Data Science Math Skills
Paul Bendich and Daniel Egger
Duke University

Bayes' Theorem, Part 1

Video companion

1 Derivation

Starting from the product rule,

$$P(A \mid B) = \frac{P(A, B)}{P(B)},$$

and multiplying by the probability of B $P(B)$, gives

$$P(A \mid B)P(B) = P(A, B).$$

Substituting the equivalent $P(B, A)$ for $P(A, B)$,

$$P(A \mid B)P(B) = P(B, A),$$

and using the product rule $P(B \mid A) = P(B, A)/P(A)$, gives

$$P(A \mid B)P(B) = P(B \mid A)P(A),$$

which when rearranged is Bayes' theorem.

Bayes' theorem:

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

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

2 Inverse probability

An inverse probability problem is one where the answer is in the form of the probability that a certain process with a certain probability parameter is being used to generate the observed data.

The symbol B is used to represent the observed data. The symbol A_i is used to represent a possible process with probability parameter θ_i.

Example: Urn 1 has 20% white marbles, and urn 2 has 10% white marbles. We observe three white marbles in a row drawn with replacement. What is the probability that we are observing urn 1? Urn 2?

What we know: parameter

A_i

Process		P(white marble)
Urn 1	A_1	20%
Urn 2	A_2	10%

θ_i

where "white marble" is the parameter.

In forward probability we are interested in the probability of an event given a known process.

In this problem, we know the outcome and want to know how probable it is that each process was involved.

Written in terms of conditional probability,

$$P(\text{process parameter} \mid \text{observed data}) = \frac{P(\text{observed data} \mid \text{process parameter}_i)P(\text{process parameter}_i)}{P(\text{data} \mid \text{process}_1)P(\text{process}_1) + P(\text{data} \mid \text{process}_2)P(\text{process}_2) + ... + P(\text{data} \mid \text{process}_n)P(\text{process}_n)}$$

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\underbrace{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + ... + P(B \mid A_n)P(A_n)}}$$

denominator is simply P(B),
but expanded into the sum rule of marginal probabilities

Example arithmetic:

First, solve for likelihoods:

Urn 1: $P(3 \text{ white marbles in a row} \mid 20\% \text{ white}) = (0.2)(0.2)(0.2) = 8/1000$
Urn 2: $P(3 \text{ white marbles in a row} \mid 10\% \text{ white}) = (0.1)(0.1)(0.1) = 1/1000$

Using principle of indifference,

$$P(A_1) = 0.5$$
$$P(A_2) = 0.5$$

we are not biased to choose an urn

because we are neutral before observing any data. These are called the "prior probability."

$$\begin{aligned} P(A_1 \mid B) &= \frac{P(B \mid A_1)P(A_1)}{P(B, A_1) + P(B, A_2)} \\ &= \frac{P(B \mid A_1)P(A_1)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2)} \\ &= \frac{(8/1000)(1/2)}{(8/1000)(1/2) + (1/1000)(1/2)} = 8/9 \end{aligned}$$

Probability of choosing 3 white marbles in a row $(0.2\%)^3$ from urn 1 (50% chance)

Possibilities from urn 1 and 2, with respective choosing probabilities

Probability that we observed urn 1: 8/9
Probability that we observed urn 2: 1/9