Probability

Ryan Miller

Outline

1. Basic definitions
   - random process, sample space, events
2. Probability laws
   - disjoint events, compliment rule, addition rule, independence, multiplication rule

Introduction

- *Statistical inference*, the process of using sample data to reach a conclusion, inherently involves *uncertainty*
  - Which cases from the population ended up in the sample data?
  - Which cases ended up in the treatment and control groups?
  - Could the data generation process have unfolded differently?

Basic Definitions

- A **random process** describes any phenomenon whose *outcome* cannot be predicted with certainty
- A **sample space** refers to the collection of possible outcomes of a random process
- An **event** describes the realization of one (or more) outcomes from a random process

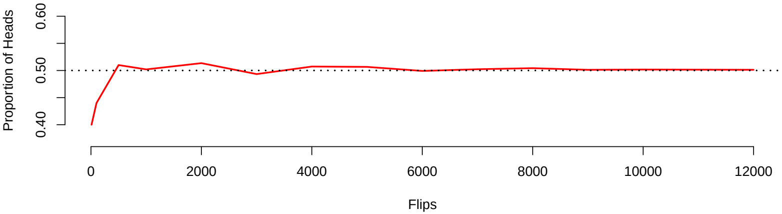
Process	Space	Event
Flipping a Coin	{H,T}	Seeing H
Rolling a 6-sided Die	{1,2,3,4,5,6}	Seeing an odd number
Person takes Vaccine	{Disease, No Disease}	No Disease

Probabilty

- **Probability** describes the long-run relative frequency of an event over infinitely many repetitions of a random process

Probabilty

- **Probability** describes the long-run relative frequency of an event over infinitely many repetitions of a random process
- The theoretical justification for this definition is the **Law of Large Numbers**, which states that the proportion of times an outcome is observed will converge to it's probability
  - For example, when flipping a fair coin we'll say $P(\text{Heads}) = 0.5$ because the proportion of heads will converge to 0.5 if the random process is repeated many times



Disjoint events

- Two events are **disjoint** or *mutually exclusive* if they cannot both occur simultaneously
  - If we flip a single coin, "Heads" and "Tails" are disjoint
  - If we roll a six-sided die, "Odd" and "≥ 4" are *not* disjoint

Disjoint events

- Two events are **disjoint** or *mutually exclusive* if they cannot both occur simultaneously
  - If we flip a single coin, "Heads" and "Tails" are disjoint
  - If we roll a six-sided die, "Odd" and "≥ 4" are *not* disjoint
- We can express the probability of disjoint events using the notation: $P(A_1 \cap A_2) = 0$
  - In words, the probability of observing both A_1 *and* A_2 simultaneously is zero

Probability distributions

- **Probability distributions** are used to map disjoint events to probabilities
  - Here is an example for the sum of two rolls of a six-sided die:

Event	2	3	4	5	6	7	8	9	10	11	12
Probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Probability distributions

- **Probability distributions** are used to map disjoint events to probabilities
  - Here is an example for the sum of two rolls of a six-sided die:

Event	2	3	4	5	6	7	8	9	10	11	12
Probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

- A *valid* probability distribution must satisfy *all* of the following:
  - All events must be *disjoint*
  - Each event must have a probability ≥ 0
  - The probability of the entire set of events (sample space) sums to exactly 1

Addition rule (disjoint events)

If two events are disjoint, the probability that *either* event occurs is given by:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

For example, for a single coin flip:

$$P(\text{Heads} \cup \text{Tails}) = P(\text{Heads}) + P(\text{Tails}) = 0.5 + 0.5 = 1$$

Addition rule (general)

If the events are *not* disjoint, the probability that either event occurs is given by:

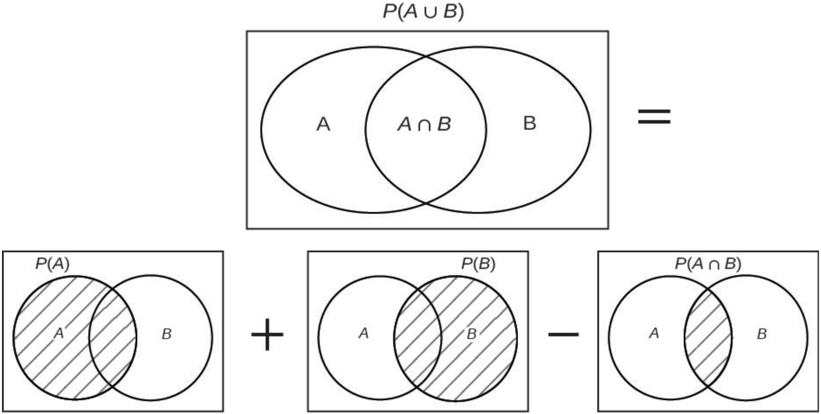
$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

For example, consider a single roll of a six-sided die:

$$P(>3 \cup \text{Even}) = P(>3) + P(\text{Even}) = 3/6 + 3/6 - 2/6 = 0.667$$

Venn diagrams

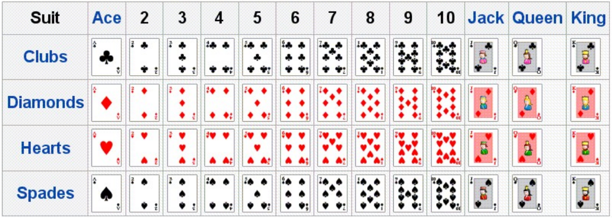
Venn diagrams provide a useful heuristic for understanding the addition rule:



Practice

A standard deck of playing cards contains 52 cards that belong to 4 different suits:

Example set of 52 poker playing cards



For the random process of drawing a single card, find the following probabilities:

1) $P(\text{Heart} \cap \text{Diamond})$
2) $P(\text{Heart} \cup \text{Diamond})$
3) $P(\text{Heart} \cup \text{Even Number})$

Practice (solution)

1) $P(\text{Heart} \cap \text{Diamond}) = 0$
2) $P(\text{Heart} \cup \text{Diamond}) = 13/52 + 13/52 = 0.5$
3) $P(\text{Heart} \cup \text{Even Number}) = 13/52 + 20/52 - 5/52 = 0.538$

Complement rule

- For any event, A, we define A^C as the **complement** of A
  - A^C represents all outcomes in the sample space that *do not* belong to A
  - For example, if A is seeing a 6 when rolling a six-sided die, A^C is rolling either a 1, 2, 3, 4, or 5

Complement rule

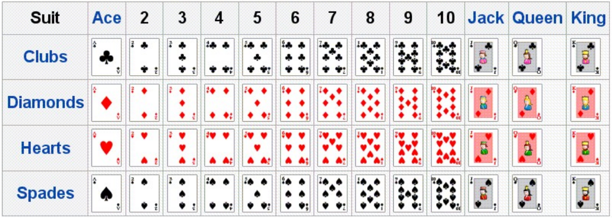
- For any event, A, we define A^C as the **complement** of A
  - A^C represents all outcomes in the sample space that *do not* belong to A
  - For example, if A is seeing a 6 when rolling a six-sided die, A^C is rolling either a 1, 2, 3, 4, or 5
- The **complement rule** states: $P(A^C) = 1 - P(A)$
  - For example, when rolling a six-sided die:

$$P(1 \cup 2 \cup 3 \cup 4 \cup 5) = 1 - P(6) = 1 - 1/6 = 0.1667$$

Practice

A standard deck of playing cards contains 52 cards that belong to 4 different suits:

Example set of 52 poker playing cards



For the random process of drawing a single card, find the following probabilities:

1) $P(\text{Heart}^C)$
2) $P(\text{Heart}^C \cup \text{Even Number})$

Practice (solution)

1) $P(\text{Heart}^C) = 1 - 13/52 = 0.75$
2) $P(\text{Heart}^C \cup \text{Even Number}) =$
   $P(\text{Heart}^C) + P(\text{Even Number}) - P(\text{Heart}^C \cap \text{Even Number}) =$
   $39/52 + 20/52 - 15/52 = 0.846$

Multiplication rule (independence)

- Two random processes are **independent** if knowing the outcome of one process provides no insight into the outcome of the other
  - Flipping a fair coin and rolling a six-sided die are *independent* random processes
  - A single student receiving a calculus test score and a physics test score are not

Multiplication rule (independence)

- Two random processes are **independent** if knowing the outcome of one process provides no insight into the outcome of the other
  - Flipping a fair coin and rolling a six-sided die are *independent* random processes
  - A single student receiving a calculus test score and a physics test score are not
- If events A_1 and A_2 arise from independent random processes, the **multiplication rule** states:

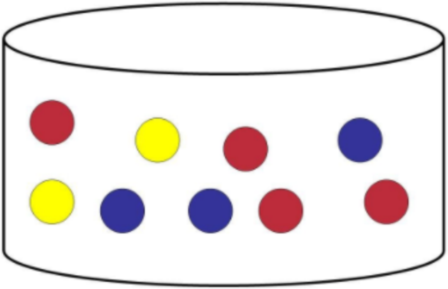
$$P(A_1 \cap A_2) = P(A_1) * P(A_2)$$

Independent vs. disjoint events

- Disjoint events generally are *never* independent (aside from the trivial case where one event has zero probability)
  - If A_1 and A_2 are *disjoint*, then $P(A_1 \cap A_2) = 0$
  - If A_1 and A_2 are *independent*, then $P(A_1 \cap A_2) = P(A_1) * P(A_2)$

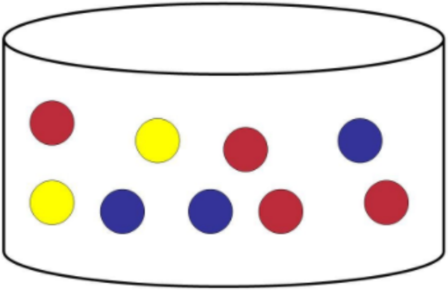
Dependent events

- When probability of an event can be influenced by another event, we must use **conditional probability**
  - A simple example is sampling from a small population
  - Let A_1 denote the event of randomly drawing a yellow ball, clearly $P(A_1) = 2/9$



Sampling from a small population

- Now let A_i represent the i^{th} draw (without replacement) from the urn being a yellow ball
  - What is $P(A_1 \text{ and } A_2)$?
  - What about $P(A_1 \text{ and } A_2 \text{ and } A_3)$?



An incorrect approach

- The multiplication rule seemingly suggests:
  - $P(A_1 \text{ and } A_2) = P(A_1) * P(A_2) = \frac{2}{9} * \frac{2}{9} \approx 0.05$
  - $P(A_1 \text{ and } A_2 \text{ and } A_3) = P(A_1) * P(A_2) * P(A_3) = (\frac{2}{9})^3 \approx 0.01$

An incorrect approach

- The multiplication rule seemingly suggests:
  - $P(A_1 \text{ and } A_2) = P(A_1) * P(A_2) = \frac{2}{9} * \frac{2}{9} \approx 0.05$
  - $P(A_1 \text{ and } A_2 \text{ and } A_3) = P(A_1) * P(A_2) * P(A_3) = (\frac{2}{9})^3 \approx 0.01$
- However, observing a yellow ball on the first draw alters the chances of getting a yellow ball on the second or third draw
  - This is most obviously evidenced by the fact that drawing 3 yellow balls is impossible!
  - Clearly the multiplication rule needs to be adjusted to work for dependent events

Multiplication Rule (general)

- In our example, it's easy to see $P(A_1) = 2/9$ and $P(A_2|A_1) = 1/8$, as well as $P(A_3|A_1, A_2) = 0$
  - These examples illustrate the concept of *conditional probability*, and they lead us to the **general multiplication rule**:

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

Multiplication Rule (general)

- In our example, it's easy to see $P(A_1) = 2/9$ and $P(A_2|A_1) = 1/8$, as well as $P(A_3|A_1, A_2) = 0$
  - These examples illustrate the concept of *conditional probability*, and they lead us to the **general multiplication rule**:

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

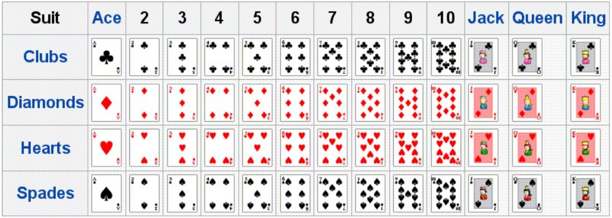
    It's often useful to rearrange this equation:

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

Practice

A standard deck of playing cards contains 52 cards that belong to 4 different suits:

Example set of 52 poker playing cards



For the random process of drawing a single card, find the following probabilities:

1) $P(\text{Heart}|\text{Red})$
2) $P(\text{Ten}| \geq \text{Seven})$

Practice (solution)

1) $P(\text{Heart}|\text{Red}) = P(\text{Heart} \cap \text{Red})/P(\text{Red}) = \frac{13/52}{26/52} = 0.5$
2) $P(\text{Ten}| \geq \text{Seven}) = P(\text{Ten} \cap \geq \text{Seven})/P(\geq \text{Seven}) = \frac{4/52}{16/52} = 0.25$

Estimating probabilities from contingency tables

- Recall that *contingency tables* are a method used to relate two categorical variables
  - We saw that *row proportions* or *column proportions* were particularly useful in describing potential associations

Estimating probabilities from contingency tables

- Recall that *contingency tables* are a method used to relate two categorical variables
  - We saw that *row proportions* or *column proportions* were particularly useful in describing potential associations
- Since we've defined probability as a long run frequency, it makes sense to use proportions observed in a sample as our *best estimate* of certain probabilities

Estimating probabilities from contingency tables

- Recall that *contingency tables* are a method used to relate two categorical variables
  - We saw that *row proportions* or *column proportions* were particularly useful in describing potential associations
- Since we've defined probability as a long run frequency, it makes sense to use proportions observed in a sample as our *best estimate* of certain probabilities

	death	not
black	38	142
white	46	152

- In the Florida Death Penalty study (the table shown above), we might estimate $P(\text{Death}|\text{WhiteOffender}) = 46/(152 + 46) = 0.232$

Marginal, joint, and conditional probabilities

Contingency tables can help us understand three distinct types of probabilities used in scenarios involving two variables (ie: two random processes)

	death	not
black	38	142
white	46	152

1) A **marginal probability** only considers a single variable, for example: $P(\text{Death}) = 84/378 = 0.222$

Marginal, joint, and conditional probabilities

Contingency tables can help us understand three distinct types of probabilities used in scenarios involving two variables (ie: two random processes)

	death	not
black	38	142
white	46	152

1) A **marginal probability** only considers a single variable, for example: $P(\text{Death}) = 84/378 = 0.222$
2) A **joint probability** simultaneous considers both variables, for example: $P(\text{Death} \cap \text{WhiteOffender}) = 46/378 = 0.165$

Marginal, joint, and conditional probabilities

Contingency tables can help us understand three distinct types of probabilities used in scenarios involving two variables (ie: two random processes)

	death	not
black	38	142
white	46	152

1) A **marginal probability** only considers a single variable, for example: $P(\text{Death}) = 84/378 = 0.222$
2) A **joint probability** simultaneous considers both variables, for example: $P(\text{Death} \cap \text{WhiteOffender}) = 46/378 = 0.165$
3) A **conditional probability** considers one variable, under the assumption that the other has already been observed, for example: $P(\text{Death}|\text{WhiteOffender}) = 46/198 = 0.232$

Practice

The table below describes survival of residents of Boston, MA in 1721 that were exposed to smallpox. Some of these residents had been inoculated using a controlled strain of smallpox:

	Lived	Died	Total
Inoculated	238	6	244
Not Inoculated	5136	884	6020
Total	5374	890	6264

State whether each of the following is a marginal, joint, or conditional probability, then estimate it using the data presented above:

1) That a resident died from their exposure
2) That a resident died given they'd been inoculated
3) That a resident had been inoculated given they've died
4) That a randomly chosen resident was both inoculated and ended up dying

Practice (solution)

1) $P(\text{Died}) = 890/6264$, marginal
2) $P(\text{Died}|\text{Inoculated}) = 6/244$, conditional
3) $P(\text{Inoculated}|\text{Died}) = 6/890$, conditional
4) $P(\text{Inoculated} \cap \text{Died}) = 6/6264$, joint

Summary

- We need to understand probability because most of the data we analyze is the consequence of one or more *random processes*:
  - Sampling from a population, randomly assigning treatment and control groups, etc.

Summary

- We need to understand probability because most of the data we analyze is the consequence of one or more *random processes*:
  - Sampling from a population, randomly assigning treatment and control groups, etc.
- At its most basic level, probability involves three major rules:
  - The *addition rule*: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - The *complement rule*: $P(A^C) = 1 - P(A)$
  - The *multiplication rule*:
    $P(A \text{ and } B) = P(A) * P(B|A) = P(A|B) * P(B)$