# Lecture 9: Probability

COSC 480 Data Science, Spring 2017

Michael Hay

# The *Ferengi*

An alien species in Star Trek notorious for extreme sexism

http://en.memory-alpha.org/wiki/The_Magnificent_Ferengi_(episode)
Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

# Tech interview question

- Ferengi want boys, so every family keeps on having children until a boy is born
    - If the newborn is a girl, have another child
    - If the newborn is a boy, stop.

***Can their strategy influence the composition of their population?***

# Probability to the rescue

- We need them to quantify uncertainty so that we can make better decisions
- Human intuitions about uncertainty are sometimes ***terrible***!


Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

# Probability

**Sample space and events** A sample space $\Omega$ is set of all possible outcomes (elementary events). An event $E$ is a subset of possible outcomes. Example: rolling a die, $\Omega = \{1, 2, 3, 4, 5, 6\}$ and an example of $E$ is rolling an even number.

**Probability** A probability function $P$ is a function mapping events to real numbers that satisfies these conditions:

1. for any event $E$, $0 \leq P(E) \leq 1$
2. $P(\Omega) = 1$ (one elementary event must happen)
3. for any set of disjoint event $E_1, E_2, \ldots, E_n$, $P\left(E_1 \cup E_2 \cup \cdots \cup E_n\right) = \sum_{i=1}^{n} P(E_i)$

Example on board: rolling a die, E=roll > 3

# Sample space for *Ferengi* family?

| Outcome |
|---|
| boy |
| girl, boy |
| girl, girl, boy |
| girl, girl, girl, boy |
| … |

# Useful properties

For any two events, $P\left(E_1 \cup E_2\right) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

**Union bound** $P\left(E_1 \cup E_2 \cup \cdots \cup E_n\right) \leq \sum_{i=1}^{n} P(E_i)$

Example on board: $E_1$ = roll is > 3, $E_3$ = roll is even

# Conditional probability

**Conditional probability** The conditional probability that event $A$ occurs given that $B$ occurs is

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

Of course, for this to make sense, it must be that $P(B) > 0$.

Example on board: $P(E_3 = \text{roll is even} \mid E_1 = \text{roll is} > 3) = ??$

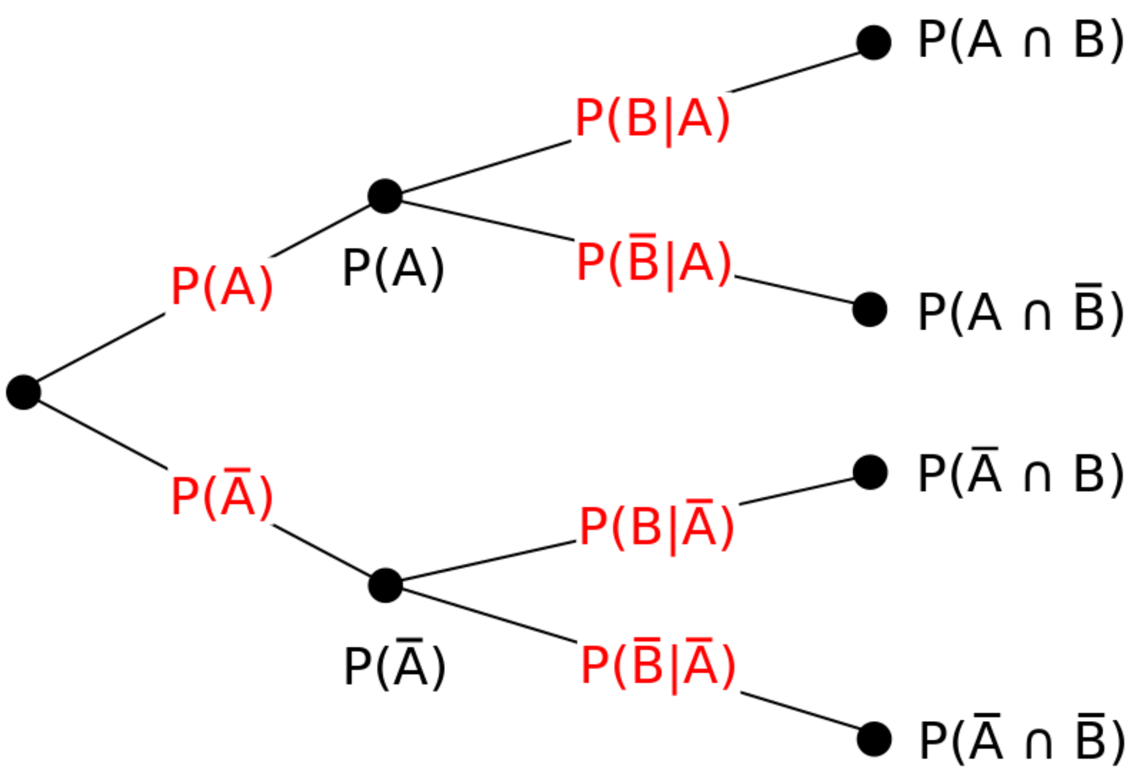
# Exercise

**Instructions:** ~1 minute to think/answer on your own; then discuss with neighbors; then I will call on one of you

- Given that a Ferengi family has at most 3 children, what is the probability that the first born child is a boy?

# Chain rule

**Chain rule** For any two events $A$ and $B$, $P(A \cap B) = P(B|A)P(A)$. It's also true that $P(A \cap B) = P(A|B)P(B)$. (Follows from definition of conditional probability and rearranging terms.)




http://en.wikipedia.org/wiki/File:Probability_tree_diagram.svg

# Independent events

**Independent events** Events $A$ and $B$ are independent if $P(A \cap B) = P(A)P(B)$. Intuitively events are independent if the occurrence of one event does not affect the probability that the other occurs. Independence implies $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

- Example: x and y are two newborns. Events "x is a girl" and "y is a boy" are independent. P(x is a girl and y is boy) = 1/2 x 1/2 = 1/4.

# Random variables

- A random variable assigns a numerical value to each possible outcome
- Example: possible outcomes in a Ferengi family, and random variables B (# boys) and G (# girls)

| Prob | Outcome | *B* | *G* |
|---|---|---|---|
| 1/2 | boy | 1 | 0 |
| $(1/2)^2$ | girl, boy | 1 | 1 |
| $(1/2)^3$ | girl, girl, boy | 1 | 2 |
| $(1/2)^4$ | girl, girl, girl, boy | 1 | 3 |
| … | … | … | … |

*What is probability of each outcome?*

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

# Probability distributions

- For a discrete random variable, we can specify its distribution by a *probability mass function* (pmf) that assigns each value a probability.

**Random variable** Formally, a random variable $X$ is a function that maps events to numbers. We can use the notation $X(\omega) = x$ to mean that event $\omega$ is mapped to $x$. The probability that $X$ takes on the value $x$ is simply: $P(X = x) = \sum_{\omega \in \Omega : X(\omega) = x} P(\omega)$.

| Prob | Outcome | ***B*** | ***G*** |
|---|---|---|---|
| 1/2 | boy | 1 | 0 |
| $(1/2)^2$ | girl, boy | 1 | 1 |
| $(1/2)^3$ | girl, girl, boy | 1 | 2 |
| $(1/2)^4$ | girl, girl, girl, boy | 1 | 3 |
| … | … | … | … |

| ***G*** | Prob |
|---|---|
| 0 | 1/2 |
| 1 | $(1/2)^2$ |
| 2 | $(1/2)^3$ |
| 3 | $(1/2)^4$ |
| | … |

| ***B*** | Prob |
|---|---|
| 1 | 1 |

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

# Exercise

**Instructions:** ~1 minute to think/answer on your own; then discuss with neighbors; then I will call on one of you

- Let *F* be a random variable equal to the fraction of boys in a Ferengi family.
- What is the probability mass function of F? (I.e., for each value of F, what is P(F)?)

# Expectation

- For discrete random variable, expected value E[X] is the average of all possible X values, weighted by probability.

**Expected value** Suppose $X$ takes on $n$ possible values $x_1, \ldots, x_n$. The expected value of a random variable is $E[X] = \sum_{i=1}^{n} x_i P(X = x_i)$.

- Intuition: if sample (many) X's and take their mean, it should be close to E[X].

- Example: X = die roll, and die is *biased:* P(X=6) = 1/2 whereas P(X=i) = 1/10 for i≠6. What is E[X]?

# Exercise

**Instructions:** ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

- Back to the Ferengi…
- Expected number of boys E[B]?
- Expected number of girls E[G]?

| Prob | *B* |
|---|---|
| 1 | 1 |

| Prob | *G* |
|---|---|
| 1/2 | 0 |
| $(1/2)^2$ | 1 |
| $(1/2)^3$ | 2 |
| $(1/2)^4$ | 3 |
| … | |

What about expected size of a family?

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

# Useful properties

**Marginalization / Total law of probability** Let $X$ and $Y$ be two random variables that take on the values $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ respectively.

$$P(X = x_i) = \sum_{j=1}^{m} P(X = x_i \cap Y = y_j)$$

**Linearity of Expectations** Given random variables $X$ and $Y$, $E[X + Y] = E[X] + E[Y]$.

# Exercise

**Instructions:** ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

- Back to the Ferengi…
- Expected number of kids in family?
- Expected % of boys in a family?
  - Hint: quite a bit more than 50%
  - E[ B/(B+G) ] does ***not*** equal E[B] / E[B+G]

| Prob | *B* |
|---|---|
| 1 | 1 |

| Prob | *G* |
|---|---|
| 1/2 | 0 |
| $(1/2)^2$ | 1 |
| $(1/2)^3$ | 2 |
| $(1/2)^4$ | 3 |
| … | |

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

# Another probability puzzle

- The probability that a woman 40 to 50 years old has breast cancer is 0.8%
- If a woman has breast cancer, the probability is 90% that she will have a positive mammogram
- If a woman does not have breast cancer, the probability is 7% that she will still have a positive mammogram
- Imagine a woman who has a positive mammogram. *What is the probability that she actually has breast cancer?*

http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

# What does the doc say?

- Reaction from a department chief at a German university teaching hospital with more than 30 years of experience:

  *"[He] was visibly nervous while trying to figure out what he would tell the woman. After mulling the numbers over, he finally estimated the woman's probability of having breast cancer, given that she has a positive mammogram, to be* ***90 percent****. Nervously, he added, 'Oh, what nonsense. I can't do this. You should test my daughter; she is studying medicine.'"*

http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

# Bayes' rule

**Bayes' rule** This simple but powerful rule follows from the definition of conditional probability and an application of the chain rule.

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

Let $\neg A$ denote the event "$A$ did not occur," we can also write it as

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

A particular useful application is when event $A$ is some hidden cause and event $B$ is some observable symptom. Given the test result is positive ($B$), what is the probability the patient has the disease ($A$)?

# Another probability puzzle

- The probability that a woman 40 to 50 years old has breast cancer is 0.8%

  P(cancer) = 0.008

- If a woman has breast cancer, the probability is 90% that she will have a positive mammogram

  P(pos|cancer) = 0.9

- If a woman does not have breast cancer, the probability is 7% that she will still have a positive mammogram

  P(pos|¬cancer) = 0.07

- Imagine a woman who has a positive mammogram. *What is the probability that she actually has breast cancer?*

  P(cancer|pos) = ???

http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University

# As for American doctors

- … 95 out of 100 estimated the woman's probability of having breast cancer to be somewhere around ***75 percent***.

http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

Credit: slide adapted with permission from Ashwin Machanavajjhalla and Jun Yang, Duke University