Bayesian Statistics Workbook

Vu Nguyen Quang Duy

Table of contents

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 occured.

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:

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

Finally, information about B doesn't always change our understanding of A. We call them **Independent events**.

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 $\overline{A}$, occurring with B:

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

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

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

1.1.4 Joint and conditional probabilities

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

$$P(A \cap B) = P(A|B)P(B) \tag{1}$$

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:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \tag{2}$$

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)

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}) \tag{3}$$

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 $L(\overline{B}|A) = P(A|\overline{B})$:

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

where by the Law of Total Probability Equation 3:

$$P(A) = P(B)L(B|A) + P(\overline{B})L(\overline{B}|A) \tag{5}$$

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)** $f(y)$. Thi 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 π ve a parameter uop which Y depends. The the conditional probability model of Y given π is specified by conditional pmf $f(y|\pi)$. This pmf specifies the conditional probability of observing y given π,

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

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 Binomal 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 depdendence of Y on π can be modeled by the Binomal model with **parameters** n and π. In mathematical notation:

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

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

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

1.1.10 Probability mass functions vs likelihood functions

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

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

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 π (e.g., π_1 or π_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 π values.

1.1.11 Bayes' Rule for variables

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

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

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

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

1.1.12 Propotionality

Since $f(y)$ is merely a normalizing constant which does not depend of π, the posterior pmf $f(\pi|y)$ is propotional 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 Excercises

1.2.1 Buidling 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 enjot 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, $P(B \mid A)$: 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, $P(B \mid A)$: 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, $P(B \mid A)$: 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, $P(B \mid A)$: he 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.