

---

## 9.1.0 Bayesian Inference

The following is a general setup for a statistical inference problem: There is an unknown quantity that we would like to estimate. We get some data. From the data, we estimate the desired quantity. In the previous chapter, we discussed the **frequentist** approach to this problem. In that approach, the unknown quantity  $\theta$  is assumed to be a fixed (non-random) quantity that is to be estimated by the observed data.

In this chapter, we would like to discuss a different framework for inference, namely the **Bayesian** approach. In the Bayesian framework, we treat the unknown quantity,  $\Theta$ , as a random variable. More specifically, we assume that we have some initial guess about the distribution of  $\Theta$ . This distribution is called the *prior distribution*. After observing some data, we update the distribution of  $\Theta$  (based on the observed data). This step is usually done using *Bayes' Rule*. That is why this approach is called the Bayesian approach. The details of this approach will be clearer as you go through the chapter. Here, to motivate the Bayesian approach, we will provide two examples of statistical problems that might be solved using the Bayesian approach.

---

### Example 9.1

Suppose that you would like to estimate the portion of voters in your town that plan to vote for Party A in an upcoming election. To do so, you take a random sample of size  $n$  from the likely voters in the town. Since you have a limited amount of time and resources, your sample is relatively small. Specifically, suppose that  $n = 20$ . After doing your sampling, you find out that 6 people in your sample say they will vote for Party A.

**Solution**

Let  $\theta$  be the true portion of voters in your town who plan to vote for Party A. You might want to estimate  $\theta$  as

$$\hat{\theta} = \frac{6}{20} = 0.3$$

In fact, in absence of any other data, that seems to be a reasonable estimate. However, you might feel that  $n = 20$  is too small. Thus, your guess is that the error in your estimation might be too high. While thinking about this problem, you remember that the data from the previous election is available to you. You look at that data and find out that, in the previous election, 40% of the people in your town voted for Party A. How can you use this data to possibly improve your estimate of  $\theta$ ? You might argue as follows:

Although the portion of votes for Party A changes from one election to another, the change is not usually very drastic. Therefore, given that in the previous election 40% of the voters voted for Party A, you might want to *model* the portion of votes for Party A in the next election as a random variable  $\Theta$  with a probability density function,  $f_{\Theta}(\theta)$ , that is mostly concentrated around  $\theta = 0.4$ . For example, you might want to choose the density such that

$$E[\Theta] = 0.4$$

Figure 9.1 shows an example of such density functions. Such a distribution shows your *prior* belief about  $\Theta$  in the absence of any additional data. That is, before taking your random sample of size  $n = 20$ , this is your guess about the distribution of  $\Theta$ .

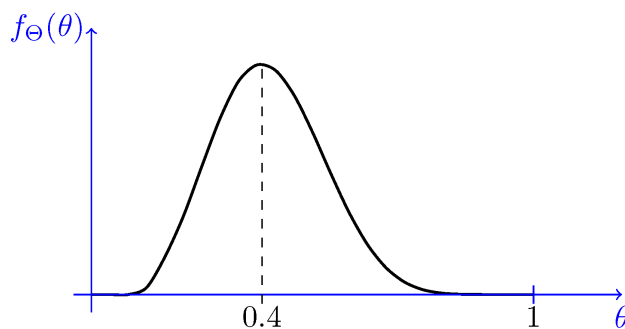


Figure 9.1 - An example of a prior distribution for  $\Theta$  in [Example 9.1](#)

Therefore, you initially have the prior distribution  $f_{\Theta}(\theta)$ . Then you collect some data, shown by  $D$ . More specifically, here your data is a random sample of size  $n = 20$  voters, 6 of whom are voting for Party A. As we will discuss in more detail, you can then proceed to find an updated distribution for  $\Theta$ , called the *posterior* distribution, using Bayes' rule:

$$f_{\Theta|D}(\theta|D) = \frac{P(D|\theta)f_{\Theta}(\theta)}{P(D)}.$$

We can now use the posterior density,  $f_{\Theta|D}(\theta|D)$ , to further draw inferences about  $\Theta$ . More specifically, we might use it to find point or interval estimates of  $\Theta$ .

---

### Example 9.2

Consider a communication channel as shown in Figure 9.2. We can model the communication over this channel as follows. At time  $n$ , a random variable  $X_n$  is generated and is transmitted over the channel. However, the channel is noisy. Thus, at the receiver, a noisy version of  $X_n$  is received. More specifically, the received signal is

$$Y_n = X_n + W_n,$$

where  $W_n \sim N(0, \sigma^2)$  is the noise added to  $X_n$ . We assume that the receiver knows the distribution of  $X_n$ . The goal here is to recover (estimate) the value of  $X_n$  based on the observed value of  $Y_n$ .

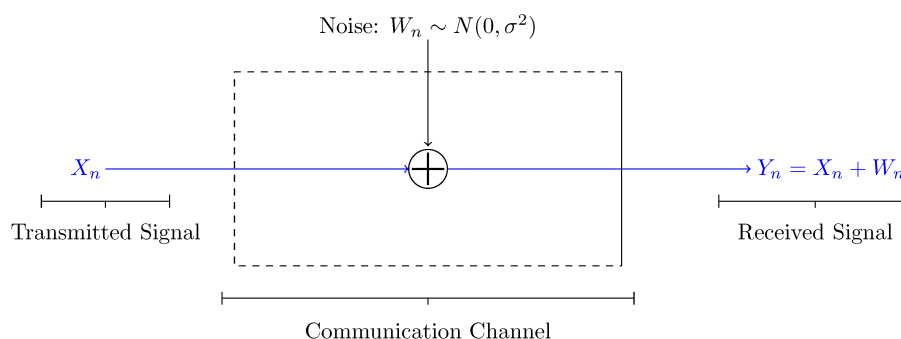


Figure 9.2 - Noisy communication channel in [Example 9.2](#)

### Solution

Again, we are dealing with estimating a random variable ( $X_n$ ). In this case, the *prior* distribution is  $f_X(x)$ . After observing  $Y_n$ , the *posterior* distribution can be written as

$$f_{X_n|Y_n}(x|y) = \frac{f_{Y_n|X_n}(y|x)f_X(x)}{f_Y(y)}.$$

Here, we have assumed both  $X$  and  $Y$  are continuous random variables. The above formula is a version of Bayes' rule. We will discuss the details of this approach shortly; however, as you'll notice, we are using the same framework as [Example 9.1](#). After finding the posterior distribution,  $f_{X_n|Y_n}(x|y)$ , we can then use it to estimate the value of  $X_n$ .

---

If you think about Examples [9.1](#) and [9.2](#) carefully, you will notice that they have similar structures. Basically, in both problems, our goal is to draw an inference about the

value of an unobserved random variable ( $\Theta$  or  $X_n$ ). We observe some data ( $D$  or  $Y_n$ ). We then use Bayes' rule to make inference about the unobserved random variable. This is generally how we approach inference problems in *Bayesian statistics*.

It is worth noting that Examples 9.1 and 9.2 are conceptually different in the following sense: In [Example 9.1](#), the choice of prior distribution  $f_{\Theta}(\theta)$  is somewhat unclear. That is, different people might use different prior distributions. In other words, the choice of prior distribution is *subjective* here. On the other hand, in [Example 9.2](#), the prior distribution  $f_{X_n}(x)$  might be determined as a part of the communication system design. In other words, for this example, the prior distribution might be known without any ambiguity. Nevertheless, once the prior distribution is determined, then one uses similar methods to attack both problems. For this reason, we study both problems under the umbrella of Bayesian statistics.

### Bayesian Statistical Inference

The goal is to draw inferences about an unknown variable  $X$  by observing a related random variable  $Y$ . The unknown variable is modeled as a random variable  $X$ , with **prior distribution**

$$\begin{aligned} f_X(x), & \quad \text{if } X \text{ is continuous,} \\ P_X(x), & \quad \text{if } X \text{ is discrete.} \end{aligned}$$

After observing the value of the random variable  $Y$ , we find the **posterior** distribution of  $X$ . This is the conditional PDF (or PMF) of  $X$  given  $Y = y$ ,

$$f_{X|Y}(x|y) \quad \text{or} \quad P_{X|Y}(x|y).$$

The posterior distribution is usually found using Bayes' formula. Using the posterior distribution, we can then find point or interval estimates of  $X$ .

Note that in the above setting,  $X$  or  $Y$  (or possibly both) could be random vectors. For example,  $X = (X_1, X_2, \dots, X_n)$  might consist of several random variables. However, the general idea of Bayesian statistics stays the same. We will specifically talk about estimating random vectors in [Section 9.1.7](#).