

Module 2

A (relatively) simple continuous example of the “Bayes paradigm”

Prof. Stephen B. Vardeman
Statistics and IMSE
Iowa State University

March 5, 2008

A Fairly Simple Continuous Example (Likelihood/Data Model)

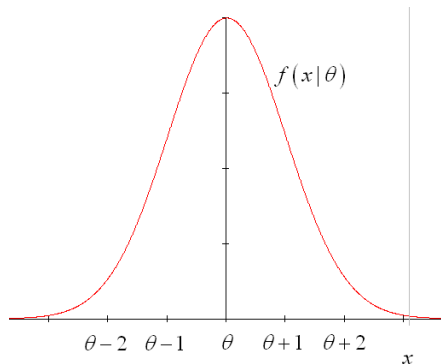
The purpose of this module is to show precisely the operation of Bayes paradigm in a simple continuous context, where the calculus works out nicely enough to produce clean formulas. Suppose that a variable X is thought to be normal with standard deviation 1, but somehow the mean, θ , is unknown. That 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.

A Fairly Simple Continuous Example (Likelihood cont.)

Figure 1 The normal probability density with mean θ and standard deviation 1



This model assumption is commonly written in "short-hand" notation as $X \sim N(\theta, 1)$

A Fairly Simple Continuous Example (Prior)

Suppose that one wants to combine some "prior" belief about θ 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 *a priori*

$$\theta \sim N(m, \gamma^2)$$

that is, that the parameter is itself normal with some mean, m , and some variance γ^2 , i.e. θ 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 θ in terms of "normal," "roughly m ," "with uncertainty characterized by a standard deviation γ .")

A Fairly Simple Continuous Example (Posterior)

Intermediate level probability calculations then imply that X and θ have jointly continuous distribution with conditionals for $\theta|X = x$ (posterior densities for θ given $X = x$)

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

i.e.

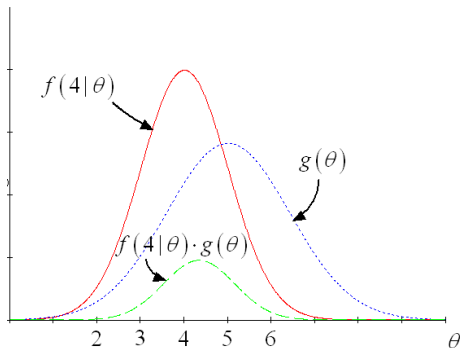
$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}$$

where one thinks of x as fixed/observed and treats the expression above as a function of θ . 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.

A Fairly Simple Continuous Example (Posterior)

Take, for a concrete example, a case where the prior is $N(5, 2)$ and one observes $X = 4$. It's easy to make the plots in Figure 2 of $f(4|\theta)$, $g(\theta)$, and the product $f(4|\theta)g(\theta)$.

Figure 2 A normal likelihood (red) a normal prior (blue) and their product (green)



A Fairly Simple Continuous Example (Posterior)

The product $f(4|\theta)g(\theta)$ is 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 $f(4|\theta)$ and $g(\theta)$, and this graphic indicates how prior and likelihood combine to produce the posterior.

The exact general result illustrated by Figure 2 is that the posterior in this model is again normal with

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

and

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

A Fairly Simple Continuous Example (Posterior)

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 precisions of the likelihood and the prior (overall/posterior precision comes from both data and prior in an "additive" way). Further, 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.

The Qualitative Nature of Bayes Analyses

The 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.

A Fairly Simple Continuous Example (Prediction)

To continue with the example

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

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

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

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

A Fairly Simple Continuous Example (Summary)

To make this all concrete, consider once again a case where the prior is $N(5, 2)$ and one observes $X = 4$. Figure 3 summarizes this example.

Figure 3 A normal likelihood if $X = 4$ (red), a prior density for θ (blue), the corresponding posterior for θ (green), and the corresponding predictive distribution for X_{new} (violet)

