

# 1 Bayesian statistics

*Bayesian statistics*, as opposed to frequentist statistics, views probabilities merely as current opinion regarding the true state of the world. As new data is brought to light such opinion will be revised to reflect the new evidence. The way to update probabilities is prescribed by Bayes' theorem.

So Bayesian probabilities are subjective, in contrast to the objective probabilities of frequentist statistics.

## 1.1 Prior and posterior probabilities

Assume we have a given probability of an event  $A$  happening. This probability  $P(A)$  is known as the *prior* in Bayesian terminology. Now, assume we know that event  $B$  has occurred - this constitutes evidence. We should now adjust our probability of event  $A$  to the conditional probability  $P(A|B)$ , which is known as the *posterior*. The two are related by Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

## 1.2 Example: Radio quality

A given corporation produces radios. Of the last 200 truckloads of radios, 128 have been "bad" and 72 "good"; In the bad truckloads 44% of the radios were defective. In the good truckloads only 15%. Now, we're faced with determining whether a new truckload of radios is good or bad. Initially, since all we have is the information that 128 out of 200 truckloads have been bad, our prior probabilities would be:

$$P(B) = \frac{128}{200} = 64\%, \quad P(G) = \frac{72}{200} = 36\% \quad (2)$$

Here,  $B$  refers to the event "Bad truckload", and  $G$  to the event "Good truckload". However, we now sample one of the radios from the truckload. This radio turns out to be defective. What are the updated, posterior probabilities of the truckload being good or bad? To answer this, we need Bayes' theorem:

$$P(B|D) = \frac{P(D|B)P(B)}{P(D)} \quad (3)$$

Here  $D$  refers to the event "Defective radio".  $P(D|B)$  is the probability of a radio in a bad truckload being defective. We know that this is 44%. We

know that  $P(B) = 64\%$ . But what is  $P(D)$ ? By the law of total probability, this is:

$$P(D) = P(D|B)P(B) + P(D|G)P(G) = 44\% \cdot 64\% + 15\% \cdot 36\% = 33.56\% \quad (4)$$

Now, we can insert into equation (3):

$$P(B|D) = \frac{44\% \cdot 64\%}{33.56\%} = 83.9\% \quad (5)$$

By symmetry, the posterior probability of a good truckload has shrunk to  $P(G|D) = 100\% - 83.9\% = 16.1\%$ . The knowledge that the sample radio is defect makes us update our view of the world.

### 1.2.1 Radio quality with odds

One could also reformulate the example above using *odds*. The odds of a bad truckload is:

$$\frac{P(B)}{P(G)} = \frac{64\%}{36\%} = 1.78 \quad (6)$$

These are the prior odds. After the reveal of the defect radio, odds are:

$$\frac{P(B|D)}{P(G|D)} = \frac{83.9\%}{16.1\%} = 5.21 \quad (7)$$

These are the posterior odds. How are the two related? Let's use Bayes' theorem to find out:

$$\frac{P(B|D)}{P(G|D)} = \frac{P(D|B)P(B)/P(D)}{P(D|G)P(G)/P(D)} = \frac{P(D|B)P(B)}{P(D|G)P(G)} \quad (8)$$

So the prior odds times the quantity  $\frac{P(D|B)}{P(D|G)}$ , which can be interpreted as a *likelihood ratio*. The relation between these three quantities can be written:

$$\text{posterior odds} = \text{likelihood ratio} \cdot \text{prior odds} \quad (9)$$

## 1.3 Calculating posterior probabilities

In general, let  $\theta$  be a parameter (or a vector of parameters) used to describe an event. In the example above,  $\theta$  covered two option: Good or Bad truckload. In general,  $\theta$  may represent many - even infinitely many - different outcomes. Let  $X$  represent new evidence. By Bayes' theorem:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \quad (10)$$

For a given set of evidence  $X$  we wish to update on,  $P(X)$  is a constant, so this may also be expressed:

$$P(\theta|X) \propto P(X|\theta)P(\theta) \quad (11)$$

Or reworded to resemble equation (9):

$$\text{posterior probability} \propto \text{likelihood} \cdot \text{prior probability} \quad (12)$$

### 1.3.1 Radio quality revisited

The radio manufacturing company from above have made further inquiries into the distribution of percentages of defective radios in a truckload. In this situation, the general parameter  $\theta$  from last section corresponds to the defective percentage, which we will call  $\pi$ . It turns out, that this percentage seems to follow a beta distribution with parameters  $\alpha = 2$  and  $\beta = 4$ :

$$P(\pi) \propto \pi(1 - \pi)^3 \quad (13)$$

Since beta distributions are continuous, this is a probability *density* of finding a truckload with a defective percentage of  $\pi$ , rather than a probability - this is our prior. We're now presented with a new truckload of radios. Without any further information, our best bet is the prior distribution. However, we now inspect the truckload, taking a random sample of 5 radios. It turns out, that 3 of those are defective. How do we account for this new information to update our distribution? According to equation (11), we need to find the likelihood  $P(X|\pi)$ . Here  $X$  refers to the evidence of getting 3 out of 5 defective radios. In other words, this is a binomial likelihood<sup>1</sup>:

$$P(X|\pi) = \binom{5}{3} \pi^3 (1 - \pi)^2 \propto \pi^3 (1 - \pi)^2 \quad (14)$$

So the posterior distribution also turns out to be a beta distribution:

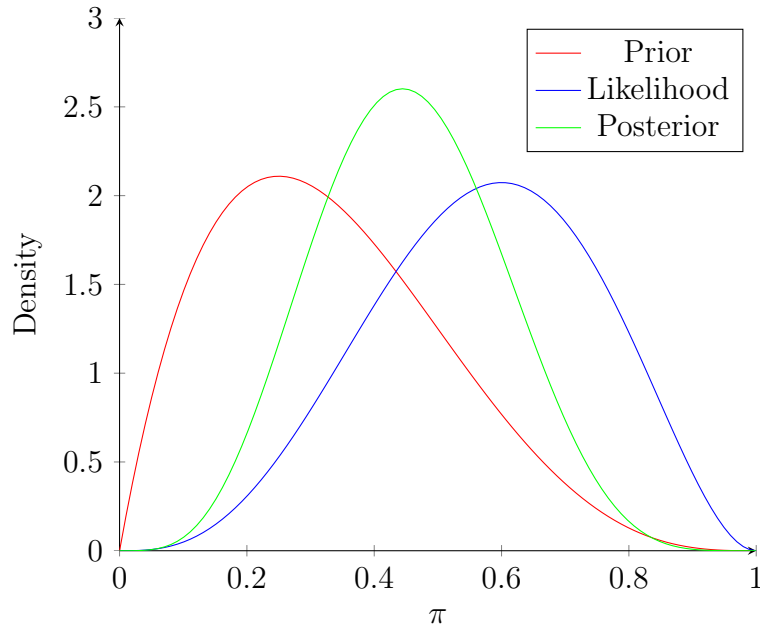
$$P(\pi|X) \propto \pi^4 (1 - \pi)^5 \quad (15)$$

This is a beta distribution with parameters  $\alpha = 5$  and  $\beta = 6$ . When both distributions are of the same type, we speak of *conjugate distributions*. Given a likelihood function, a prior that makes the distributions conjugate is known as a *conjugate prior*.

---

<sup>1</sup>This assumes that the truckload is large enough that the consecutive picking of radios to test does not affect  $\pi$  significantly.

The graph below shows the three distributions for comparison<sup>2</sup>. The prior is relatively optimistic regarding quality, and the likelihood less so. The posterior is a compromise between the two.




---

<sup>2</sup>The likelihood isn't a distribution, since it does not sum to 1. However, for clarity, in this figure it has been scaled as if it did.