BST 6200 Spatial Statistics and Disease Mapping

Bayesian Methods

Steven E. Rigdon

Spring 2020

Motivating Example

Medical devices are produced at three factories, in Atlanta (A), Boston (B), and Chicago (C). Factory A makes half of the devices, Factory B makes 30%, and Factory C makes 20%.

Defect rates vary across factories. In Factory A, 2% of devices are defective. The defective rates in Factories B and C are 3% and 5%, respectively.

All devices are shipped to a warehouse in Detroit before being sent to users.

Q1: What is the probability that an item selected at random in the warehouse is defective?

Q2: An item is found to be defective. What is the probability that it came from Factory A? From Factory B? From Factory C?

Notation

Notation:

$\cap$ means set intersection, the logical "and".

$\cup$ means set union or the logical "or".

$P(A)$ means the probability of event A.

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

If A and B are *mutually exclusive*, then $A \cap B = \emptyset$ and $P(A \cup B) = P(A) + P(B)$

The conditional probability of A given B is

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

If events A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$. We can then show that $P(A \cap B) = P(A)P(B)$.

Q1: What is the probability that an item selected at random in the warehouse is defective?

Q2: An item is found to be defective. What is the probability that it came from Factory A? From Factory B? From Factory C?

Law of Total Probability

Suppose $A_1, A_2, \ldots, A_k$ are mutually exclusive and when taken together cover the entire space of possible outcomes. (IOW, every time the random experiment is done, one and only one of the A_i occurs.)

Let B be any set. Then

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_k \cap B)$$

and

$$\begin{aligned}
P(B) &= P\big((A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_k \cap B)\big) \\
&= P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_k \cap B) \\
&= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_k)P(B|A_k) \\
&= \sum_{i=1}^{k} P(A_i)P(B|A_i)
\end{aligned}$$

Bayes Theorem

$$
\begin{aligned}
P(A_j|B) &= \frac{P(A_j \cap B)}{P(B)} \\
&= \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)}
\end{aligned}
$$

Notice how Bayes Theorem "turns around" conditional probabilities.

Thinking about Bayes theorem in context of inference:

- Prior: probabilities for each factory having made the device
- Unknown Parameter: which factory the device came from
- Data: whether an item is defective
- Posterior: probability for each factory having made the device given the data

Moving toward Bayesian Inference: Example

Suppose I have three coins:

1. a "fake" coin where both sides are tails; this coin will always result in a tail.
2. a fair coin; this coin has probabilities of $1/2$ of being a head on every toss.
3. a "fake" coin where both sides are heads; this coin will always result in a head.

I select one coin at random, i.e., each with probability $1/3$.

(a) What is the probability that on three tosses I obtain three heads?

(b) What is the probability that I have coin number 2 (the fair coin) given that I obtain three heads?

(c) Repeat (b) for coin 1 and coin 3.

Solutions:

Thinking about Bayes theorem in context of inference:

- Prior: probabilities of selecting each coin
- Unknown Parameter: which coin was selected
- Data: how many heads we observed
- Posterior: probability that the selected coin is number 1, 2, or 3, given the number of observed heads.

Moving closer to inference with Bayes Theorem

Suppose I run a small clinical trial. Of interest is the proportion of patients who are cured by the treatment. (This is a poorly designed trial because there is no control, but it is just for illustration.)

Let θ be the probability that a person is cured.

We don't know θ, so we run $n = 10$ trials and we observe $X = 7$ successes.

What can we say about θ after observing $X = 7$?

How Bayesian statisticians look at this problem:

We don't know θ.

We are uncertain about θ.

We express our uncertainty in probabilistic terms. (Classical statisticians, or frequentists, don't do this!)

Before observing any data, we are completely ignorant about the value of θ, so let's assume that θ can be any of the numbers $\theta = 0.00, 0.01, 0.02, \ldots, 0.99, 1.00$.

Since we are completely ignorant, we'll assign a prior probability of $\frac{1}{101}$ to each of these 101 numbers.

Find the Posterior Probability Distribution

1. What is the probability of observing $X = 7$?
2. What is the probability that $\theta = 0.50$ given that we observe $X = 7$?
3. Repeat part (2) for an arbitrary value of θ.
4. Sketch a graph of the posterior probabilities from part (3).

Solution

A continuous prior for θ

Why couldn't θ be *any* number between 0 and 1?

Suppose θ has a uniform prior distribution over the interval $[0, 1]$. IOW,

$$p(\theta) = \begin{cases} 1, & 0 \leq \theta \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

The *likelihood function*, denoted $L(\theta|x)$ is the (joint) PDF or PMF of X for a given value of θ.

$$L(\theta|x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

The Posterior Distribution

$$p(\theta|x) = \frac{p(\theta)L(\theta|x)}{\int p(\theta)L(\theta|x)\ d\theta}$$

Interpreting the Posterior Distribution

1. The posterior distribution reflects what we believe about θ after observing the data.
2. If we want a point estimate of θ, we can take the posterior mean. (Other options are available.)
3. If we want an interval estimate for θ, we could find values a and b such that $P(a < \theta < b) = 1 - \alpha$ for a given α.