Sta 111 - Summer II 2017
Probability and Statistical Inference

2. Probability

Lu Wang

Duke University, Department of Statistical Science

July 4, 2017

Outline

- An *event* or a *random process* is a situation in which we know what outcomes could happen, but we don't know which particular outcome will happen.
  - Examples: coin tosses, die rolls, iTunes shuffle, whether the stock market goes up or down tomorrow, etc.
- There are several possible interpretations of *probability* but they (almost) completely agree on the mathematical rule probability must follow.
  - $P(A)$ = Probability of event A
  - $0 \leq P(A) \leq 1$
- *Frequentist interpretation:*
  - The probability of an outcome is the proportion of times the outcome would occur if we observed the event an infinite number of times.
- *Bayesian interpretation:*
  - A Bayesian interprets probability as a subjective degree of belief: For the same event, two separate people could have different viewpoints and so assign different probabilities.

Law of large numbers states that as more observations are collected, the proportion of occurrences with a particular outcome, $\hat{p}_n$, converges to the probability of that outcome, p.

When tossing a *fair* coin, if heads come up on each of the first 10 tosses, what do you think the chance is that another head will come up on the next toss? 0.5, less than 0.5, or more than 0.5?

$$\underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{?}$$

- The probability is still 0.5, or there is still a 50% chance that another head will come up on the next toss.

$$P(H \text{ on } 11^{th} \text{ toss}) = P(T \text{ on } 11^{th} \text{ toss}) = 0.5$$

- The common misunderstanding of the LLN is that random processes are supposed to compensate for whatever happened in the past; this is just not true and is also called *gambler's fallacy*.

Disjoint and non-disjoint outcomes

Disjoint (mutually exclusive) outcomes: Cannot happen at the same time.

- The outcome of a single coin toss cannot be a head and a tail.
- A student both cannot fail and pass a class.
- A single card drawn from a deck cannot be an ace and a queen.

Non-disjoint outcomes: Can happen at the same time.

- A student can get an A in Stats and A in Econ in the same semester.

General addition rule

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

- For disjoint events $P(A \text{ and } B) = 0$, so the above formula simplifies to $P(A \text{ or } B) = P(A) + P(B)$.

Union of non-disjoint events

What is the probability of drawing a jack or a red card from a well shuffled full deck?

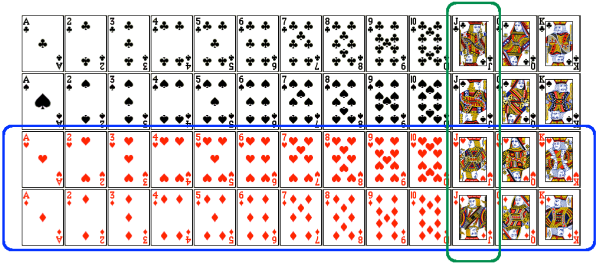

Figure from *http://www.milefoot.com/math/discrete/counting/cardfreq.htm*.

Practice

What is the probability that a randomly sampled student thinks marijuana should be legalized or they agree with their parents' political views?

	Share Parents' Politics		
Legalize MJ	No	Yes	Total
No	11	40	51
Yes	36	78	114
Total	47	118	165

(a) $\frac{40+36-78}{165}$

(b) $\frac{114+118-78}{165}$

(c) $\frac{78}{165}$

(d) $\frac{78}{188}$

(e) $\frac{11}{47}$

A *probability distribution* lists all possible outcomes and the probabilities with which they occur.

- The probability distribution for the gender of one kid:

Event	Male	Female
Probability	0.5	0.5

- Rules for probability distributions:
  1. The outcomes listed must be disjoint
  2. Each probability must be between 0 and 1
  3. The probabilities must total 1

- The probability distribution for the genders of two kids:

Sample space is the collection of all possible outcomes of a trial.

- A couple has one kid, what is the sample space for the gender of this kid? $S = \{M, F\}$
- A couple has two kids, what is the sample space for the gender of these kids?

Complementary events are two mutually exclusive events whose probabilities add up to 1.

- A couple has one kid. If we know that the kid is not a boy, what is gender of this kid? $\{\cancel{M}, F\} \rightarrow$ Boy and girl are *complementary* outcomes.
- A couple has two kids, if we know that they are not both girls, what are the possible gender combinations for these kids?

Two events are *independent* if knowing the outcome of one provides no useful information about the outcome of the other.

- Outcomes of two tosses of a coin are independent.
  - Knowing that the coin landed on a head on the first toss does not provide any useful information for determining what the coin will land on in the second toss.
- Outcomes of two draws from a deck of cards (without replacement) are dependent.
  - Knowing that the first card drawn from a deck is an ace does provide useful information for determining the probability of drawing an ace in the second draw.

Practice

Between January 9-12, 2013, SurveyUSA interviewed a random sample of 500 NC residents asking them whether they think widespread gun ownership protects law abiding citizens from crime, or makes society more dangerous. 58% of all respondents said it protects citizens. 67% of White respondents, 28% of Black respondents, and 64% of Hispanic respondents shared this view. Which of the below is true?
Opinion on gun ownership and race ethnicity are most likely

(a) complementary

(b) mutually exclusive

(c) independent

(d) dependent

(e) disjoint

http://www.surveyusa.com/client/PollReport.aspx?g=a5f460ef-bba9-484b-8579-1101ea26421b

Checking for independence

If P(A occurs, given that B is true) = $P(A \mid B) = P(A)$, then A and B are independent.

P(protects citizens) = 0.58

P(randomly selected NC resident says gun ownership protects citizens, given that the resident is white) =
P(protects citizens | White) = 0.67

P(protects citizens | Black) = 0.28

P(protects citizens | Hispanic) = 0.64

P(protects citizens) varies by race/ethnicity, therefore opinion on gun ownership and race ethnicity are most likely dependent.

Product rule for independent events

$$P(A \text{ and } B) = P(A) \times P(B)$$

Or more generally, $P(A_1 \text{ and } \cdots \text{ and } A_k) = P(A_1) \times \cdots \times P(A_k)$

You toss a coin twice, what is the probability of getting two tails in a row?

$$P(\text{T on the first toss}) \times P(\text{T on the second toss}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Disjoint vs. independent

- *Disjoint (mutually exclusive) outcomes* cannot happen at the same time
  - A voter cannot register as a Democrat and a Republican at the same time
  - For disjoint A and B: $P(A \text{ and } B) = 0$

- If A and B are *independent events*, having information on A does not tell us anything about B (and vice versa)
  - If A and B are independent:
    - $P(A \mid B) = P(A)$
    - $P(A \text{ and } B) = P(A) \times P(B)$

Graded questions:

- Ch 2: 2.8, 2.18, 2.22
- Ch 3: 3.4, 3.12, 3.36

Practice questions:

- Defining probability: 2.1, 2.3, 2.5, 2.7, 2.13
- Conditional probability: 2.15, 2.19, 2.21, 2.23
- Normal distribution: 3.3, 3.5, 3.9, 3.11, 3.17
- Binomial distribution: 3.25, 3.27, 3.29, 3.33

Note: You can use the back of the book to check your work on the practice questions.