Introduction to Probability

Random processes

- 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.
- It can be helpful to model a process as random even if it is not truly random.

- A random process has a set of possible outcomes
- Each outcome happens with some probability

Probability

Frequentist interpretation:

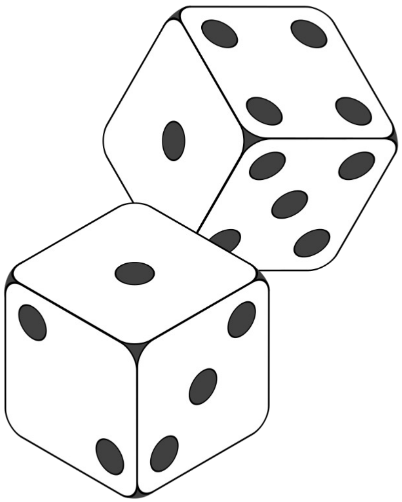
- The probability of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.

Bayesian interpretation:

- A Bayesian interprets probability as a subjective degree of belief: For the same outcome, two separate people could have different viewpoints and so assign different probabilities.

Random Events and Probability

Suppose we roll <u>two fair dice</u>…

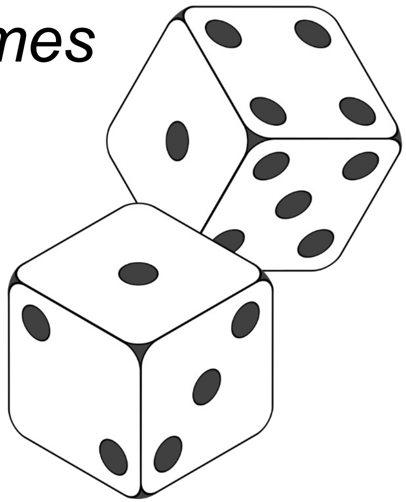


- ➢ What are the possible outcomes?
- ➢ What is the *probability* of rolling **even** numbers?
- ➢ What is the *probability* of rolling **odd** numbers?
- ➢ If one die rolls 1, then what is the probability of the second die also rolling 1?
- ➢ How to mathematically formulate outcomes and their probabilities?

*…this is an **experiment** or **random process**.*

Formulate as <u>probability space</u> having <u>3 components</u>

Random Events and Probability



1 A **sample space** Ω: *set of all possible outcomes* of the experiment.

Dice Example: All combinations of dice rolls,

$$\Omega = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}$$

2 An **event space** $\mathcal{F}$: Family of sets representing <u>allowable events</u>, where each set in $\mathcal{F}$ is a subset of the sample space Ω.

Dice Example: Event that we roll <u>even numbers</u>,

$$E = \{(2,2),(2,4),\ldots,(6,4),(6,6)\} \in \mathcal{F}$$

Sample space

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

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

Sample space

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

- A couple has one kid, what is the sample space for the sex of this kid? S = {M, F}
- A couple has two kids, what is the sample space for the sex of these kids? S = {MM, FF, FM, MF}

Disjoint and non-disjoint events

Disjoint (mutually exclusive) events: 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 events: Can happen at the same time.

- A student can get an 10 in Stats and 10 in Calculus in the same semester.

Outcomes in the sample space are disjoint

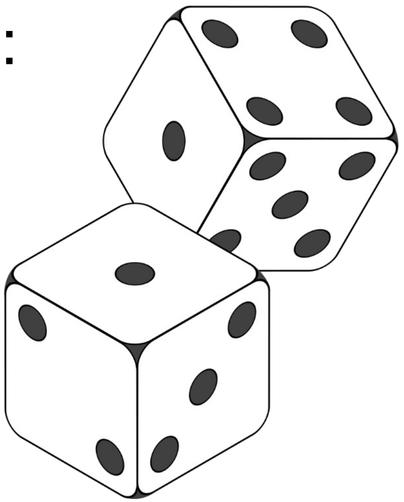
Random Events and Probability

3 A **probability function** $P : \mathcal{F} \rightarrow \mathbf{R}$ satisfying:

1. For any event E, $0 \leq P(E) \leq 1$
2. $P(\Omega) = 1 \text{ and } P(\emptyset) = 0$
3. For any *finite* or *countably infinite* sequence of pairwise mutually disjoint events $E_1, E_2, E_3, \ldots$

$$P\left(\bigcup_{i \geq 1} E_i\right) = \sum_{i \geq 1} P(E_i)$$

Axioms of Probability



(Fair) Dice Example: Probability that we roll even numbers,

$$P((2,2) \cup (2,4) \cup \ldots \cup (6,6)) = P((2,2)) + P((2,4)) + \ldots + P((6,6))$$

$$= \frac{1}{36} + \frac{1}{36} + \ldots + \frac{1}{36} = \frac{9}{36}$$

9 Possible outcomes, each with equal probability of occurring

Uniform Sample Spaces

- *Sample spaces with finite or countably infinite outcomes are called discrete sample spaces (vs. continuous sample spaces).*
- *In discrete sample spaces, if all outcomes of an experiments are considered equally likely, then for each event A,*
  - $P(A) = \frac{|A|}{|\Omega|}$
- *Example: We toss a (fair) die twice. Ω has 36 elements:*
- *11, 12, . . . , 16, 21, . . . , 26, . . . , 61, . . . , 66*
- *Let's say we are interested in the event "A: at least one 6"*
- *To assign a probability to A, we need to count the number of points in Ω, and A.*

Permutations/Combinations

Permutations

Number of distinct ways to order n objects: $n!$

Combinations

Number of distinct ways to pick k objects from a collection of n objects

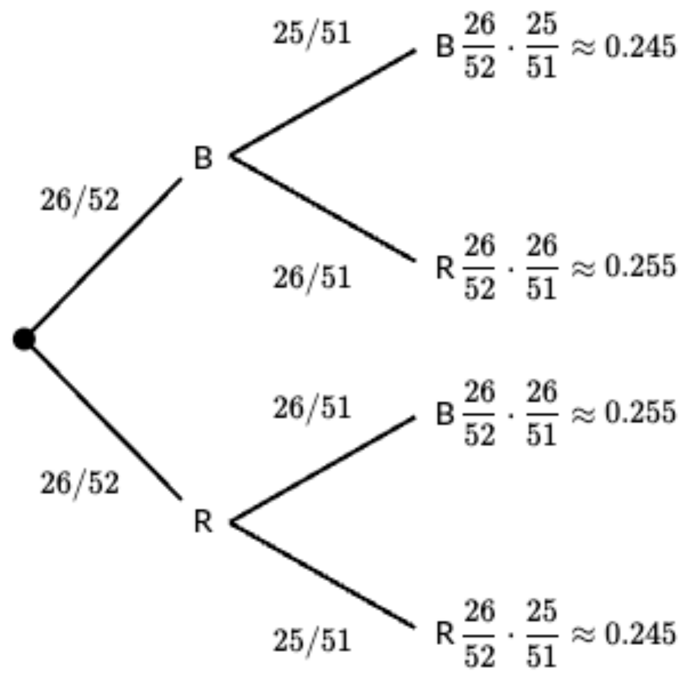
$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$

Multiplication Rule

- r *steps*
- n_r *choices in each step*
- *Number of choices:* $n_1 \times n_2 \times \cdots \times n_r$

Example

Probability of drawing two black cards from a deck



Union of events

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

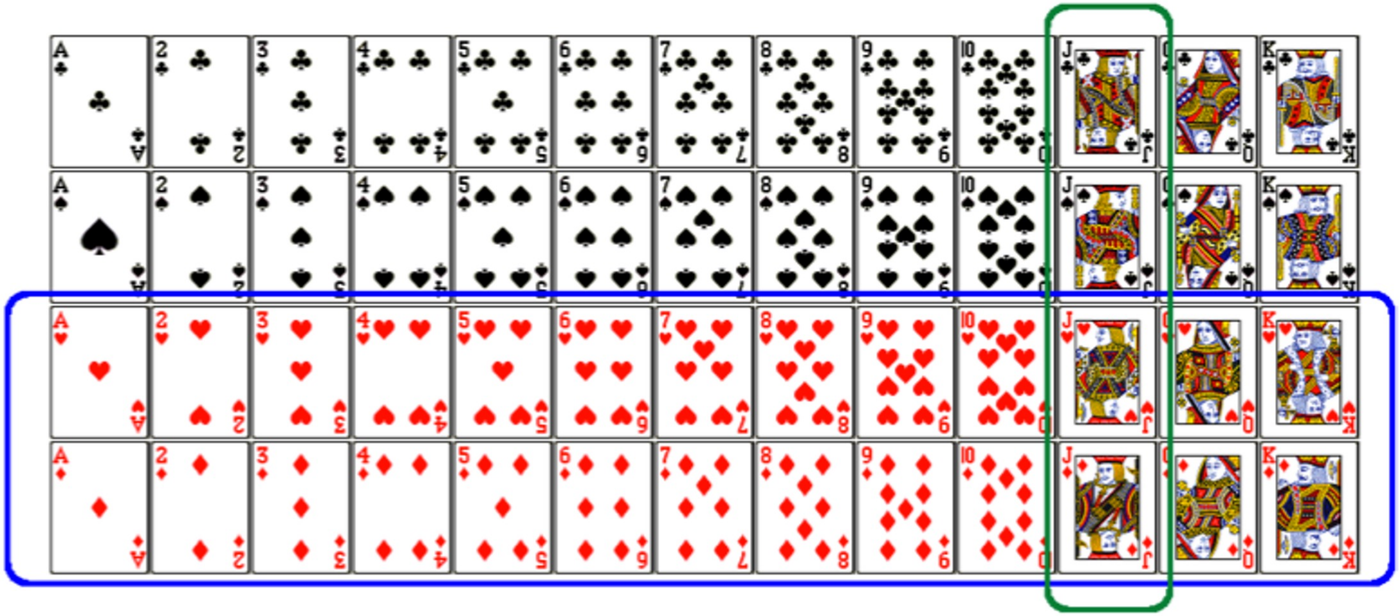
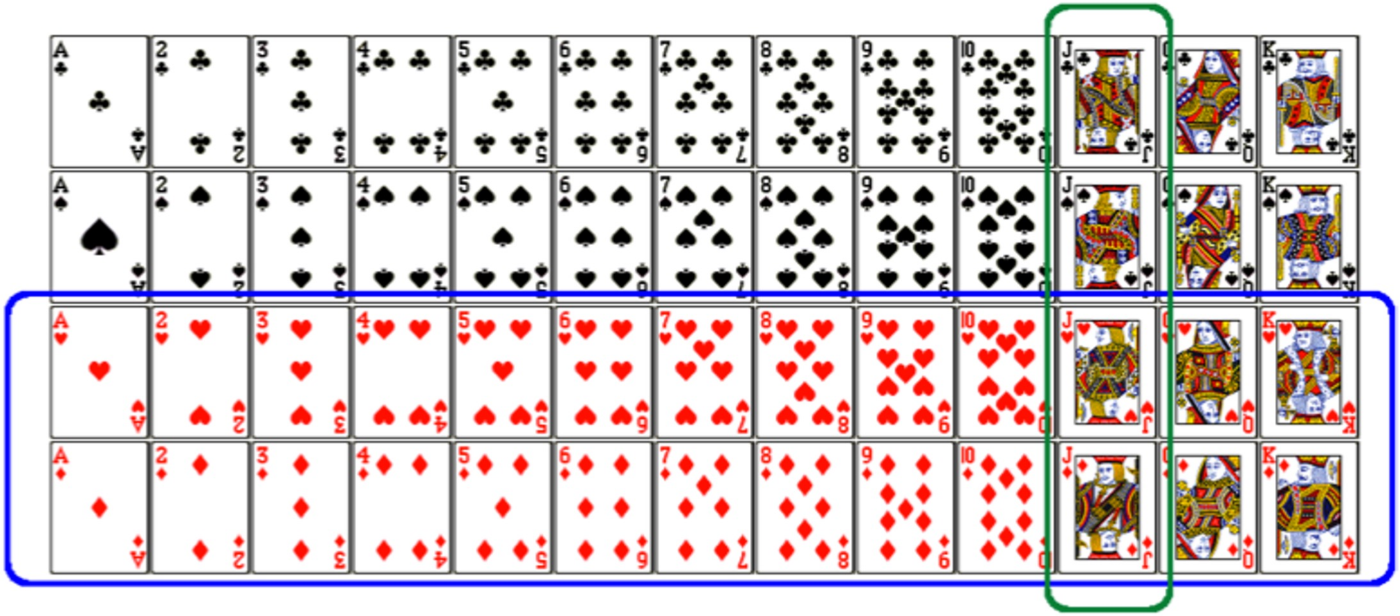


Figure from *http://www.milefoot.com/math/discrete/counting/cardfreq.htm*

Union of events

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



$$P(jack\ or\ red) = P(jack) + P(red) - P(jack\ and\ red)$$

$$= \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52}$$

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

Probability of unions

General addition rule

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

$$P(A\ or\ B) = P(A \cup B)$$
$$P(A\ and\ B) = P(A \cap B)$$

Note: For disjoint events $P(A\ and\ B) = 0$, so the above formula simplifies to $P(A\ or\ B) = P(A) + P(B)$

Practice

What is the probability that a randomly sampled student thinks marijuana should be legalized <u>or</u> 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) (40 + 36 - 78) / 165

(b) (114 + 118 - 78) / 165

(c) 78 / 165

(d) 78 / 188

(e) 11 / 47

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) (40 + 36 - 78) / 165

(b) (114 + 118 - 78) / 165

(c) 78 / 165

(d) 78 / 188

(e) 11 / 47

Independence

Definition

Two events A and B are independent if

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

We denote this as

$$A \perp B$$

Example

- You roll a die. Consider the following two events
  - A : "The outcome is an even number"
  - B : "The outcome is one of the numbers {1, 2, 3, 4}"
  - C : "The outcome is one of the numbers {1, 2, 3}"

- Find $P(A), P(B), and\ P(A \cap B)$
- Are A and B independent?
- Find $P(A), P(C), and\ P(A \cap C)$
- Are A and C independent?

Conditional Probability

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

Consider again the following two events:

- A : "The outcome is an even number"
- B : "The outcome is one of the numbers {1, 2, 3, 4}"

- You want to bet on event A. How much are you willing to bet?
- I roll the die and tell you that event B has happened (hence, the outcome is one of {1, 2, 3, 4}.
- How much are you willing to bet now?
- We just described the conditional probability

$$P(A = True|B = true)$$

Conditional Probability

Conditional Probability:

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

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

Independence

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

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

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

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

Independence

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

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

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

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 | 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.

Determining dependence based on sample data

- If conditional probabilities calculated based on sample data suggest dependence between two variables, the next step is to conduct a hypothesis test to determine if the observed difference between the probabilities is likely or unlikely to have happened by chance.
- If the observed difference between the conditional probabilities is large, then there is stronger evidence that the difference is real.
- If a sample is large, then even a small difference can provide strong evidence of a real difference.

We saw that P(protects citizens | White) = 0.67 and P(protects citizens | Hispanic) = 0.64. Under which condition would you be more convinced of a real difference between the proportions of Whites and Hispanics who think gun widespread gun ownership protects citizens?
n = 500 or n = 50,000

Conditional Independence

A and B may be independent in the universe where C has occurred.

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

$$P(A|B,C) = P(A|C)$$

Conditional Independence vs indpendence

- You roll two dice
  - A: First die gets a 1
  - B: Second die gets a 1
  - C: Sum of the two dice is 5

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

- Given C:
  - $P(A \cap B|C) =$
  - $P(A|C) =$
  - $P(B|C) =$

Conditional Independence

- Find events A, B, C such that
  - A and B are dependent
  - A and B are independent given C

Disjoint and complementary events

Disjoint events are events that cannot happen at the same time.

Are disjoint events independent?

Two events are complementary if they are disjoint and their probabilities sum to 1.

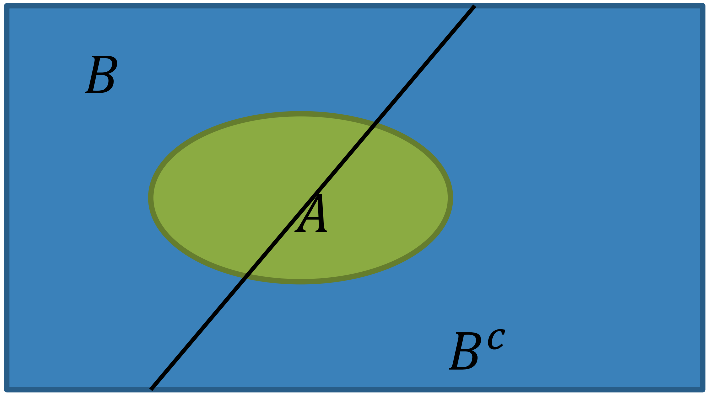
$$B^c = \Omega \setminus B$$

More than two events that are disjoint and their probabilities sum to 1 form a partition of the sample space

Law of Total Probability

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)$$



Bayes' Rule

Bayes' Rule

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

Breast cancer screening

- American Cancer Society estimates that about 1.7% of women have breast cancer. *http://www.cancer.org/cancer/cancerbasics/cancer-prevalence*
- Susan G. Komen For The Cure Foundation states that mammography correctly identifies about 78% of women who truly have breast cancer. *http://ww5.komen.org/BreastCancer/AccuracyofMammograms.html*
- An article published in 2003 suggests that up to 10% of all mammograms result in false positives for patients who do not have cancer. *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1360940*

Note: These percentages are approximate, and very difficult to estimate.

Inverting probabilities

When a patient goes through breast cancer screening there are two competing claims: patient had cancer and patient doesn't have cancer. If a mammogram yields a positive result, what is the probability that patient actually has cancer?

Inverting probabilities

When a patient goes through breast cancer screening there are two competing claims: patient had cancer and patient doesn't have cancer. If a mammogram yields a positive result, what is the probability that patient actually has cancer?

What we know

P(BC+) = 0.017

P(T+|BC+) = 0.78

P(T+|BC-) =0.1

What we want to know

P(BC+|T+) = ?

Bayes' Rule

Bayes' Rule

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

$$P(A_1|B)$$

$$= \frac{P(B|A_1)\ \ P(A_1)}{P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A\_n)}$$

Practice

Suppose a woman who gets tested once and obtains a positive result wants to get tested again. If the second test is positive, what is the probability she has cancer?

(a) 0.0936

(b) 0.088

(c) 0.48

(d) 0.52

Application activity: Bayes Rule

A common epidemiological model for the spread of diseases is the SIR model, where the population is partitioned into three groups: Susceptible, Infected, and Recovered. This is a reasonable model for diseases like chickenpox where a single infection usually provides immunity to subsequent infections. Sometimes these diseases can also be difficult to detect.

Imagine a population in the midst of an epidemic where 60% of the population is considered susceptible, 10% is infected, and 30% is recovered. The only test for the disease is accurate 95% of the time for susceptible individuals, 99% for infected individuals, but 65% for recovered individuals. (Note: In this case accurate means returning a negative result for susceptible and recovered individuals and a positive result for infected individuals).

If an individual has tested positive, what is the probability that they are actually infected?

Recap

- *Probability is a way to quantify the probability with which an event occurs.*
- *For discrete sample spaces, it is pretty easy to define a probability measure over the set of all possible events.*
- *We can use the axioms of probability to prove several properties of probability.*

- *Two events are called independent when knowing the value of one doesn't influence the probability of the value of the other.*
- *The conditional probability of A given B denotes the probability of event A in a world where B has occurred.*
- *Bayes rule connects P(A|B) and P(B|A). These two are confused but they are not the same*