

Template

December 5, 2021

In this video I will present 4 examples on deriving the posterior distribution given the prior and the observed outcome. Keep in mind that the main idea in Bayesian modelling is the notion of updating what we know about the distribution of the model parameters after the data becomes available. In other words, we want to study the distribution of the parameters conditional on the observed data.

Example 1

Suppose that we have n observations x_1 up to x_n and we model them as draws from an exponential distribution with rate parameter λ . The only complication is that we want to model the situation in a way that allows for the scale parameter λ itself to be random. For this example, that us suppose that our prior belief is λ being a random variable from a gamma distribution with shape parameter 7 and rate parameter 5.

$$\begin{aligned} X_i | \lambda &\stackrel{iid}{\sim} \text{Exp}(\lambda) \quad \text{for } i = 1, \dots, n \\ \lambda &\sim \text{Gamma}(5, 7) \end{aligned}$$

The first step is to work out the likelihood. Since we are assuming that each of the iid observations comes from an exponential distribution whenever λ is given, the likelihood denoted by L is given by this product.

$$L(\mathbf{x} | \lambda) = \prod_{i=1}^n f(x_i | \lambda) = \lambda^n \exp \left(-\lambda \sum_{i=1}^n x_i \right)$$

Next the density of the prior distribution looks like this

$$p(\lambda) = \frac{5^7}{\Gamma(7)} \lambda^{7-1} \exp(-5\lambda)$$

The last step is to work out what the posterior distribution of the parameter which is

$$\text{density}(\text{parameter} | \text{data}).$$

This formalizes the idea of how your knowledge of the model parameter have changed or reinforced after seeing the observed data.

In this example it is the density of λ given the vector of observations \mathbf{x} . I usually like to call my posterior π .

$$\text{density}(\text{parameter} | \text{data}) = \pi(\lambda | \mathbf{x})$$

To work that out we need to apply the Bayes theorem, which is

$$\pi(\lambda|\mathbf{x}) = \frac{L(\mathbf{x}|\lambda)p(\lambda)}{\int L(\mathbf{x}|\lambda)p(\lambda) d\lambda}$$

The integral on the bottom is normalizing constant which ensures the whole expression integrates to 1. For the purpose of today, we are not interested in this integral, so it turns out that is actually enough to say that posterior density is proportional to the numerator expression.

$$\pi(\lambda|\mathbf{x}) = \frac{L(\mathbf{x}|\lambda)p(\lambda)}{\int L(\mathbf{x}|\lambda)p(\lambda) d\lambda} \propto L(\mathbf{x}|\lambda)p(\lambda)$$

So plugging what we have we get the following. A trick that I used here is to ignore any multiplicative constants unrelated to the parameter λ . This is ok because we studying the posterior up to some proportionality constant. Finally notice that we end up with an expression that is equal to a gamma distribution. Sometimes we are lucky that the posterior density is some well-studied one. Sometimes we are not so lucky, then we need to put in some more effort.

That aside, knowing what the posterior density is, we can use it to form predictions. One simple way to do so is to simulate a lot of λ from this distribution, then for each λ we simulate a corresponding x , from there we can make some predictions about x for example by calculating the average.

$$\begin{aligned} \pi(\lambda|\mathbf{x}) &\propto L(\mathbf{x}|\lambda)p(\lambda) \\ &= \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) \times \frac{5^7}{\Gamma(7)} \lambda^{7-1} \exp(-5\lambda) \\ &\propto \lambda^{7+n-1} \exp\left(-\left[5 + \sum_{i=1}^n x_i\right] \lambda\right) \\ \lambda|\mathbf{x} &\sim \text{Gamma}(7+n, 5 + \sum_{i=1}^n x_i) \end{aligned}$$

In other words, a simple way to make predictions on x is

- for k in $\{1, 2, \dots, m\}$:
- simulate $\lambda_k \sim \text{Gamma}(7+n, 5 + \sum_{i=1}^n x_i)$
- simulate $x_k \sim \text{Exp}(\lambda_k)$
- calculate average of x_1, \dots, x_m

So in summary, once you have decided on your model, there are three steps to follow

1. write down the likelihood conditional on the parameter
2. write down the prior
3. work out the posterior using the Bayes' theorem

Example 2

Let us try example. Suppose x_1 up to x_n are iid normal with mean μ and a known variance of $1/2$. Further the prior belief for the mean is exponential with rate 2. Find the posterior density for μ

We have

$$\begin{aligned} x_i &\stackrel{iid}{\sim} \mathcal{N}(\mu, 1/2) \\ \mu &\sim \text{Exp}(2) \end{aligned}$$

so that

$$\begin{aligned} L(\mathbf{x}|\mu) &= \prod_{i=1}^n f(x_i|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{\pi}} \exp[-(x_i - \mu)^2] \\ &= \pi^{-n/2} \exp\left(-\sum_{i=1}^n (x_i - \mu)^2\right) \\ &= \pi^{-n/2} \exp\left(-n\mu^2 + 2\mu \sum_{i=1}^n x_i + \sum_{i=1}^n x_i^2\right) \\ &= \pi^{-n/2} \exp\left(-n\mu^2 + 2n\bar{x}\mu + \sum_{i=1}^n x_i^2\right) \\ \therefore p(\mu) &= 2 \exp(-2\mu) \mathbb{I}\{\mu > 0\} \\ \therefore \pi(\mu|\mathbf{x}) &\propto (2\pi)^{-n/2} \exp\left(-n\mu^2 + 2n\bar{x}\mu + \sum_{i=1}^n x_i^2\right) 2 \exp(-2\mu) \mathbb{I}\{\mu > 0\} \\ &\propto \exp\left(-n\mu^2 + 2n(\bar{x} - 1/n)\mu\right) \mathbb{I}\{\mu > 0\} \\ &= \exp\left(-n\mu^2 + 2n(\bar{x} - 1/n)\mu - n(\bar{x} - 1)^2 + n(\bar{x} - 1)^2\right) \mathbb{I}\{\mu > 0\} \\ &\propto \exp(-n[\mu - (\bar{x} - 1/n)]^2) \mathbb{I}\{\mu > 0\} \\ \therefore \mu|\mathbf{x} &\sim \mathcal{N}(\bar{x} - 1/n, 1/(2n)) \mathbb{I}\{\mu > 0\}. \end{aligned}$$

So it is a normal with mean $\bar{x} - 1$, variance $1/2n$ constrained to $\mu > 0$.

Example 3

In the last two examples, we will look at some regression type model.

Let us suppose that the pairs (x_1, y_1) up to (x_n, y_n) are observed and that given β_0, β_1 and σ^2 we have. One thing to note is that since we are working with regression type model, everything will turn out to be conditional on x_i .

$$\begin{aligned} y_i|\beta_0, \beta_1, \sigma^2, x_i &\stackrel{iid}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2) \quad \text{for } i = 1, \dots, n \\ \beta_j|\sigma^2 &\stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2) \quad \text{for } j = 0, 1 \\ \sigma^2 &\sim \text{InvGamma}(5, 7) \end{aligned}$$

One thing to note is that since we are working with regression type model, everything will turn out to be conditional on x_i .

Following the three steps

$$\begin{aligned}
L(\mathbf{y}|\beta_0, \beta_1, \sigma^2, \mathbf{x}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \\
&= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \\
&= (2\pi\sigma^2)^{n/2} \exp\left(-\frac{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \\
\therefore p(\beta_0, \beta_1|\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\beta_0^2}{2\sigma^2}\right) \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\beta_1^2}{2\sigma^2}\right) \\
&= (2\pi\sigma^2)^{-1} \exp\left(-\frac{\beta_0^2 + \beta_1^2}{2\sigma^2}\right) \\
p(\sigma^2) &= \frac{7^5}{\Gamma(5)} (\sigma^2)^{-5-1} \exp\left(-\frac{7}{\sigma^2}\right) \\
\therefore \pi(\beta_0, \beta_1, \sigma^2|\mathbf{y}, \mathbf{x}) &\propto (2\pi\sigma^2)^{n/2} \exp\left(-\frac{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \\
&\quad \times (2\pi\sigma^2)^{-1} \exp\left(-\frac{\beta_0^2 + \beta_1^2}{2\sigma^2}\right) \\
&\quad \times \frac{7^5}{\Gamma(5)} (\sigma^2)^{-5-1} \exp\left(-\frac{7}{\sigma^2}\right)
\end{aligned}$$

We can see already that this joint distribution is a beast. Fortunately there are techniques like Markov chain Monte Carlo which allows us to approximate simulate draws from this kind of beast. So with a bit of coding, we can still perform inference from it.

Example 4

Our last example takes the following form.

$$\begin{aligned}
y_i|\beta_0, \beta_1, \sigma, x_i &\stackrel{iid}{\sim} \mathbf{N}(\beta_0 + \beta_1 x_i, \sigma^2) \quad \text{for } i = 1, \dots, n \\
\beta_j &\stackrel{iid}{\sim} \mathbf{N}(0, 20^2) \quad \text{for } j = 0, 1 \\
\sigma &\sim \text{HalfCauchy}(0, 10)
\end{aligned}$$

so that

$$\begin{aligned}
p(\beta_0, \beta_1) &= \frac{1}{\sqrt{2\pi \times 20^2}} \exp\left(-\frac{\beta_0^2}{2 \times 20^2}\right) \times \frac{1}{\sqrt{2\pi \times 20^2}} \exp\left(-\frac{\beta_1^2}{2 \times 20^2}\right) \\
p(\sigma) &= \frac{2}{\pi \times 10} \times \frac{1}{1 + \sigma^2/10^2}
\end{aligned}$$

Notice that unlike the previous example where we specify a prior for σ^2 , this time we specify a prior to σ .

Following the three steps

$$\begin{aligned}
L(\mathbf{y}|\beta_0, \beta_1, \sigma^2, \mathbf{x}) &= (2\pi\sigma^2)^{n/2} \exp\left(-\frac{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \\
p(\beta_0, \beta_1) &= \frac{1}{\sqrt{2\pi} \times 20^2} \exp\left(-\frac{\beta_0^2}{2 \times 20^2}\right) \times \frac{1}{\sqrt{2\pi} \times 20^2} \exp\left(-\frac{\beta_1^2}{2 \times 20^2}\right) \\
p(\sigma) &= \frac{2}{\pi \times 10} \times \frac{1}{1 + \sigma^2/10^2} \\
\therefore \pi(\beta_0, \beta_1, \sigma|\mathbf{y}, \mathbf{x}) &= (2\pi\sigma^2)^{n/2} \exp\left(-\frac{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \\
&\quad \times \frac{1}{\sqrt{2\pi} \times 20^2} \exp\left(-\frac{\beta_0^2}{2 \times 20^2}\right) \times \frac{1}{\sqrt{2\pi} \times 20^2} \exp\left(-\frac{\beta_1^2}{2 \times 20^2}\right) \\
&\quad \times \frac{2}{\pi \times 10} \times \frac{1}{1 + \sigma^2/10^2}
\end{aligned}$$