4. Bayesian Inference

4.1 The Bayesian approach to inference

The Bayesian approach to inference assumes that all forms of uncertainty (including parameter values) are expressed in terms of probability. A Bayesian model is a representation of all random quantities by prior distributions which are meant to capture our beliefs about the random experiment before seeing any data. After observing some data, we apply Bayes' Rule to obtain the posterior distribution for the unobserved quantities, which takes incorporates both the prior distribution and the data. Using the posterior distribution we can perform inference.

Suppose a random experiment will result in data $\mathbf{X} = (X_1, \ldots, X_n)$ where $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x|\theta)$ and $\theta \in \Theta$ is an unknown parameter. In the Bayesian approach θ is considered to be a random quantity whose randomness can be described by a probability distribution we call the **prior distribution**, denoted by $\pi(\theta)$. The prior distribution is chosen before we observe any data using only information about θ. Suppose next that we observed data $\mathbf{x} = (x_1, \ldots, x_n)$. These data can be used to update our belief about the value of θ, which we call the **posterior distribution**, denoted by $\pi(\theta|\mathbf{x})$, using Bayes' rule, i.e.,

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_\Theta f(\mathbf{x}|\theta)\pi(\theta)\,\mathrm{d}\theta}, \tag{4.1}$$

thus the posterior distribution is the conditional distribution of the parameter θ given the sample $\mathbf{X} = \mathbf{x}$.

Because $X_1, \ldots, X_n$ are i.i.d, their joint pdf/pmf is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta).$$

Using the posterior distribution for inference is straightforward:

- For a point estimate, we use the expected value of θ under the posterior distribution, i.e.,

$$\tilde{\theta} = \int_\Theta \theta\, \pi(\theta|\mathbf{x})\,\mathrm{d}\theta.$$

- For a confidence interval of level $1-\alpha$ we take the $\alpha/2$ and $1-\alpha/2$ left quantiles of the posterior distribution, i.e., if L and U are such that

$$\int_{-\infty}^{L} \pi(\theta|\mathbf{x})\,\mathrm{d}\theta = \alpha/2 \quad \text{and} \quad \int_{-\infty}^{U} \pi(\theta|\mathbf{x})\,\mathrm{d}\theta = 1-\alpha/2,$$

  then the corresponding confidence interval is $[L, U]$.

Example 4.1. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$, $\theta \in [0,1]$. Then $f(x|\theta) = \theta^x(1-\theta)^{1-x}$ so

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}.$$

Suppose we choose a prior distribution as $\theta \sim \text{U}(0,1)$, i.e., a uniform distribution in the interval $(0,1)$. Then $\pi(\theta) = 1$ for $\theta \in (0,1)$ and

$$f(\mathbf{x}|\theta)\pi(\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}.$$

Here we introduce a new function: the integral

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x,$$

is called the **beta function** for $\alpha, \beta > 0$. This function is available in Python in `scipy.special.beta`. We can see that

$$\int_0^1 f(\mathbf{x}|\theta)\pi(\theta)\,\mathrm{d}\theta = \int_0^1 \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\,\mathrm{d}\theta = B\left(\sum x_i + 1, n - \sum x_i + 1\right),$$

so the posterior distribution from (4.1) is

$$\pi(\theta|\mathbf{x}) = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{B\left(\sum x_i + 1, n - \sum x_i + 1\right)}.$$

This pdf can be recognised as a form of the beta distribution defined below.

Definition 4.1 (Beta distribution)**.**
The random variable with pdf

$$f(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \ x \in (0,1), \ \alpha,\beta > 0,$$

is said to have the beta distribution with parameters α and β, written $X \sim \text{Beta}(\alpha,\beta)$. The mean of this distribution is $\mathsf{E}\,x = \frac{\alpha}{\alpha+\beta}$.

The beta distribution is available in Python in `scipy.stats.beta`.

This suggests that the posterior distribution for θ is the beta distribution with parameters $\alpha = \sum x_i + 1$ and $\beta = n - \sum x_i + 1$.

Suppose that with $n = 12$ trials, we observe 4 successes, i.e., $\sum x_i = 4$. Then $\alpha = 5$ and $\beta = 9$, so the posterior distribution for θ becomes $\theta \sim \text{Beta}(5,9)$. The prior and posterior distributions in this case are plotted in Figure 4.1. It can be seen that as the prior distribution has no preference for a particular value of θ, the posterior distribution suggests that a more likely value for θ is in the range 0.2 to 0.5.

We can derive a point estimate for θ by calculating the posterior mean. Using the formula for the mean of the beta distribution, we have $\tilde{\theta} = \frac{5}{5+9} = 0.357$. Comparing this with the MLE, $\hat{\theta} = \frac{4}{12} = 0.333$, we see that the two estimates are similar but the Bayesian estimate $\tilde{\theta}$ is closer to 0.5. This is because of the influence from the prior which had mean 0.5.

For a confidence interval for θ, let $\beta_{\alpha/2}$ and $\beta_{1-\alpha/2}$ be the $\alpha/2$ and $1-\alpha/2$ left quantiles of the $\text{Beta}(5,9)$ distribution which can be obtained using Python, e.g. for $\alpha = 5\%$ we have

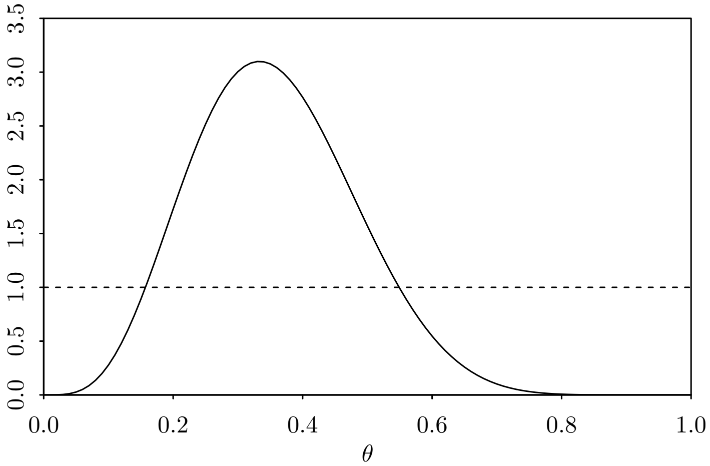

Figure 4.1: Prior pdf (dashed line) and posterior pdf (solid line) for Example 4.1.

```
>>> import scipy.stats
>>> scipy.stats.beta.ppf([.025,.975],5,9)
array([ 0.13857934,  0.61426166])
```

so $\beta_{\alpha/2} = 0.139$ and $\beta_{1-\alpha/2} = 0.614$. In general, the level $1-\alpha$ confidence interval is $[\beta_{\alpha/2}, \beta_{1-\alpha/2}]$. ▶

Example 4.2. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\theta, \sigma^2)$ where σ^2 has a known value and θ is unknown and assume a prior $\theta \sim \mathrm{N}(\mu, \tau^2)$. Then,

$$
\begin{aligned}
f(\mathbf{x}|\theta) &= \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(x_i - \theta)^2 \right\} \right] \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} x_i^2 - 2\theta \sum_{i=1}^{n} x_i + n\theta^2 \right) \right\}. \\
\pi(\theta) &= (2\pi\tau^2)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\tau^2}(\theta - \mu)^2 \right\} \\
&= (2\pi\tau^2)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\tau^2}(\theta^2 - 2\theta\mu + \mu^2) \right\}.
\end{aligned}
$$

Multiplying the two, we can see that the posterior density will have the form

$$
\pi(\theta|\mathbf{x}) = \exp\left\{ -\frac{1}{2}(A\theta^2 - 2B\theta + C) \right\}
$$

where

$$
A = \frac{n}{\sigma^2} + \frac{1}{\tau^2}, \qquad\qquad B = \frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2},
$$

and we don't need to know the value of C which does not depend on θ. Next, we perform an exercise called "completing the square" whereby we write the term in the brackets as

$$
A\theta^2 - 2B\theta + C = A\left(\theta - \frac{B}{A}\right)^2 + C',
$$

for some other C' which does not depend on θ. This form can be recognised as the pdf of the normal distribution with mean $\lambda = B/A$ and variance $\omega^2 = 1/A$ where we find

$$\omega^2 = \frac{1}{A} = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}, \qquad \lambda = \frac{B}{A} = \frac{\mu\sigma^2 + n\tau^2\bar{x}}{\sigma^2 + n\tau^2}.$$

Suppose that in a sample of $n = 5$ observations we find $\bar{x} = 4.2$ with $\sigma^2 = 1$, and we assume $\theta \sim \mathrm{N}(0, 5)$, i.e., $\mu = 0$ and $\tau^2 = 5$. The we can calculate the posterior mean to be $\lambda = \frac{0\times1+5\times5\times(4.2)}{1+5\times5} = 4.038$ and $\omega^2 = \frac{1\times5}{1+5\times5} = 0.192$. A plot of the prior and posterior distributions for θ are shown in Figure 4.2.

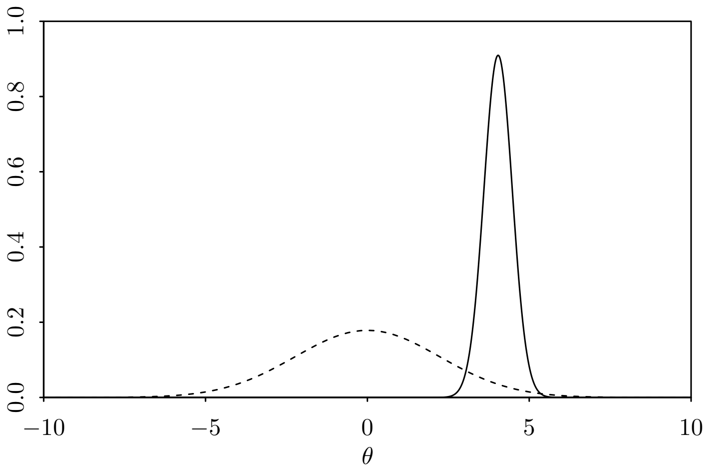

Figure 4.2: Prior pdf (dashed line) and posterior pdf (solid line) for Example 4.2.

A 95% confidence interval for θ can be obtained by selecting the appropriate quantiles of the posterior distribution. If we let $Z = \frac{\theta-\lambda}{\omega}$, then $Z \sim \mathrm{N}(0, 1)$ and $z_{0.975} = 1.96$, $z_{0.025} = -1.96$. Therefore the confidence interval for θ is $L = \lambda + z_{0.025} \times \omega = 4.038 - (1.96)\sqrt{0.192} = 2.741$ and $U = \lambda + z_{0.975} \times \omega = 4.038 + (1.96)\sqrt{0.192} = 5.336$, so $[2.741, 5.336]$. ▶

4.2 Exercises

1. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Bernoulli}(\theta)$, $\theta \in [0, 1]$ with prior distribution $\theta \sim \mathrm{Beta}(\alpha, \beta)$ for given values of α and β.

   a) Derive the posterior distribution for θ in this case.

   b) Suppose we choose $\alpha = 2$, $\beta = 4$ and, after $n = 10$ trials, we observe 7 successes. Derive the posterior estimate for θ and use Python to get a 95% confidence interval.

2. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Poisson}(\theta)$, $\theta > 0$, with prior distribution $\theta \sim \mathrm{Gamma}(\alpha, \beta)$ for given values of α (shape parameter) and β (rate parameter).

   a) Derive the posterior distribution for θ in this case.

   b) Suppose we choose $\alpha = 1$, $\beta = 1$ and, with $n = 5$, we observe $\sum x_i = 3.4$. Derive the posterior estimate for θ and use Python to get a 95% confidence interval.