Lecture 1: Conditional Probability

Jeff Rouder

June, 2022

Consider the dreaded disease **Dipsidoodleitis**. Dipsidoodleitis is detected by a blood test which measures dipsidoodlamines in the blood. Values for people without the disease are normally distributed with $M = 70$ and $sd = 10$. Value for people with the disease are $M = 84$, $sd = 14$. Baserates are know as follows: 10% of people have the disease.

You happen to have a dipsidoodleamine value of 88. What is the probability that you are diseas free?

Several Ideas:

- Probability
- Random variables
- Parameters
- Model notation
- Discrete vs. Continuous Random Variables
- Probability mass vs. density functions
- Joint Distributions
- Marginal vs. Conditional Probability

Outcomes

Flip a Coin Once:

- H
- T

Flip a Coin Twice:

- (H,H)
- (H,T)
- (T,H)
- (T,T)

Sample Space

Collection (set) of all outcomes.

- Flip a Coin Once: X={H,T}
- Flip Twice: X={(H,H),(H,T),(T,H),(T,T)}

Events

Combinations of outcomes

The event of at least one head in two flips A={(H,H),(H,T),(T,H)}

Your Turn

1. List all events for the two-flip experiment. There are 16 of them. (Hint: The null set is an event.)
2. Which event is that the coin is a trick coin where both sides are heads or both sides are tails?

Probability

Two Domains:

- Rules
- Soul (yes, probability has a soul)

Probability Rules, Kolmogorov Axioms

1. $Pr(A) \geq 0$, where A is any event,
2. $Pr(\text{sample space}) = 1$, and
3. if $A \cap B = \emptyset$, then $Pr(A \cup B) = Pr(A) + Pr(B)$.

Your Turn

Suppose we flip a coin twice. Here are the probabilities on the outcomes. $P(\{(H, H)\}) = .5$, $P(\{(H, T)\}) = .2$, $P(\{(T, H)\}) = .15$. Use the Kolmogov axioms to find:

1. The probabilities on all events.
2. Probability of at least one tail
3. Probability of a trick coin

Probability

- Rules describe relative weight
- Works say on the weight of my body divided into overlapping parts
- But, we need more

Soul of Probability

Frequentist (Conventional):

- Probability is proportion in the long run
- Probability is a property of the natural world
    - Probability of coin is a property of the coin, much like its radius, weight, or alloy composition.
- Probability cannot be changed
- Probability is only knowable exactly in the long run
- Probability is not placed on unknown values that are fixed, say a parameter or a model.

Soul of Probability:

Bayesian

- Probability is degree of plausibility
- Probability is subjective
- Probability is our mental construct, not part of the world per se
- Probabilities may be placed on parameters, models, individual coin flips, etc.

Updating Probability?

There are three horses in a race. Each has a 1/3rd chance of winning. We are the house and we are going to let people make bets. Each bet costs $1, and if you pick the winning horse, the bet pays $2.90. We make 10 cents per bet. Now, right before betting opens, Horse 3 eats a bad apple, gets sick, and withdraws. How should we adjust probability and bets so that we make 10 cents per bet?

Bayesian Goal:

- Treat beliefs as probabilities
- Update beliefs rationally in light of data using rules of conditional probability

Joint Probabilty

```
##          Men Women
## Afr Am 0.010 0.016
## Asian  0.109 0.202
## Latinx 0.057 0.083
## White  0.098 0.217
## Other  0.088 0.121
```

- Probability of cross, “and”, intersection
- All cells sum to 1.0

Law of Total Probability

Let X be a sample space. Let $B_1, B_2, \ldots, B_N$ be a collection of disjoint sets ($B_i \cap B_j = \emptyset$) that cover the sample sample $X \subseteq \cup B_i$. Then,

$$Pr(A) = Pr(A \cap B_1) + Pr(A \cap B_2) + \ldots = \sum_i P(A \cap B_i).$$

Law of Total Probability Example

```
##           Men Women
## Afr Am 0.010 0.016
## Asian  0.109 0.202
## Latinx 0.057 0.083
## White  0.098 0.217
## Other  0.088 0.121
```

What is the probability White?

$P(W) = P(W \cap M) + P(W \cap F) = .098 + .217 = .315.$

Your Turn

```
##            Men Women
## Afr Am 0.010 0.016
## Asian  0.109 0.202
## Latinx 0.057 0.083
## White  0.098 0.217
## Other  0.088 0.121
```

What is the probability that a given person is a woman according to the table?

Marginal Probability

Probability of an event calculated from the joint through Law of Total Probability

Your Turn

```
##          Men Women
## Afr Am 0.010 0.016
## Asian  0.109 0.202
## Latinx 0.057 0.083
## White  0.098 0.217
## Other  0.088 0.121
```

- Use Law of Total Probability to Derive Marginal Probability Distrubiton of Ethnicity?

Your Turn (Answer)

```
margEth=apply(joint,1,sum)
print(round(margEth,3))
```

```
## Afr Am  Asian Latinx  White  Other
##  0.026  0.310  0.140  0.315  0.209
```

```
sum(margEth)
```

```
## [1] 1
```

Definition of Conditional Probability

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

Your Turn

```
##           Men Women
## Afr Am 0.010 0.016
## Asian  0.109 0.202
## Latinx 0.057 0.083
## White  0.098 0.217
## Other  0.088 0.121
```

- What is the $P(AfAm|F)$?
- What is $P(F|AfAm)$?

Your Turn

```
##           Men Women
## Afr Am 0.010 0.016
## Asian  0.109 0.202
## Latinx 0.057 0.083
## White  0.098 0.217
## Other  0.088 0.121
```

- What is the conditional distribution of gender given ethnicity

Your Turn (Answer)

```
tab=cbind(joint,margEth)
print(round(tab,3))
```

```
##           Men Women margEth
## Afr Am 0.010 0.016   0.026
## Asian  0.109 0.202   0.310
## Latinx 0.057 0.083   0.140
## White  0.098 0.217   0.315
## Other  0.088 0.121   0.209
```

Your Turn (Answer)

```
print(round(joint/margEth,3))
```

```
##           Men Women
## Afr Am 0.400 0.600
## Asian  0.350 0.650
## Latinx 0.407 0.593
## White  0.311 0.689
## Other  0.420 0.580
```

Your Turn

```
##           Men Women
## Afr Am 0.010 0.016
## Asian  0.109 0.202
## Latinx 0.057 0.083
## White  0.098 0.217
## Other  0.088 0.121
```

- What is the conditional distribution of ethnicity given gender?

Your Turn

- Is $P(A)$ alwas larger than $P(A \cap B)$?
- Is $P(A|B)$ always larger than $P(A)$?

Note:

Since

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

We can re-express intersection as product of conditional and marginal:

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

Law of Total Probability (conditional format)

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$$

We will use this form a lot soon

Law of Conditional Probability

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

Random Variables and Random-Variable Notation:

Random variables map from events to numbers. Example, let Y be the number of heads for two flips.

- What are the possible values of Y?
- What are the probabilities on these values if the coin is fair?

Random Variable Notation

Let Y be a random variable that denotes the number of heads. Suppose we specify Y as

$$Y \sim \text{Binomial}(\theta, n)$$

where θ is the parameter of interest and n is the number of trials.

- Discrete Distribution, $y = 0, 1, 2, \ldots, n$.
- total number of successes on n trials
- one parameter, θ, probability success on a single trial

YOUR TURN:

For $Y \sim \text{Binomial}(\theta, 2)$ (two flips with probability θ)

- How do outcomes map to values of Y?
- What are the probabilities over the outcomes in natural space?
- What are the probabilities over values of Y?

Bernoulli

$Y \sim \text{Bernoulli}(\theta)$

- single coin flip rather than many
- same as $Y \sim \text{Binomial}(\theta, 1)$
- The Bernoulli is a discrete distribution, it lives on two points: $y = 0$ and $y = 1$.
- The Bernoulli describes dichotomous events, like coin flips.
- The Bernoulli has a single parameter, the probability of success.

Probabilities for Bernoulli

$$Pr(Y = y \mid \theta) = \theta^y (1-\theta)^{(1-y)}, \quad y = 0, 1$$

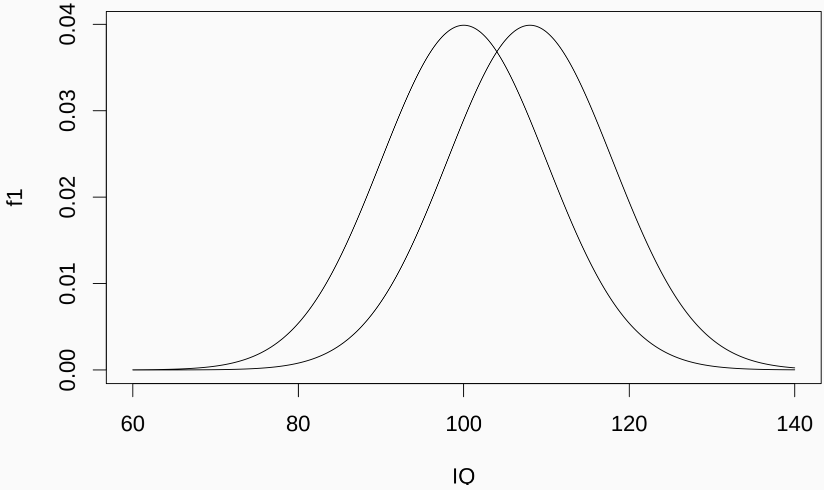
Normal Distribution

Everyone knows the normal. Here it is for say IQ across a population

$$Y \sim \text{Normal}(\mu, \sigma^2)$$

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

Normal Distribution



Normal Distribution

1. The normal is a continuous distribution, it lives on the interval $(-\infty, \infty)$
2. The normal has a location parameter μ and a variance parameter σ^2
3. The normal is used because of its convenience.
4. Effects shift the distribution.

Working the Problem

Consider the dreaded disease **Dipsidoodleitis**. Dipsidoodleitis is detected by a blood test which measures dipsidoodlamines in the blood. Values for people without the disease are normally distributed with $M = 70$ and $sd = 10$. Value for people with the disease are $M = 84$, $sd = 14$. Baserates are know as follows: 10% of people have the disease.

You happen to have a dipsidoodleamine value of 88. What is the probability that you are diseas free?

Notation Step: Let Let Y denote be a random variable that denotes the dipsidoodlamine level. Let D be a random variable that denotes whether an individual has the disease or not. ($D = 0, 1$ for no disease and disease, respectively).

Model Step What kind of model statements can we make? Are they conditional or marginal statements?

- $D \sim \text{Bernoulli}(.1)$
- $Y|D = 0 \sim \text{Normal}(70, 10^2)$
- $Y|D = 1 \sim \text{Normal}(84, 14^2)$

Working The Problem

We want to know $Pr(D = 0|Y = 86)$

Apply Law of Conditional Probability

$$P(D = 0|Y = 88) = \frac{P(Y = 88|D = 0)P(D = 0)}{P(Y = 88)}$$

- Numerator is plug=and-chug
  - $P(Y = 88|D = 0) = f(y = 88, \mu = 70, \sigma^2 = 10^2) = .0079$
  - $P(D = 0) = .9$
  - $P(Y = 88|D = 0)P(D = 0) = .0071$

Working The Problem

- Denominator is $P(Y = 88)$
- Use Law of Total Probability
- $P(Y = 88) = P(Y = 88|D = 0)P(D = 0) + P(Y = 88|D = 1)P(D = 1)$
- $P(Y|D = 1) = f(y = 88, \mu = 84, \sigma^2 = 14^2) = .0274$
- $P(D = 1) = .0027$
- $P(Y = 88) = .0071 + .0027 = .0098$

Working The Problem

- $P(D=0|Y=88) = \frac{P(Y=88|D=0)P(D=0)}{P(Y=88)} = \frac{.0071}{.0098} = .72$

Your Turn

- Plot the marginal density of Y, that is $Pr(Y = y)$ for all values of Y.
- Is it a normal? Why or Why not.

Your Turn

- Medical tests have a "false positive" and "false negative" rate.
- These are always reposted as conditionals, that is false positive is the probability of a positive result given that the tested-attribute is not present.
- Do you think it is better to report these as conditions or joints? A joint probability would be Pr(positive and the tested-attribute is not present)?