

# Bayesian Statistics Workbook

Vu Nguyen Quang Duy

## Table of contents

<b>1</b>	<b>Bayes' Rules</b>	<b>1</b>
1.1	Chapter Summary . . . . .	1
1.1.1	Conditional vs unconditional probability . . . . .	1
1.1.2	Independent events . . . . .	2
1.1.3	Probability vs likelihood . . . . .	2
1.1.4	Joint and conditional probabilities . . . . .	3
1.1.5	Law of Total Probability (LTP) . . . . .	3
1.1.6	Bayes' Rule for events . . . . .	3
1.1.7	Discrete probability model . . . . .	4
1.1.8	Conditional probability model of data $Y$ . . . . .	4
1.1.9	The Binomial model . . . . .	4
1.1.10	Probability mass functions vs likelihood functions . . . . .	5
1.1.11	Bayes' Rule for variables . . . . .	5
1.1.12	Proportionality . . . . .	5
1.2	Exercises . . . . .	7
1.2.1	Building up to Bayes' Rule . . . . .	7

## 1 Bayes' Rules

### 1.1 Chapter Summary

#### 1.1.1 Conditional vs unconditional probability

Let  $A$  and  $B$  be two events, The **unconditional probability** of  $A$ , measures the probability of observing  $A$ , without any knowledge of  $B$ . In contrast, the **conditional probability** of  $A$  given  $B$ ,  $P(A|B)$ , measures the probability of observing  $A$  in light of the information that  $B$  occurred.

Conditional probabilities are fundamental to Bayesian analyses. In general, comparing the conditional vs unconditional probabilities,  $P(A|B)$  vs  $P(A)$ , reveals the extent to which information about  $B$  informs our understanding of  $A$ . In some cases, the certainty of an event  $A$  might *increase* or *decrease* in light of new data  $B$ . In other words:

$$P(A|B) > P(A) \text{ Or } P(A|B) < P(A)$$

The *order* of conditioning is also important. Since they measure two it's typically the case that:

$$P(A|B) \neq P(B|A)$$

### 1.1.2 Independent events

Two events  $A$  and  $B$  are **independent** if and only if the occurrence of  $B$  does not tell us anything about the occurrence of  $A$ :

$$P(A|B) = P(A)$$

### 1.1.3 Probability vs likelihood

When  $B$  is known, the **conditional probability function**  $P(\cdot|B)$  allows us to compare the probabilities of an unknown event,  $A$  and  $\bar{A}$ , occurring with  $B$ :

$$P(A|B) \text{ vs } P(\bar{A}|B)$$

When  $A$  is known, the **likelihood function**  $L(\cdot|A) = P(A|\cdot)$  allows us to evaluate the relative compatibility of data  $A$  with events  $B$  or  $\bar{B}$ :

$$L(B|A) \text{ vs } L(\bar{B}|A)$$

#### 1.1.4 Joint and conditional probabilities

For events  $A$  and  $B$ , the joint probability of  $A \cap B$  is calculated by weighting the conditional probability of  $A$  given  $B$  by the marginal probability of  $B$ :

$$P(A \cap B) = P(A|B)P(B) \quad (1)$$

Thus when  $A$  and  $B$  are *independent*,

$$P(A \cap B) = P(A)P(B)$$

Dividing both sides of Equation 1 by  $P(B)$ , and assuming  $P(B) \neq 0$ , reveals the definition of the conditional probability of  $A$  given  $B$ :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2)$$

Thus, to evaluate the chance that  $A$  occurs in light of information  $B$  we can consider the chance that they occur together,  $P(A \cap B)$ , relative to the chance that  $B$  occurs at all,  $P(B)$

#### 1.1.5 Law of Total Probability (LTP)

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) \quad (3)$$

#### 1.1.6 Bayes' Rule for events

For events  $A$  and  $B$ , the posterior probability of  $B$  given  $A$  follows by combining Equation 1 with Equation 2 and recognizing that we can evaluate data  $A$  through the likelihood function,  $L(B|A) = P(A|B)$  and  $L(\bar{B}|A) = P(A|\bar{B})$ :

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)L(B|A)}{P(A)} \quad (4)$$

where by the Law of Total Probability Equation 3:

$$P(A) = P(B)L(B|A) + P(\bar{B})L(\bar{B}|A) \quad (5)$$

More generally,

$$\text{posterior} = \frac{\text{prior} \cdot \text{likelihood}}{\text{normalizing constant}}$$

### 1.1.7 Discrete probability model

Let  $Y$  be a discrete random variable. The probability model of  $Y$  is specified by a **probability mass function (pmf)**  $f(y)$ . This pmf defines the probability of any given outcome  $y$ ,

$$f(y) = P(Y = y)$$

and has the following properties:

- $0 \leq f(y) \leq 1$  for all  $y$ , and
- $\sum_{\text{all } y} f(y) = 1$ , i.e., the probabilities of all possible outcomes of  $y$  sum to 1.

### 1.1.8 Conditional probability model of data $Y$

Let  $Y$  be a discrete random variable and  $\pi$  be a parameter upon which  $Y$  depends. The conditional probability model of  $Y$  given  $\pi$  is specified by conditional pmf  $f(y|\pi)$ . This pmf specifies the conditional probability of observing  $y$  given  $\pi$ ,

$$f(y|\pi) = P(Y = y|\pi)$$

and has the following properties:

- $0 \leq f(y|\pi) \leq 1$  for all  $y$ , and
- $\sum_{\text{all } y} f(y|\pi) = 1$

### 1.1.9 The Binomial model

Let random variable  $Y$  be the *number of successes* in a *fixed number of trials*  $n$ . Assume that the trials are *independent* and that the *probability of success* in each trial is  $\pi$ . Then the conditional dependence of  $Y$  on  $\pi$  can be modeled by the Binomial model with **parameters**  $n$  and  $\pi$ . In mathematical notation:

$$Y|\pi \sim \text{Bin}(n, \pi)$$

where “ $\sim$ ” can be read as “modeled by”. Correspondingly, the Binomial model is specified by **conditional pmf**

$$f(y|\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y} \text{ for } y \in \{0, 1, 2, \dots, n\} \quad (6)$$

### 1.1.10 Probability mass functions vs likelihood functions

When  $\pi$  is known, the **conditional pmf**  $f(\cdot|\pi)$  allows us to compare the probabilities of different possible values of data  $Y$  (e.g.,  $y_1$  or  $y_2$ ) occurring with  $\pi$ :

$$f(y_1|\pi) \text{ vs } f(y_2|\pi)$$

When  $Y = y$  is known, the **likelihood function**  $L(\cdot|y) = f(y|\cdot)$  allows us to compare the relative likelihood of observing data  $y$  under different possible values of  $\pi$  (e.g.,  $\pi_1$  or  $\pi_2$ ):

$$L(\pi_1|y) \text{ vs } L(\pi_2|y)$$

Thus,  $L(\cdot|y)$  provides the tool we need to evaluate the relative compatibility of data  $Y = y$  with various  $\pi$  values.

### 1.1.11 Bayes' Rule for variables

For any variables  $\pi$  and  $Y$ , let  $f(\pi)$  denote the prior pmf of  $\pi$  and  $L(\pi|y)$  denote the likelihood function of  $\pi$  given observed data  $Y = y$ . Then the posterior pmf of  $\pi$  given data  $Y = y$  is:

$$f(\pi|y) = \frac{\text{prior} \cdot \text{likelihood}}{\text{normalizing constant}} = \frac{f(\pi)L(\pi|y)}{f(y)} \quad (7)$$

where, by the Law of Total Probability, the overall probability of observing data  $Y = y$  across all possible  $\pi$  is:

$$f(y) = \sum_{\text{all } \pi} f(\pi)L(\pi|y)$$

### 1.1.12 Proportionality

Since  $f(y)$  is merely a normalizing constant which does not depend of  $\pi$ , the posterior pmf  $f(\pi|y)$  is proportional to the product of  $f(\pi)$  and  $L(\pi|y)$ :

$$f(\pi|y) = \frac{f(\pi)L(\pi|y)}{f(y)} \propto f(\pi)L(\pi|y)$$

That is,

$$\text{posterior} \propto \text{prior} \cdot \text{likelihood}$$

The significance of this proportionality is that all the information we need to build the posterior model is held in the prior and likelihood.

## 1.2 Exercises

### 1.2.1 Building up to Bayes' Rule

#### Exercise 1.1. *Comparing the prior and posterior*

For each scenario below, you're given a pair of events,  $A$  and  $B$ . Explain what you believe to be the relationship between the posterior and prior probabilities of  $B$ :  $P(B|A) > P(B)$  or  $P(B|A) < P(B)$

- a)  $A$  = you just finished reading Lambda Literary Award-winning author Nicole Dennis-Benn's first novel, and you enjoyed it!  $B$  = you will also enjoy Benn's newest novel.
- b)  $A$  = it's 0 degrees Fahrenheit in Minnesota on a January day.  $B$  = it will be 60 degrees tomorrow.
- c)  $A$  = the authors only got 3 hours of sleep last night.  $B$  = the authors make several typos in their writing today.
- d)  $A$  = your friend includes three hashtags in their tweet.  $B$  = the tweet gets retweeted.

#### Solution

a) **Answer:**  $P(B|A) > P(B)$

- The prior probability,  $P(B)$ : The general probability of enjoying Benn's newest novel before reading any of her previous work.
- The posterior probability,  $P(B | A)$ : The updated probability of enjoying Benn's newest novel, given that her first novel was read and enjoyed.

The event  $A$  (enjoying the first novel) is positive evidence that provides a reason to increase belief in event  $B$  (enjoying the newest novel). A favorable experience with the author's work makes the updated belief (the posterior) stronger and therefore higher than the initial belief (the prior).

b) **Answer:**  $P(B|A) < P(B)$

- The prior probability,  $P(B)$ : The general probability that it will be 60 degrees tomorrow.
- The posterior probability,  $P(B | A)$ : The updated probability that it will be 60 degrees tomorrow, given that it was 0 degrees Fahrenheit yesterday.

The event  $A$  (a temperature of 0°F yesterday) is negative evidence that provides a reason to decrease the belief in event  $B$  (a temperature of 60°F tomorrow). A temperature of 0°F makes it significantly less likely that the temperature will be a relatively mild 60°F the next day. This new information acts as negative evidence, causing a decrease in the belief of event  $B$ .

c) **Answer:**  $P(B|A) > P(B)$

- The prior probability,  $P(B)$ : The general probability that the authors will make several typos in their writing today.
- The posterior probability,  $P(B | A)$ : The updated probability that the authors will make several typos, given they only got 3 hours of sleep last night.

The event  $A$  (only 3 hours of sleep) is positive evidence that increases the probability of event  $B$  (making typos). Lack of sleep is a well-known factor that impairs cognitive function and attention to detail, making errors like typos more probable. The updated belief is therefore higher than the initial belief.

d) **Answer:**  $P(B|A) > P(B)$

- The prior probability,  $P(B)$ : The general probability that the tweet will be retweeted. This is the baseline likelihood without knowing anything about the tweet's content or format.
- The posterior probability,  $P(B | A)$ : the updated probability that the tweet will be retweeted, given that it includes three hashtags.

The event  $A$  (including three hashtags) is positive evidence that increases the probability of event  $B$  (the tweet being retweeted). Research on social media engagement shows that tweets with hashtags, especially a moderate number, tend to have wider reach and higher engagement, which includes retweets. Therefore, the updated belief is higher than the initial belief.

---

**Exercise 1.2.** *Marginal, conditional, or joint?*

Define the following events for a resident of a fictional town:

- $A$  = drives 10 miles per hour above the speed limit,
- $B$  = gets a speeding ticket,
- $C$  = took statistics at the local college,
- $D$  = has used R,
- $E$  = likes the music of Prince,
- $F$  = is a Minnesotan.

Several facts about these events are listed below. Specify each of these facts using probability notation, paying special attention to whether it's a marginal, conditional, or joint probability.



- a) 73% of people that drive 10 miles per hour above the speed limit get a speeding ticket.
- b) 20% of residents drive 10 miles per hour above the speed limit.
- c) 15% of residents have used R.
- d) 91% of statistics students at the local college have used R.
- e) 38% of residents are Minnesotans that like the music of Prince.
- f) 95% of the Minnesotan residents like the music of Prince.

### Solution

- a) **Answer:** conditional probability

This fact gives the probability of getting a speeding ticket ( $B$ ) given that a person is already driving 10 miles per hour above the speed limit ( $A$ ).

The probability notation for this is:  $P(B|A) = 0.73$

- b) **Answer:** marginal probability

This facts only describe the proportion of residents that drives 10 miles per hour above the speed limit ( $A$ ) without any conditions.

The probability notation for this is:  $P(A) = 0.20$

- c) **Answer:** marginal probability

This facts only describe the proportion of residents that have used R ( $D$ ) without any conditions.

The probability notation for this is:  $P(D) = 0.15$

- d) **Answer:** conditional probability

This fact It states the probability of a person having used R ( $D$ ) given that they took statistics at the local college ( $C$ ).

The probability notation for this is:  $P(D|C) = 0.91$

- e) **Answer:** joint probability

This is a joint probability because it refers to the likelihood of two events happening at the same time: being a Minnesotan and liking the music of Prince.

The probability notation for this is:  $P(E \cap F) = 0.38$

- f) **Answer:** conditional probability

This is a conditional probability. It states the probability of a person liking the music of Prince ( $E$ ) given that they are a Minnesotan ( $F$ ).

The probability notation for this is:  $P(E|F) = 0.95$