Recitation Note: Calculating Bayes Posteriors

Tyler Maunu
maunut@mit.edu

September 3, 2019

This note will discuss the calculation of posterior distributions in the context of Bayes estimation. First, we will quickly review the set-up for the Bayesian way of thinking about estimation. Then we will go through a few examples to show how to apply this reasoning.

Introduction

As a review, estimation in Bayesian frame utilizes Bayes' theorem, which we'll state here for reference.

Theorem 1 (Bayes')**.** *Suppose that A and B are events on a probability space. Then,*

$$P(A|B)P(B) = P(B|A)P(A). \tag{1}$$

The more familiar form when $P(B) \neq 0$ is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{2}$$

In the following problems, we assume that we are given some data $X_1, \dots, X_n$, and that these data follow a distribution that depends on a parameter θ. In the Bayesian framework, we think of θ as a random variable as well. In this case, we write the following:

$$X_1, \dots, X_n | \theta \sim P_\theta, \ \theta \sim \pi \tag{3}$$

Here, P_θ is the condition distribution of X given θ, and π is the distribution of θ. The posterior distribution is $\pi(\theta | X_1, \dots, X_n)$, and captures our uncertainty in θ after we observe a bunch of data. We use Bayes theorem to write

$$\pi(\theta | X_1, \dots, X_n) = \frac{P(X_1, \dots, X_n | \theta)\pi(\theta)}{P(X_1, \dots, X_n)}. \tag{4}$$

Notice that $P(X_1, \dots, X_n | \theta)$ is the likelihood of our sample. Further, $P(X_1, \dots, X_n)$ does not depend on θ. In determining the posterior distribution, we thus usually write

$$\pi(\theta | X_1, \dots, X_n) \propto L(X_1, \dots, X_n | \theta)\pi(\theta), \tag{5}$$

and identify the distribution by the functional form of $L \cdot \pi$. This is because of the general property that only one probability distribution is proportional to a functional form. In other words, if I define a function $g(x)$ such that

$$g(x) \geq 0, \int_{-\infty}^{\infty} g(x)dx < \infty, \tag{6}$$

then there is a unique probability density function $f(x)$ that is proportional to g. This $f(x)$ is defined by

$$f(x) = g(x) \cdot \frac{1}{\int_{-\infty}^{\infty} g(x)dx}. \tag{7}$$

Question

Find the posterior distribution and Bayes estimator for the following observed and prior distributions.

1. $X_1, \ldots, X_n | p \overset{i.i.d.}{\sim} \textsf{Geometric}(p)$, $p \sim \textsf{Beta}(a, b)$
2. $X_1, \ldots, X_n | \theta \overset{i.i.d.}{\sim} \textsf{Pareto}(\theta)$, $\theta \sim \textsf{Gamma}(a, b)$
3. $X_1, \ldots, X_n | v \overset{i.i.d.}{\sim} N(0, v)$, $v \sim \textsf{Inverse Gamma}(a, b)$

Solution

1. In this case, the pdf of X_i is given by

$$f(x|p)(1-p)^{x-1}p, \ x = 1, 2, \ldots. \tag{8}$$

and the prior is given by

$$\pi(p) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}, \ p \in [0, 1], \tag{9}$$

where $B(a, b)$ is the beta function, for which we do not have an analytical expression in general (just think of it as our normalizing constant). Thus, using our "likelihood times prior" procedure, we find the following functional form for our posterior:

$$\pi(p|X_1 = x_1, \ldots, X_n = x_n) \propto L(X_1 = x_1, \ldots, X_n = x_n|p)\pi(p) \tag{10}$$

$$\left(\prod_{i=1}^{n}(1-p)^{x_i-1}p\right) \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}$$

$$\propto (1-p)^{\sum_{i=1}^{n} x_i - n + b - 1} p^{n+a-1}$$

Notice that, as we simplify our expression, we drop any constant terms that do not depend on p. The game is to now identify what distribution this is. Notice that the form is $(1 - p)$ to some power times p to some power, for $p \in [0, 1]$. But this is precisely the functional form of the $\textsf{Beta}$ distribution given in (9). In fact, we recognize this to be a $\textsf{Beta}(n+a, \sum_{i=1}^{n} x_i - n + b)$ distribution.

For our Bayes estimator, we use the posterior mean (although other choices of estimator can be motivated, but that is a separate topic – after you find the posterior distribution, what you do with it is up to you!). Now, either calculating the expectation of a Beta distribution ourselves, or looking it up in our favorite source (like Wikipedia), we find that our Bayes estimator is

$$\hat{p}_{Bayes} = \frac{n+a}{n+a+\sum_{i=1}^n x_i - n} = \frac{n+a}{a+b+\sum_{i=1}^n x_i}. \tag{11}$$

2. We repeat our procedure used in the previous problem. The pdf of X_i is given by

$$f(x|\theta) = \theta x^{-\theta-1},\ x \geq 1, \tag{12}$$

and the pdf of θ is give by

$$\pi(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta},\ \theta > 0. \tag{13}$$

The posterior is then

$$\begin{aligned}\pi(\theta|X_1 = x_1, \ldots, X_n = x_n) &\propto L(X_1 = x_1, \ldots, X_n = x_n|\theta)\pi(\theta) \\ &= \left(\prod_{i=1}^n \theta x_i^{-\theta-1}\right)\frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta} \\ &\propto \theta^{n+a-1}e^{-b\theta}\left(\prod_{i=1}^n x_i\right)^{-\theta-1} \\ &= \theta^{n+a-1}e^{-b\theta}e^{-(\theta+1)\left(\sum_{i=1}^n \log(x_i)\right)} \\ &\propto \theta^{n+a-1}e^{-\theta\left(b+\sum_{i=1}^n \log(x_i)\right)}.\end{aligned} \tag{14}$$

We recognize this as the functional form of a $\mathsf{Gamma}(n+a, b+\sum_{i=1}^n \log(x_i))$ distribution.

Looking up the mean of a Gamma distribution, we find our Bayes estimator:

$$\hat{\theta}_{Bayes} = \frac{n+a}{b+\sum_{i=1}^n \log(x_i)}. \tag{15}$$

3. Finally, the familiar pdf of X_i is given by

$$f(x|v) = \frac{1}{\sqrt{2\pi v}}e^{-\frac{x^2}{2v}},\ x \in, \tag{16}$$

and the pdf of v is given by

$$\pi(v) = \frac{b^a}{\Gamma(a)}v^{-a-1}e^{-b/v},\ v > 0. \tag{17}$$

The posterior is then (skipping a couple steps this time!)

$$\begin{aligned}\pi(v|X_1 = x_1, \ldots, X_n = x_n) &\propto L(X_1 = x_1, \ldots, X_n = x_n|v)\pi(v) \\ &= \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi v}}e^{-\frac{x_i^2}{2v}}\right)\frac{b^a}{\Gamma(a)}v^{-a-1}e^{-b/v} \\ &\propto v^{-\frac{n}{2}-a-1}e^{-\frac{1}{v}\left(b+\frac{1}{2}\sum_{i=1}^n x_i^2\right)}\end{aligned} \tag{18}$$

This is the functional form of an $\mathsf{Inverse\ Gamma}(n/2 + a, b + (\sum_i x_i^2/2))$ distribution. Our Bayes estimator is

$$\hat{v}_{Bayes} = \frac{b + (\sum_i x_i^2)/2}{n/2 + a}. \tag{19}$$

Notice that, if $a = b = 0$, then we arrive back at the usually maximum likelihood estimate of the variance.

Discussion

All of the examples above are a bit special in that the posterior distributions have nice forms. This will not always be the case in practice. When the posterior distribution is the same as the prior distribution, then we call the prior a *conjugate prior* for the given data distribution. All of the above are examples of conjugate priors.