to the product of the prior density and the likelihood (the density of $X$ treated as a function of the parameter). This is the continuous equivalent of making a table of products and dividing by "row totals" to get posterior distributions.

As it turns out, the posterior (the conditional distribution of $\theta | X = x$) is again normal. Take, for a concrete example, a case where the prior is $N(5, 2)$ and
one observes $X = 4$. It’s easy enough to use some engineering mathematics software to make the plots in Figure 2 of $f(4|\theta), g(\theta)$, and the product $f(4|\theta)g(\theta)$. The product is itself proportional to a normal probability density. Qualitatively, it seems like this normal density may have a mean somewhere between the locations of the peaks of of $f(4|\theta)$ and $g(\theta)$, and this graphic indicates how prior and likelihood combine to produce the posterior.
Figure 2: A normal likelihood (red) a normal prior (blue) and their product (green)
The exact result illustrated by Figure 2 is that the posterior in this model is again normal with

\[
\text{posterior mean} = \left( \frac{1}{1 + \frac{1}{\gamma^2}} \right) x + \left( \frac{1}{1 + \frac{1}{\gamma^2}} \right) m
\]

and

\[
\text{posterior variance} = \frac{1}{1 + \frac{1}{\gamma^2}}
\]

This is an intuitively appealing result in the following way. First, it is common in Bayes contexts to call the reciprocal of a variance a "precision." So in this model, the likelihood has precision

\[
\frac{1}{1} = 1
\]
while the "prior precision" is
\[ \frac{1}{\gamma^2} \]
The sum of these is
\[ 1 + \frac{1}{\gamma^2} \]
so that the posterior precision is the sum of the precision of the likelihood and the prior ... overall/posterior precision comes from both data and prior in an "additive" way. Then, the posterior mean is a weighted average of the observed value, \( x \), and the prior mean, \( m \), where the weights are proportional to the respective precisions of the likelihood and the prior.

These relationships between precisions of prior, likelihood and posterior, and how the first two "weight" the data and prior in production of a posterior are, in their exact form, special to this kind of "normal-normal" model. But
they indicate how Bayes analyses generally go. "Flat"/"uninformative"/large-variance/small-precision priors allow the data to drive an analysis. "Sharp"/
"peaked"/"informative"/small-variance/large-precision priors weight prior beliefs strongly and require strong sample evidence to move an analysis off of those beliefs.

To continue with the example

\[ X \sim N(\theta, 1) \text{ and } \theta \sim N(m, \gamma^2) \]

there remains the matter of a predictive posterior for \( X_{\text{new}} \) (e.g. from the same distribution as \( X \)). If given \( \theta \) it make sense to model \( X \) and \( X_{\text{new}} \) as independent draws from the same \( N(\theta, 1) \) distribution, it’s easy to figure out an appropriate predictive posterior for \( X_{\text{new}} \) given \( X = x \). That is, we know what the posterior of \( \theta \) is, and \( X_{\text{new}} \) is just "\( \theta \) plus \( N(0, 1) \) noise." That is,
the predictive posterior is

$$X_{\text{new}} | X = x \sim \mathcal{N} \left( \left( \frac{1}{1 + \frac{1}{\gamma^2}} \right) x + \left( \frac{\frac{1}{\gamma^2}}{1 + \frac{1}{\gamma^2}} \right) m, 1 + \frac{1}{1 + \frac{1}{\gamma^2}} \right)$$

i.e. it is again normal with the same posterior mean as $\theta$, but a variance increased from the posterior variance of $\theta$ by 1 (the variance of an individual observation). This is again qualitatively right. One can not know more about $X_{\text{new}}$ than one knows about $\theta$.

To make this all concrete, consider once again a case where the prior is $\mathcal{N}(5, 2)$ and one observes $X = 4$. Figure 3 shows the likelihood, and all of the prior, posterior, and posterior predictive distributions.
Figure 3: A normal likelihood if $X = 4$ (red), a prior density for $\theta$ (blue), the corresponding posterior for $\theta$ (green), and the corresponding predictive posterior for $X_{\text{new}}$ (violet)