Lab 8

Probability and Bayesian Estimation

Revised from
Zhiwei Li

Probability and logic

- Suppose you know the probability of event A and event B happening to be $P(A)$ and $P(B)$. How do you describe the probability of:
  - $P(A \text{ and } B)$
  - $P(A \text{ or } B)$

Probability and logic

- Suppose you know the probability of event A and event B happening to be $P(A)$ and $P(B)$. How do you describe the probability of:
  - $P(A \text{ and } B) = P(A)P(B)$ *if they are independent events*
  - $P(A \text{ or } B) = P(A) + P(B)$ *if they are mutually exclusive*
  - $P(A \textit{ or } B) = P(A) + P(B) - P(A)P(B)$ *if independent*

Independence and Marginals

	A	**not A**	
B			P(B)
not B			
	P(A)		

Independence and Marginals

Fill in marginals:

	A	**not A**	
B			P(B)
not B			1-P(B)
	P(A)	1-P(A)	1

Independence and Marginals

If A and B are independent:

	A	**not A**	
B	P(A)P(B)		P(B)
not B			1-P(B)
	P(A)	1-P(A)	1

Independence and Marginals

Fill in based on marginals:

	A	not A	
B	P(A)P(B)	P(B)-P(A)P(B) = P(B)[1-P(A)]	P(B)
not B	P(A)-P(A)P(B) = P(A)[1-P(B)]	[1-P(A)][1-P(B)]	1-P(B)
	P(A)	1-P(A)	1

Independence and Marginals

That is, *all* pairs are independent

	A	not A	
B	P(A)P(B)	P(B)-P(A)P(B) = P(B)[1-P(A)]	P(B)
not B	P(A)-P(A)P(B) = P(A)[1-P(B)]	[1-P(A)][1-P(B)]	1-P(B)
	P(A)	1-P(A)	1

Conjunction fallacy

- You flip a coin for 6 times. Which of the following sequence is most likely given the coin is fair?
  - THTTT
  - HTHTTT
  - THHHHH

Conjunction fallacy

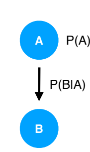
- You flip a coin for 6 times. Which of the following sequence is most likely given the coin is fair?
  - THTTT, probability $=(1/2)^5$
  - HTHTTT, probability $=(1/2)^6$
  - THHHHH, probability $=(1/2)^6$

Conjunction fallacy

- You flip a coin for 6 times. Which of the following sequence is most likely given the coin is fair?
  - THTTT, probability $=(1/2)^5$
  - HTHTTT, probability $=(1/2)^6$
  - THHHHH, probability $=(1/2)^6$
  - 1 head in 5 flips, probability = $\binom{5}{1}(1/2)^1(1/2)^{5-1} \approx 0.1562$
  - 2 heads in 6 flips, probability = $\binom{6}{2}(1/2)^2(1/2)^{6-2} \approx 0.2344$

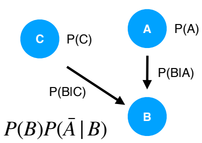
Probability and logic

- Suppose you know the events A and B *are not independent*. How do you describe the probability of:
  - $P(A$ and $B)$
  - $P(A$ or $B)$
  - $P(A$ or B, but not both)



Probability and logic

- Suppose you know the events A and B *are not independent*.
  How do you describe the probability of:
  - $P(A \text{ and } B) = P(A)P(B\,|\,A)$
  - $P(A \text{ or } B) = P(A) + P(\bar{A})P(B\,|\,\bar{A})$
  - $P(A \text{ or } B\text{, but not both}) = P(A)P(\bar{B}\,|\,A) + P(B)P(\bar{A}\,|\,B)$



Prosecutor's fallacy

- Your Covid test result is positive. The test sensitivity is 90% (i.e., P(positive test | infected) = 0.9), please start to panic
- You look at the NY Covid dashboard and learn that the current infection rate is 2%, how panicked should you be?
- Your Covid test result is negative. The test specificity is 80% (i.e., P(negative test | not infected) = 0.8. How relieved should you be?

Prosecutor's fallacy

	disease	no disease
test negative	0.10	0.80
test postive	0.90	0.20
prevalence	0.02	0.98

Prosecutor's fallacy

	disease	no disease
test negative	0.10	0.80
test postive	0.90	0.20
prevalence	0.02	0.98

	disease	no disease	marginal
test negative			
test postive			
marginal			100,000

Prosecutor's fallacy

	disease	no disease
test negative	0.10	0.80
test postive	0.90	0.20
prevalence	0.02	0.98

	disease	no disease	marginal
test negative			
test postive			
marginal	2,000	98,000	100,000

Prosecutor's fallacy

	disease	no disease
test negative	0.10	0.80
test postive	0.90	0.20
prevalence	0.02	0.98

	disease	no disease	marginal
test negative	200	78,400	
test postive	1,800	19,600	
marginal	2,000	98,000	100,000

Prosecutor’s fallacy

	disease	no disease
test negative	0.10	0.80
test postive	0.90	0.20
prevalence	0.02	0.98

	disease	no disease	marginal
test negative	200	78,400	78,600
test postive	1,800	19,600	21,400
marginal	2,000	98,000	100,000

P(infected I test positive) = P(infected & test positive)/P(test positive) = 1,800/21,400 = 8.4%

Or P(infected I test positive) = P(test positive I infected)P(infected)/P(test positive)
= 0.9 * 0.02 / .214 = 8.4%

Prosecutor’s fallacy

	disease	no disease
test negative	0.10	0.80
test postive	0.90	0.20
prevalence	0.02	0.98

	disease	no disease	marginal
test negative	200	78,400	78,600
test postive	1,800	19,600	21,400
marginal	2,000	98,000	100,000

P(not infected I test negative) = P(not infected & test negative)/P(test negative)
= 78,400/78,600 = 99.75%

Or P(not infected I test negative) = P(test negative I not infected)P(not infected)/P(test neg)
= 0.8 * 0.98 / .786 = 99.75%

Reply to prosecutor’s fallacy

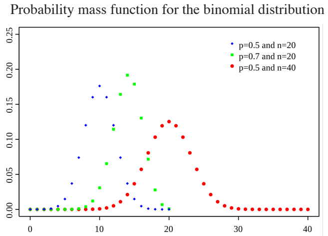
- Your Covid test result is positive. The test sensitivity is 90% (i.e., P(positive test I infected) = 0.9), please start to panic.
- You look at the NY Covid dashboard and learn that the current infection rate is 2%, how panicked should you be?
  - The disease rate is low and the false-positive rate cannot be ignored.
- Your Covid test result is negative. The test specificity is 80% (i.e., P(negative test I not infected) = 0.8. How relieved should you be?
  - Quite relieved, despite the poor specificity of the test

Coin flips and the Bernoulli distribution

- The letter "e" has a frequency of 12% in English words. How likely is its appearance in first names? Let’s ask 10 students in the room.
  - How many student’s names do you expect to have the letter “e”?
  - What’s the variance of this number?

The binomial distribution

Probability mass function for the binomial distribution



https://en.wikipedia.org/wiki/Binomial_distribution

- The letter “e” has a frequency of 12% in English words. How likely is its appearance in first names? Let’s ask 10 students in the room.
  - Bino(N,p) where N = 10, p = .12

Bernoulli distribution

X ~ Binomial(p, n=1)

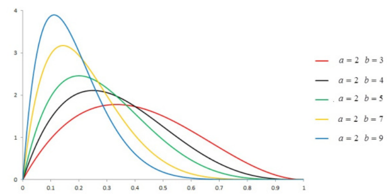
mean=p
var=$p(1-p)$.

- The letter “e” has a frequency of 12% in English words. How likely is its appearance in first names? Let’s ask 10 students in the room.
  - $X \sim$ Bernoulli(p).
  - $Y = X_1 + X_2 + \cdots + X_{10}$ (the sum of 10 independent random variables distributed the same as X)
  - E(Y)?
  - std(Y)? (thinking in terms of variance will be easier)

Posterior distribution after a coin flip

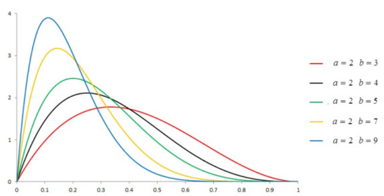
- The letter “e” has a frequency of 12% in English words. How likely is its appearance in first names? Let’s ask 10 students in the room.
  - How far is your estimate from our real experimental result?
  - What’s the posterior probability of letter “e” appearing in a first name?

Beta distribution



- Support (range of x): [0,1]
- Interpreted as the Bernoulli distribution parameter *p.*
- “Conjugate prior” of the Bernoulli distribution.
- Parameter: a,b => Beta(p|a,b)

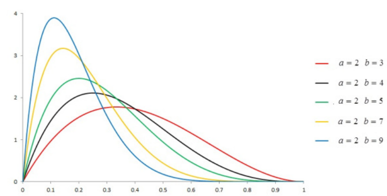
Beta distribution intuition



Beta(p|a,b):

- a-1 = number of “hits” already encountered
- b-1 = number of “misses” already encountered

Beta distribution update



posterior = likelihood * prior

Beta = Bernoulli*Beta

$$P(x\,|\,\text{data}) \propto P(\text{data}\,|\,x)P(x)$$

What if we add 5 more data points that “hit” (by just asking people I know who, by chance, have “e” in their first names)?

- How will the distribution move?

- What should be the mode of the posterior?

Conjugate priors

- Beta is the conjugate prior of Bernoulli.
- Normal is the conjugate prior of itself!