Lecture 2A
MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

Today's agenda

Today's lecture will cover

- Bayes' theorem
- Use Bayes' theorem in Bayesian inference to compute posterior probabilities with discrete priors

Review of Bayes' theorem

- Bayes' theorem was formulated by Thomas Bayes in the 18th century.
- It's a basic part of probability theory.
- It's also essential for Bayesian statistics.

Bayes' theorem

- Recall that Bayes' theorem allows us to 'invert' conditional probabilities.
- Suppose we have events A and B, with $p(B) > 0$. Then

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

This is so because

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

and hence

$$P(A \mid B)\, P(B) = P(A \cap B) = P(B \mid A)\, P(A) \quad \text{multiplication rule}$$

Bayes' theorem

- Let Ω be the sample space. Suppose it is partitioned into a set of mutually exclusive and exhaustive events $A_1, A_2, \ldots, A_m$. (i.e. at least one must occur and no two can occur).
- The event B happens under any of the hypotheses A_i with a known conditional probability $P(B \mid A_i)$.
- Then we can write

$$P(A_i \mid B) = \frac{P(B \mid A_i)\, P(A_i)}{P(B)} = \frac{P(B \mid A_i)\, P(A_i)}{\sum_{j=1}^{m} P(B \mid A_j)\, P(A_j)}$$

- **Why?**

Diagnostic test example

Why does it matter?

Suppose HIV has prevalence of $1/2000$ in the population.
Suppose a test for HIV has 90% sensitivity and 95% specificity.

- So $a = P($ test $+$ve $|$ HIV $+$ve$) = 0.9$, and
- $b = P($ test -ve $|$ HIV -ve$) = 0.95$

Suppose a patient is screened and has a positive test.
Represent this information with a tree and use Bayes' theorem to compute

- What is the probability that someone who tests positive is HIV positive?
- What is the probability that someone who tests positive is HIV negative?

Diagnostic test example

- By Bayes' Theorem

$$P(\text{HIV +ve} \mid \text{ test +ve}) = \frac{P(\text{ test +ve} \mid \text{HIV +ve})P(\text{HIV +ve})}{P(\text{ test +ve})} \approx 1\%$$

  much less than the sensitivity of the test, $P(\text{ test +ve} \mid \text{HIV +ve})$, but higher than $P(\text{HIV +ve}) = 1/2000$.

- Mixing up $P(A|B)$ with $P(B|A)$ is the Prosecutor's Fallacy; a small probability of evidence given innocence need NOT mean a small probability of innocence given evidence.

Prosecutor's fallacy: Sally Clark

- After the sudden death of two baby sons, Sally Clark (above, center) was sentenced to life in prison in 1999 in the UK.
- Among other errors, expert witness Prof Roy Meadow (above right) had wrongly interpreted the small probability of two cot deaths as a small probability of Clark's innocence.
- After a long campaign, including refutation of Meadow's statistics, Clark was released and cleared in 2003
- She was unable to recover from the effects of her conviction. She died in 2007 from alcohol poisoning. See Convicted on Statistics?

Terminology and Bayes' theorem in tabular form

Diagnostic test example

- Data: the results of our experiment. In this case, the test is positive
- Hypotheses: The hypotheses are the possible answers to the question being asked. In this case they are: the subject is HIV positive and the subject HIV negative.
- Prior probabilities: The priors are the probabilities of the hypotheses prior to collecting data. In this case, before seeing the test result, the probability that someone is HIV +ve and the probability that someone is HIV negative in the general population

$$P(\text{HIV +ve}) = 1/2000, \quad P(\text{HIV -ve}) = 1999/2000$$

Diagnostic test example

- Likelihood: The likelihood is the probability of the data assuming that the hypothesis is true. In this case there are two likelihoods, one for each hypothesis

$$P(\text{test +ve}|\text{HIV +ve}) = 0.90 \quad P(\text{test +ve}|\text{HIV -ve}) = 0.05$$

- Posterior probabilities: The posteriors are the probabilities of the hypotheses given the data. In this case

$$P(\text{HIV +ve}|\text{test +ve}) = 0.0089 \quad P(\text{HIV -ve}|\text{test +ve}) = 0.9911$$

Diagnostic test example

By Bayes' theorem

$$P(\text{HIV +ve}|\text{test +ve}) = \frac{P(\text{ test +ve} \mid \text{HIV +ve})P(\text{HIV +ve})}{P(\text{ test +ve})}.$$

$$P(\text{HIV -ve}|\text{test +ve}) = \frac{P(\text{ test +ve} \mid \text{HIV -ve})P(\text{HIV -ve})}{P(\text{ test +ve})}.$$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{total prob. of data}}.$$

Terminology and Bayes' theorem in tabular form

Diagnostic test example: Calculation using a Bayesian update table

- We organise all of these neatly in a Bayesian updating table

Hypothesis	Prior	Likelihood	Bayes numerator	Posterior
HIV +ve	1/2000	0.90	0.00045	0.0089
HIV -ve	1999/2000	0.05	0.049975	0.9911
Total	1	NO SUM TO 1	0.050425	1

- Law of total probability: $P(\text{data}) = P(\text{test +ve}) =$ sum of Bayes numerator column $= 0.050425$
- Bayes theorem:

$$P(\text{HIV +ve} \mid \text{test +ve}) = \frac{P(\text{ test +ve} \mid \text{HIV +ve})P(\text{HIV +ve})}{P(\text{ test +ve})}$$
$$= \frac{\text{likelihood} \times \text{prior}}{\text{total prob. of data}}$$

Bayes' theorem

- We can express Bayes' theorem

$$P(\text{hypothesis} \mid \text{data}) = \frac{P(\text{data}|\text{hypothesis})P(\text{hypothesis})}{P(\text{data})}$$

- With the terminology

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{total prob. of data}}.$$

- With the data fixed, the denominator just serves to normalise the posterior to 1. So we can express the Bayes' theorem as

$$\text{posterior} \propto \text{likelihood} \times \text{prior}.$$

- Bayesian updating: The process of going from the prior to the posterior is called Bayesian updating. Bayesian updating uses the data to update our initial beliefs about the hypotheses.

Board Question: Coins

- There are three types of coins which have different probabilities of heads
  - Type A coins are fair, with probability 0.5 of heads.
  - Type B are bent and have probability 0.6 of heads.
  - Type C are bent and have probability 0.9 of heads.

  Suppose I have a drawer containing 5 coins: 2 of type A, 2 of type B, and 1 of type C. I pick a coin at random, and without showing you the coin I flip it once and get heads.
- Use Bayes' theorem to compute the probabilities that the coin is type A, type B or type C.
- Identify the data, hypotheses, likelihoods, prior probabilities and posterior probabilities.
- Make a Bayesian update table and compute the posterior probabilities that the chosen coin is each of the three coins.

Board Question: Coins

Food for thought

- Suppose that you didn't know how many coins of each type were in the drawer. You picked one at random and got heads.
- How would you go about deciding which coin type if any was most supported by the data?

Board Question: Dice

- Five dice in the drawer: 4-sided, 6-sided, 8-sided, 12-sided, 20-sided.
- Suppose I picked one at random and, without showing it to you, rolled it and reported a 13.
- Make a Bayesian update table and compute the posterior probabilities that the chosen die is each of the five dice.
- Same question if I rolled a 5.

The Bayes variation

- Sometimes it is more convenient to work with random variables.
- Let X and Y are continuous random variables with joint density $f(x, y)$

$$f(x \mid y) = \frac{f(y \mid x)\, f(x)}{f(y)}$$
$$= \frac{f(y \mid x)\, f(x)}{\int f(y \mid x')\, f(x')\, dx'}$$

- The formulae follow from standard results about conditional and marginal pdfs.

The Bayes variation

- If X and Y are discrete replace pdf with pmf and integral with sum

$$P(X = x \mid Y = y) = \frac{P(Y = y \mid X = x)P(X = x)}{\sum_{x'} P(Y = y \mid X = x')P(X = x')}.$$

- If X continuous and Y discrete

$$f(x|Y = y) = \frac{f(x)P(Y = y \mid x)}{\int f(x')P(Y = y \mid x')\, dx'}.$$

- If X discrete and Y continuous

$$P(X = x|y) = \frac{P(X = x)f(y \mid x)}{\sum_{x'} P(X = x')f(y \mid x')}.$$

Bayesian inference

- Probability model $p(y \mid \theta)$ depends on a set of parameters θ.
- Have data y, assumed to be generated by this probability model.
- These two parts are the same as frequentist, likelihood-based inference.
- In frequentist, θ is fixed and $p(y \mid \theta)$ assigns a probability to Y for each fixed valued of θ

Bayesian inference

- In Bayesian inference, all uncertainty is specified by probability distributions.
- This includes uncertainty about the parameters, θ
- So we start with a probability distribution for the parameters $p(\theta)$, called the prior distribution
- The prior is a subjective distribution, based on experimenter's belief, and is formulated before the data y are seen.

Bayesian inference

- Let y be the observed data (the result of the experiment, e.g., test is positive)
- We then update the prior distribution for θ using y.
- This updating is done using Bayes' theorem.

$$p(\theta \mid y) = \frac{p(\theta)\, p(y \mid \theta)}{p(y)},$$

where the observed data enters through the likelihood $p(y \mid \theta)$.

- We don't normally need to find $p(y)$, which is given by

$$p(y) = \int p(\theta')\, p(y \mid \theta')\, d\theta' \ \text{ or } \ \sum_{\theta'} p(\theta')\, p(y \mid \theta')$$

What does it mean?

$$p(\theta \mid y) \propto p(\theta)\, p(y \mid \theta) \tag{1}$$

Posterior $\propto$ prior $\times$ likelihood

- $p(y \mid \theta)$ is the likelihood and it the probability of data y given the true θ
- Start with initial beliefs/information about θ, $p(\theta)$ - this is the prior distribution formulated before the data are seen.
- Bayesian updating: Update the prior distribution using the data y, using (1)
- The updated prior, $p(\theta \mid y)$ is called the posterior distribution .
- We base our inferences about θ based on this posterior distribution.

Bayesian updating with discrete data, discrete prior

Diagnostic test example

- We can redo the diagnostic test example, using discrete pmf of the data and the parameters (hypotheses).
- We need to assign values to events (HIV +ve is 1 and HIV -ve is 0).
- Let's use the following notation
- θ is the value of the hypothesis. In this case, $\theta = 1$ means HIV +ve and $\theta = 0$ means HIV -ve. (θ is a Bernoulli random variable)
- $p(\theta)$ is the prior pmf of the hypothesis. In this case,

$$p(\theta = 1) = 1/2000 \quad p(\theta = 0) = 1999/2000$$

Bayesian updating with discrete data, discrete prior

Diagnostic test example

- Data: $x = 1$ means the test is positive.
- Likelihood. the probability of the data $x = 1$, given the true θ (This is not a pmf). In this case,

$$p(x = 1|\theta = 1) = 0.90 \quad p(x = 1|\theta = 0) = 0.05$$

- $p(\theta = 1|x = 1)$ and $p(\theta = 0|x = 1)$ are the posterior pmf of the θ given the data $x = 1$

Bayesian updating with discrete data, discrete prior

Diagnostic test example

- The Bayesian update table with pmf prior and discrete data is

Hypothesis	prior	likelihood	Bayes numerator	posterior
θ	$p(\theta)$	$p(x=1\|\theta)$	$p(x=1\|\theta)p(\theta)$	$p(\theta\|x=1)$
$\theta=1$	$1/2000$	0.90	0.00045	$p(\theta=1\|x=1)=0.0089$
$\theta=0$	$1999/2000$	0.05	0.049975	$p(\theta=0\|x=1)=0.9911$
Total	1	NOT SUM TO 1	$p(x=1)=0.050425$	1

- Law of total probability: $p(x=1) = p(x=1|\theta=1)p(\theta=1) + p(x=1|\theta=0)p(\theta=0) = 0.050425$.

- Bayes' theorem: $p(\theta=1|x=1) = \frac{p(x=1|\theta=1)p(\theta=1)}{p(x=1)} = 0.0089$.
- Similarly for $p(\theta=0|x=1) = 0.9911$.

Borard question

- Using the notation for discrete pmf $p(\theta)$ etc., redo example with coins and dice.