Probability 1

Jacob M. Montgomery

2018

Probability 1

Probability and inference

- The goal of inference is to take some observed data or known facts and backwards induct something about the world.

Probability and inference

- The goal of inference is to take some observed data or known facts and backwards induct something about the world.
  - For instance, we might want to survey a random subset of American citizens (our data) and estimate the true attitudes of the entire American electorate (the parameter).

Probability and inference

- The goal of inference is to take some observed data or known facts and backwards induct something about the world.
  - For instance, we might want to survey a random subset of American citizens (our data) and estimate the true attitudes of the entire American electorate (the parameter).
  - Alternatively, a game theoretic model may require actors to estimate the location of the median voter given the sequence of prior election outcomes $x = (x_1, x_2, \ldots, x_n)$ and candidate positions $y = (y_1, y_2, \ldots, y_n)$.

- Probability theory is exactly the reverse.

- Probability theory is exactly the reverse.
    - Here we *know* the basic features of the data generating process (the parameters) and want to understand what the data is likely to look like.

- Probability theory is exactly the reverse.
  - Here we *know* the basic features of the data generating process (the parameters) and want to understand what the data is likely to look like.
  - For instance, we might have a fair coin and we want to understand the likelihood of flipping 20 heads before the first tail shows up.
- Obviously, most of the things you are going to be doing in your career will be about inference. Nonetheless, you *really* need to have a grasp of probabilty theory first.

What is probability theory?

- Probability is essentially thinking clearly about counting.

What is probability theory?

- Probability is essentially thinking clearly about counting.
- For the simplest problems, all you need to know is the number of ways that some set of outcomes X could happen versus the total number of ways things could have turned out.

What is probability theory?

- Probability is essentially thinking clearly about counting.
- For the simplest problems, all you need to know is the number of ways that some set of outcomes X could happen versus the total number of ways things could have turned out.
- So to begin with, you just need to focus on getting a handle on the basic concepts of:
  - How to count events
  - How to think about and handle sets
  - How counting and sets relate to the concept of "probability"
  - Conditional probability, independence, and Bayes' law

Advanced Counting

A preliminary

- For $n \in \mathbb{N}$,
$$n! = \prod_{k=1}^{n} k$$

Advanced Counting

A preliminary

- For $n \in \mathbb{N}$,
$$n! = \prod_{k=1}^{n} k$$
- Example: $3! = 3 \cdot 2 \cdot 1 = 6$

Advanced Counting

A preliminary

- For $n \in \mathbb{N}$,
$$n! = \prod_{k=1}^{n} k$$
- Example: $3! = 3 \cdot 2 \cdot 1 = 6$
- Example: $0! = 1$

Advanced Counting

A preliminary

- For $n \in \mathbb{N}$,
$$n! = \prod_{k=1}^{n} k$$
- Example: $3! = 3 \cdot 2 \cdot 1 = 6$
- Example: $0! = 1$
- Example: for $x \geq y$,

$$\frac{x!}{(x-y)!} = \frac{x(x-1) \cdot \ldots \cdot (x-y+1) \cdot (x-y) \cdot (x-y-1) \cdot \ldots \cdot 1}{(x-y) \cdot (x-y-1) \cdot \ldots \cdot 1}$$

Advanced Counting

A preliminary

- For $n \in \mathbb{N}$,

$$n! = \prod_{k=1}^{n} k$$

- Example: $3! = 3 \cdot 2 \cdot 1 = 6$
- Example: $0! = 1$
- Example: for $x \geq y$,

$$\begin{aligned}
\frac{x!}{(x-y)!} &= \frac{x(x-1) \cdot \ldots \cdot (x-y+1) \cdot (x-y) \cdot (x-y-1) \cdot \ldots \cdot 1}{(x-y) \cdot (x-y-1) \cdot \ldots \cdot 1} \\
&= x \cdot (x-1) \cdot \ldots \cdot (x-y+1)
\end{aligned}$$

Fundamental Theorem of Counting

- If there are k characteristics, each with n_k alternatives, there are $\prod_{i=1}^{k} n_k$ possible outcomes.

Fundamental Theorem of Counting

- If there are k characteristics, each with n_k alternatives, there are $\prod_{i=1}^{k} n_k$ possible outcomes.
- We often need to count the number of ways to choose a subset from some set of possiblities. The number of outcomes depends on two characteristics of the process: does the *order* matter and is *replacement* allowed?

- If there are n objects and we select $k < n$ of them, how many different outcomes are possible?
    1. Ordered, with replacement: n^k

- If there are n objects and we select $k < n$ of them, how many different outcomes are possible?
  1. Ordered, with replacement: n^k
  2. Ordered, without replacement: $\frac{n!}{(n-k)!}$

- If there are n objects and we select $k < n$ of them, how many different outcomes are possible?
  1. Ordered, with replacement: n^k
  2. Ordered, without replacement: $\frac{n!}{(n-k)!}$
  3. Unordered, with replacement: $\frac{(n+k-1)!}{(n-1)!k!} = \binom{n+k-1}{k}$

- If there are n objects and we select $k < n$ of them, how many different outcomes are possible?
    1. Ordered, with replacement: n^k
    2. Ordered, without replacement: $\frac{n!}{(n-k)!}$
    3. Unordered, with replacement: $\frac{(n+k-1)!}{(n-1)!k!} = \left(\begin{array}{c} n+k-1 \\ k \end{array} \right)$
    4. Unordered, without replacement: $(n \text{ choose } k)$: $\frac{n!}{(n-k)!k!} = \left(\begin{array}{c} n \\ k \end{array} \right)$

- Ordered events are sometimes referred to as permutations, while unordered events are combinations.
- You will almost always be working with combinations.

Sets

- **Set**: A set is any well defined collection of elements. If x is an element of S, $x \in S$.

Types of sets

1. Countably finite: a set with a finite number of elements, which can be mapped onto positive integers.

$$S = \{1, 2, 3, 4, 5, 6\}$$

Types of sets

1. Countably finite: a set with a finite number of elements, which can be mapped onto positive integers.

$$S = \{1, 2, 3, 4, 5, 6\}$$

2. Countably infinite: a set with an infinite number of elements, which can still be mapped onto positive integers.

$$S = \{1, \frac{1}{2}, \frac{1}{3}, \dots\}$$

3. Uncountably infinite: a set with an infinite number of elements, which cannot be mapped onto positive integers.

$$S = \{x : x \in [0, 1]\}$$

3. Uncountably infinite: a set with an infinite number of elements, which cannot be mapped onto positive integers.

$$S = \{x : x \in [0, 1]\}$$

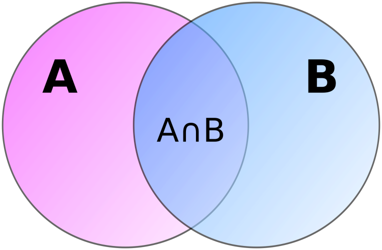
4. Empty: a set with no elements.

$$S = \{\}$$

or

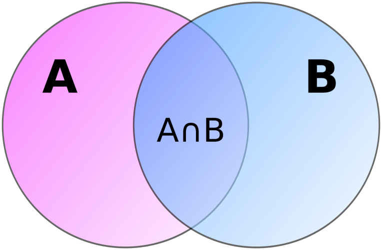
$$S = \emptyset$$

Set operations



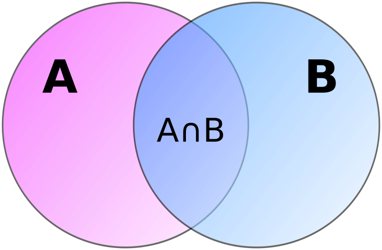
- **Union**: The union of two sets A and B, $A \cup B$, is the set containing all of the elements in A or B.

Set operations



- **Union**: The union of two sets A and B, $A \cup B$, is the set containing all of the elements in A or B.
- **Intersection**: The intersection of sets A and B, $A \cap B$, is the set containing all of the elements in both A and B.

Set operations



- **Union**: The union of two sets A and B, $A \cup B$, is the set containing all of the elements in A or B.
- **Intersection**: The intersection of sets A and B, $A \cap B$, is the set containing all of the elements in both A and B.
- **Complement**: If set A is a subset of S, then the complement of A, denoted A^C, is the set containing all of the elements in S that are not in A.

Properties of set operations:

1. Commutative: $A \cup B = B \cup A$, $A \cap B = B \cap A$

Properties of set operations:

1. Commutative: $A \cup B = B \cup A$, $A \cap B = B \cap A$
2. Associative: $A \cup (B \cup C) = (A \cup B) \cup C$,
   $A \cap (B \cap C) = (A \cap B) \cap C$

Properties of set operations:

1. Commutative: $A \cup B = B \cup A$, $A \cap B = B \cap A$
2. Associative: $A \cup (B \cup C) = (A \cup B) \cup C$,
   $A \cap (B \cap C) = (A \cap B) \cap C$
3. Distributive: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
   $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Properties of set operations:

1. Commutative: $A \cup B = B \cup A$, $A \cap B = B \cap A$
2. Associative: $A \cup (B \cup C) = (A \cup B) \cup C$,
   $A \cap (B \cap C) = (A \cap B) \cap C$
3. Distributive: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
   $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
4. de Morgan's laws: $(A \cup B)^C = A^C \cap B^C$, $(A \cap B)^C = A^C \cup B^C$

Disjointedness and partitions

- Sets are disjoint when they do not intersect, such that $A \cap B = \{\emptyset\}$. A collection of sets is pairwise disjoint if, for all $i \neq j$, $A_i \cap A_j = \{\emptyset\}$.

Disjointedness and partitions

- Sets are disjoint when they do not intersect, such that $A \cap B = \{\emptyset\}$. A collection of sets is pairwise disjoint if, for all $i \neq j$, $A_i \cap A_j = \{\emptyset\}$.
- A collection of sets form a partition of set S if they are pairwise disjoint and they cover set S, such that $\bigcup_{i=1}^{k} A_i = S$.

Probability

- Probability is an expression of uncertainty.

Probability

- Probability is an expression of uncertainty.
- Modern probability theory is a way of estimating our uncertainty about some future events given specific assumed properties of the world.

Probability

- Probability is an expression of uncertainty.
- Modern probability theory is a way of estimating our uncertainty about some future events given specific assumed properties of the world.
- This is a formalization of basic human intuition about how to handle risk.

Sample Space

- A set or collection of all possible outcomes from some process.
- Outcomes in the set can be discrete elements (countable) or points along a continuous interval (uncountable).

Sample Space

- A set or collection of all possible outcomes from some process.
- Outcomes in the set can be discrete elements (countable) or points along a continuous interval (uncountable).
- Examples:

    1. **Discrete**: the numbers on a die, the number of possible wars that could occur each year, whether a vote cast is republican or democrat.

Sample Space

- A set or collection of all possible outcomes from some process.
- Outcomes in the set can be discrete elements (countable) or points along a continuous interval (uncountable).
- Examples:
  1. **Discrete**: the numbers on a die, the number of possible wars that could occur each year, whether a vote cast is republican or democrat.
  2. **Continuous**: GNP, arms spending, age.

Probability Distribution/Function

- A probability *function* on a sample space S is a mapping $\Pr(A)$ from events in S to the real numbers.
- It is just like any other function.

Probability Distribution/Function

- A probability *function* on a sample space S is a mapping $\Pr(A)$ from events in S to the real numbers.
- It is just like any other function.
- We have some event/sample space S we have a probability space (e.g., the probability of event x happening is some number in $[0, 1]$) and we have the function that translates x into the probability space that we denote $p(x)$ or $f(x)$.

Example

- Let's say we are flipping two coins.
- These are the things that can happen is $S = HH, HT, TH, TT$.
- We are interested in the number of heads
- The **sample space is then** $x = 0, 1, 2$

Example

- Let's say we are flipping two coins.
- These are the things that can happen is $S = HH, HT, TH, TT$.
- We are interested in the number of heads
- The **sample space is then** $x = 0, 1, 2$

Outcome	X=x
HH	2
HT	1
TH	1
TT	0

Example

- Let's say we are flipping two coins.
- These are the things that can happen is $S = HH, HT, TH, TT$.
- We are interested in the number of heads
- The **sample space is then** $x = 0, 1, 2$

Outcome	X=x
HH	2
HT	1
TH	1
TT	0

x	p(x)
TT $\rightarrow$ 0	.25
HT, TH $\rightarrow$ 1	.50
HH $\rightarrow$ 2	.25
Sum	1.00

Exercise: Rolling two fair dice

1. Write out the sample space
2. Write out the empirical probability function

Axioms of Probability

- Probability functions will satisfy the following three axioms (due to Kolmogorov).

Axioms of Probability

- Probability functions will satisfy the following three axioms (due to Kolmogorov).
- Define the number $\Pr(A)$ correponding to each event A in the sample space S such that:

  1. Axiom: For any event A, $\Pr(A) \geq 0$.

Axioms of Probability

- Probability functions will satisfy the following three axioms (due to Kolmogorov).
- Define the number $\Pr(A)$ correponding to each event A in the sample space S such that:
    1. Axiom: For any event A, $\Pr(A) \geq 0$.
    2. Axiom: $\Pr(S) = 1$

Axioms of Probability

- Probability functions will satisfy the following three axioms (due to Kolmogorov).
- Define the number $\Pr(A)$ correponding to each event A in the sample space S such that:

1. Axiom: For any event A, $\Pr(A) \geq 0$.
2. Axiom: $\Pr(S) = 1$
3. Axiom: For any sequence of disjoint events $A_1, A_2, \ldots$ (of which there may be infinitely many),

$$\Pr\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \Pr(A_i)$$

Basic Theorems of Probability

- Using these three axioms, we can define all of the common theorems of probability.
    1. $\Pr(\emptyset) = 0$

Basic Theorems of Probability

- Using these three axioms, we can define all of the common theorems of probability.
  1. $\Pr(\emptyset) = 0$
  2. $\Pr(A^C) = 1 - \Pr(A)$

Basic Theorems of Probability

- Using these three axioms, we can define all of the common theorems of probability.
  1. $\Pr(\emptyset) = 0$
  2. $\Pr(A^C) = 1 - \Pr(A)$
  3. For any event A, $0 \leq \Pr(A) \leq 1$.

Basic Theorems of Probability

- Using these three axioms, we can define all of the common theorems of probability.
  1. $\Pr(\emptyset) = 0$
  2. $\Pr(A^C) = 1 - \Pr(A)$
  3. For any event A, $0 \leq \Pr(A) \leq 1$.

4. If $A \subset B$, then $\Pr(A) \leq \Pr(B)$.
5. For any two events A and B,
   $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
6. For any sequence of n events (which need not be disjoint)

$$A_1, A_2, \ldots, A_n$$

   ,

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr(A_i)$$

Example 1

Let's assume we have an evenly-balanced, six-sided die. Then,

1. Sample space $S =$

Example 1

Let's assume we have an evenly-balanced, six-sided die. Then,

1. Sample space $S = \{1, 2, 3, 4, 5, 6\}$
2. $\Pr(1) = \cdots = \Pr(6) = 1/6$

Example 1

Let's assume we have an evenly-balanced, six-sided die. Then,

1. Sample space $S = \{1, 2, 3, 4, 5, 6\}$
2. $\Pr(1) = \cdots = \Pr(6) = 1/6$
3. $\Pr(\emptyset) = \Pr(7) = 0$
4. $\Pr\left(\{1, 3, 5\}\right) =$

Example 1

Let's assume we have an evenly-balanced, six-sided die. Then,

1. Sample space $S = \{1, 2, 3, 4, 5, 6\}$
2. $\mathrm{Pr}(1) = \cdots = \mathrm{Pr}(6) = 1/6$
3. $\mathrm{Pr}(\emptyset) = \mathrm{Pr}(7) = 0$
4. $\mathrm{Pr}\left(\{1, 3, 5\}\right) = 1/6 + 1/6 + 1/6 = 1/2$

5. $\Pr\left(\overline{\{1,2\}}\right) =$

5. $\Pr\left(\overline{\{1,2\}}\right) = \Pr(\{3,4,5,6\}) = 2/3$
6. Let $B = S$ and $A = \{1,2,3,4,5\} \subset B$. Then $\Pr(A) = 5/6 < \Pr(B) = 1$.

5. $\Pr\left(\overline{\{1,2\}}\right) = \Pr(\{3,4,5,6\}) = 2/3$
6. Let $B = S$ and $A = \{1,2,3,4,5\} \subset B$. Then $\Pr(A) = 5/6 < \Pr(B) = 1$.
7. Let $A = \{1,2,3\}$ and $B = \{2,4,6\}$. Then $A \cup B = \{1,2,3,4,6\}$, $A \cap B = \{2\}$, and

$$\Pr(A \cup B) \quad =$$

5. $\Pr\left(\overline{\{1,2\}}\right) = \Pr(\{3,4,5,6\}) = 2/3$
6. Let $B = S$ and $A = \{1,2,3,4,5\} \subset B$. Then $\Pr(A) = 5/6 < \Pr(B) = 1$.
7. Let $A = \{1,2,3\}$ and $B = \{2,4,6\}$. Then $A \cup B = \{1,2,3,4,6\}$, $A \cap B = \{2\}$, and

$$\begin{aligned}\Pr(A \cup B) &= \Pr(A) + \Pr(B) - \Pr(A \cap B) \\ &= 3/6 + 3/6 - 1/6\end{aligned}$$

5. $\Pr\left(\overline{\{1,2\}}\right) = \Pr(\{3,4,5,6\}) = 2/3$
6. Let $B = S$ and $A = \{1,2,3,4,5\} \subset B$. Then $\Pr(A) = 5/6 < \Pr(B) = 1$.
7. Let $A = \{1,2,3\}$ and $B = \{2,4,6\}$. Then $A \cup B = \{1,2,3,4,6\}$, $A \cap B = \{2\}$, and

$$
\begin{aligned}
\Pr(A \cup B) &= \Pr(A) + \Pr(B) - \Pr(A \cap B) \\
&= 3/6 + 3/6 - 1/6 \\
&= 5/6
\end{aligned}
$$

Example 2

Let a trial be an equally likely selection of one of X, Y, Z for exit poll.

Example 2

Let a trial be an equally likely selection of one of X, Y, Z for exit poll.

1. What's the sample space?

Example 2

Let a trial be an equally likely selection of one of X, Y, Z for exit poll.

1. What's the sample space? $\{X, Y, Z\}$
2. For event $A =$ "select X", calculate $p(A)$.

Example 2

Let a trial be an equally likely selection of one of X, Y, Z for exit poll.

1. What's the sample space? $\{X, Y, Z\}$
2. For event $A =$ "select X", calculate $p(A)$. $p(X) = \frac{1}{3}$
3. Calculate $p(A^C)$.

Example 2

Let a trial be an equally likely selection of one of X, Y, Z for exit poll.

1. What's the sample space? $\{X, Y, Z\}$
2. For event $A =$ "select X", calculate $p(A)$. $p(X) = \frac{1}{3}$
3. Calculate $p(A^C)$. $p(Y \cup Z) = \frac{2}{3}$
4. Calculate $p(A \cup A^C)$.

Example 2

Let a trial be an equally likely selection of one of X, Y, Z for exit poll.

1. What's the sample space? $\{X, Y, Z\}$
2. For event $A =$ "select X", calculate $p(A)$. $p(X) = \frac{1}{3}$
3. Calculate $p(A^C)$. $p(Y \cup Z) = \frac{2}{3}$
4. Calculate $p(A \cup A^C)$. $p(X \cup (Y \cup Z)) = 1$

Conditional Probability and Bayes Law

- Let A, B be two events with $p(A) > 0$,

Conditional Probability and Bayes Law

- Let A, B be two events with $p(A) > 0$,
  - the *conditional probability* of B given A is

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

Conditional Probability and Bayes Law

- Let A, B be two events with $p(A) > 0$,
  - the *conditional probability* of B given A is

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

  - Equivalently,

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

Conditional Probability and Bayes Law

- Let A, B be two events with $p(A) > 0$,
  - the *conditional probability* of B given A is
  $$p(B|A) = \frac{p(B \cap A)}{p(A)}$$
  - Equivalently,
  $$p(B \cap A) = p(A)p(B|A)$$
  - Relatedly,
  $$p(A \cap B \cap C) = p(A)p(B|A)p(C|A \cap B)$$

Conditional Probability and Bayes Law

- Let A, B be two events with $p(A) > 0$,
  - the *conditional probability* of B given A is

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

  - Equivalently,

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

  - Relatedly,

$$p(A \cap B \cap C) = p(A)p(B|A)p(C|A \cap B)$$

  - Conditioning information can be subtlely important

Example: Older Child Paradox

- Consider picking a two-child family at random, with equiprobable $S = \{FF, MM, FM, MF\}$.

Example: Older Child Paradox

- Consider picking a two-child family at random, with equiprobable $S = \{FF, MM, FM, MF\}$.
- Let $B = FF$, so $p(B) = .25$.

Example: Older Child Paradox

- Consider picking a two-child family at random, with equiprobable $S = \{FF, MM, FM, MF\}$.
- Let $B = FF$, so $p(B) = .25$.
- Let A be "at least one girl". Calculate $p(B|A)$.

Example: Older Child Paradox

- Consider picking a two-child family at random, with equiprobable $S = \{FF, MM, FM, MF\}$.
- Let $B = FF$, so $p(B) = .25$.
- Let A be "at least one girl". Calculate $p(B|A)$.

$$p(B|A) \quad = \quad \frac{p(B \cap A)}{p(A)} = \frac{1/4}{3/4} = \frac{1}{3}$$

- Let C be "older child is girl". Calculate $p(B|C)$.

Example: Older Child Paradox

- Consider picking a two-child family at random, with equiprobable $S = \{FF, MM, FM, MF\}$.
- Let $B = FF$, so $p(B) = .25$.
- Let A be "at least one girl". Calculate $p(B|A)$.

$$p(B|A) \quad = \quad \frac{p(B \cap A)}{p(A)} = \frac{1/4}{3/4} = \frac{1}{3}$$

- Let C be "older child is girl". Calculate $p(B|C)$.

$$p(B|C) \quad = \quad \frac{p(B \cap C)}{p(C)} = \frac{1/4}{2/4} = \frac{1}{2}$$

Multiplicative Law of Probability

- The probability of the intersection of two events A and B is

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

Multiplicative Law of Probability

- The probability of the intersection of two events A and B is

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

- Follows directly from the definition of conditional probability.

Law of Total Probability

- Let S be the sample space of some experiment and let the disjoint k events $B_1, \ldots, B_k$ partition S.

Law of Total Probability

- Let S be the sample space of some experiment and let the disjoint k events $B_1, \ldots, B_k$ partition S.
- If A is some other event in S, then the events $AB_1, AB_2, \ldots, AB_k$ will form a partition of A and we can write A as

$$A = (AB_1) \cup \cdots \cup (AB_k)$$

- Since the k events are disjoint,

$$\begin{aligned}
\Pr(A) &= \sum_{i=1}^{k} \Pr(A, B_i) \\
&= \sum_{i=1}^{k} \Pr(B_i)\Pr(A|B_i)
\end{aligned}$$

- Since the k events are disjoint,

$$\begin{aligned} \Pr(A) &= \sum_{i=1}^{k} \Pr(A, B_i) \\ &= \sum_{i=1}^{k} \Pr(B_i) \Pr(A|B_i) \end{aligned}$$

- Sometimes it is easier to calculate the conditional probabilities and sum them than it is to calculate $\Pr(A)$ directly.

Bayes Law

- From conditional probability,

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

Bayes Law

- From conditional probability,

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

- But also,

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

Bayes Law

- From conditional probability,

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

- But also,

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

- Since we know $p(B \cap A) = p(A \cap B)$,

Bayes Law

- From conditional probability,

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

- But also,

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

- Since we know $p(B \cap A) = p(A \cap B)$,

$$p(A)p(B|A) \;\; = \;\; p(B)p(A|B)$$

Bayes Law

- From conditional probability,

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

- But also,

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

- Since we know $p(B \cap A) = p(A \cap B)$,

$$\begin{aligned} p(A)p(B|A) &= p(B)p(A|B) \\ p(B|A) &= \frac{p(B)p(A|B)}{p(A)} \end{aligned}$$

Then, from the Law of Total Probability,

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

Then, from the Law of Total Probability,

$$
\begin{aligned}
p(B|A) &= \frac{p(B)p(A|B)}{p(A)} \\
&= \frac{p(B)p(A|B)}{p(A \cap B) + p(A \cap B^C)}
\end{aligned}
$$

Then, from the Law of Total Probability,

$$\begin{aligned}
p(B|A) &= \frac{p(B)p(A|B)}{p(A)} \\
&= \frac{p(B)p(A|B)}{p(A \cap B) + p(A \cap B^C)} \\
&= \frac{p(B)p(A|B)}{p(B)p(A|B) + p(B^C)p(A|B^C)}
\end{aligned}$$

More generally, where the B_j form a partition,

$$p(B_j|A) = \frac{p(B_j)p(A|B_j)}{\sum_{i=1}^{k} p(B_i)p(A|B_i)}$$

More generally, where the B_j form a partition,

$$p(B_j|A) = \frac{p(B_j)p(A|B_j)}{\sum_{i=1}^{k} p(B_i)p(A|B_i)}$$

These are *Bayes' Law* or *Bayes' Theorem* or *Bayes' Rule*.

Thinking about Bayes' Rule

- Assume that events $B_1, \ldots, B_k$ form a partition of the space S.
- Then

$$\Pr(B_j|A) = \frac{\Pr(A, B_j)}{\Pr(A)} = \frac{\Pr(B_j)\Pr(A|B_j)}{\sum_{i=1}^{k} \Pr(B_i)\Pr(A|B_i)}$$

- If there are only two states of B, then this is just

$$\Pr(B_1|A) = \frac{\Pr(B_1)\Pr(A|B_1)}{\Pr(B_1)\Pr(A|B_1) + \Pr(B_2)\Pr(A|B_2)}$$

- If this was a continuouse distribution we could write this as:

$$\Pr(B_j|A) = \frac{\Pr(A, B_j)}{\Pr(A)} = \frac{\Pr(B)\Pr(A|B)}{\int \Pr(A, B)\Pr(B)}$$

- Note that the denominator has an indefinite integral, meaning that there is an unknown integration constant to consider.
- On the other hand, the denominator will be constant wrt B...

In Bayesian modeling and data analysis,

$$p(\theta|y) \quad \propto \quad p(\theta)p(y|\theta)$$

In Bayesian modeling and data analysis,

$$\begin{aligned} p(\theta|y) &\propto p(\theta)p(y|\theta) \\ \text{posterior} &\propto \text{prior} \cdot \text{likelihood} \end{aligned}$$

Example: Rare conditions and “accurate” tests

- A test for cancer correctly detects it 90% of the time, but incorrectly identifies a person as having cancer 10% of the time. If 10% of all people have cancer at any given time, what is the probability that a person who tests positive actually has cancer?

$$p(B|A) = \frac{p(B)p(A|B)}{p(B)p(A|B) + p(B^C)p(A|B^C)}$$

Example: Rare conditions and "accurate" tests

- A test for cancer correctly detects it 90% of the time, but incorrectly identifies a person as having cancer 10% of the time. If 10% of all people have cancer at any given time, what is the probability that a person who tests positive actually has cancer?

$$
\begin{aligned}
p(B|A) &= \frac{p(B)p(A|B)}{p(B)p(A|B) + p(B^C)p(A|B^C)} \\
p(Y|+) &= \frac{p(Y)p(+|Y)}{p(Y)p(+|Y) + p(Y^C)p(+|Y^C)}
\end{aligned}
$$

Example: Rare conditions and “accurate” tests

- A test for cancer correctly detects it 90% of the time, but incorrectly identifies a person as having cancer 10% of the time. If 10% of all people have cancer at any given time, what is the probability that a person who tests positive actually has cancer?

$$
\begin{aligned}
p(B|A) &= \frac{p(B)p(A|B)}{p(B)p(A|B) + p(B^C)p(A|B^C)} \\
p(Y|+) &= \frac{p(Y)p(+|Y)}{p(Y)p(+|Y) + p(Y^C)p(+|Y^C)} \\
&= \frac{.1 \cdot p(+|Y)}{.1 \cdot p(+|Y) + .9 \cdot p(+|Y^C)}
\end{aligned}
$$

Example: Rare conditions and "accurate" tests

- A test for cancer correctly detects it 90% of the time, but incorrectly identifies a person as having cancer 10% of the time. If 10% of all people have cancer at any given time, what is the probability that a person who tests positive actually has cancer?

$$\begin{aligned}
p(B|A) &= \frac{p(B)p(A|B)}{p(B)p(A|B) + p(B^C)p(A|B^C)} \\
p(Y|+) &= \frac{p(Y)p(+|Y)}{p(Y)p(+|Y) + p(Y^C)p(+|Y^C)} \\
&= \frac{.1 \cdot p(+|Y)}{.1 \cdot p(+|Y) + .9 \cdot p(+|Y^C)} \\
&= \frac{.1 \cdot .9}{.1 \cdot .9 + .9 \cdot .1}
\end{aligned}$$

Example: Rare conditions and "accurate" tests

- A test for cancer correctly detects it 90% of the time, but incorrectly identifies a person as having cancer 10% of the time. If 10% of all people have cancer at any given time, what is the probability that a person who tests positive actually has cancer?

$$
\begin{aligned}
p(B|A) &= \frac{p(B)p(A|B)}{p(B)p(A|B) + p(B^C)p(A|B^C)} \\
p(Y|+) &= \frac{p(Y)p(+|Y)}{p(Y)p(+|Y) + p(Y^C)p(+|Y^C)} \\
&= \frac{.1 \cdot p(+|Y)}{.1 \cdot p(+|Y) + .9 \cdot p(+|Y^C)} \\
&= \frac{.1 \cdot .9}{.1 \cdot .9 + .9 \cdot .1} \\
&= \frac{.09}{.09 + .09}
\end{aligned}
$$

Example: Rare conditions and "accurate" tests

- A test for cancer correctly detects it 90% of the time, but incorrectly identifies a person as having cancer 10% of the time. If 10% of all people have cancer at any given time, what is the probability that a person who tests positive actually has cancer?

$$
\begin{aligned}
p(B|A) &= \frac{p(B)p(A|B)}{p(B)p(A|B) + p(B^C)p(A|B^C)} \\
p(Y|+) &= \frac{p(Y)p(+|Y)}{p(Y)p(+|Y) + p(Y^C)p(+|Y^C)} \\
&= \frac{.1 \cdot p(+|Y)}{.1 \cdot p(+|Y) + .9 \cdot p(+|Y^C)} \\
&= \frac{.1 \cdot .9}{.1 \cdot .9 + .9 \cdot .1} \\
&= \frac{.09}{.09 + .09} \\
&= .5
\end{aligned}
$$

Independence

- If the occurrence or nonoccurrence of either events A and B have no effect on the occurrence or nonoccurrence of the other, then A and B are **independent**.

Independence

- If the occurrence or nonoccurrence of either events A and B have no effect on the occurrence or nonoccurrence of the other, then A and B are **independent**.
- If A and B are independent, then
  1. $\Pr(A|B) = \Pr(A)$
  2. $\Pr(B|A) = \Pr(B)$
  3. $\Pr(A \cap B) = \Pr(A)\Pr(B)$

Pairwise independence

- A set of more than two events $A_1, A_2, \ldots, A_k$ is **pairwise independent** if $\Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j)$, $\forall i \neq j$.

Pairwise independence

- A set of more than two events $A_1, A_2, \ldots, A_k$ is **pairwise independent** if $\Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j)$, $\forall i \neq j$.
- Note that this does *not* necessarily imply that $\Pr(\bigcap_{i=1}^{k} A_i) = \prod_{i=1}^{K} \Pr(A_i)$.

- Consider two flips of a fair coin.

- Consider two flips of a fair coin. $\{HH, HT, TH, TT\}$. Let
    1. $A_1 = H\square$
    2. $A_2 = \square H$
    3. $A_3 =$ exactly one H
- These are pairwise independent

- Consider two flips of a fair coin. $\{HH, HT, TH, TT\}$. Let
    1. $A_1 = H\square$
    2. $A_2 = \square H$
    3. $A_3 =$ exactly one H
- These are pairwise independent, but not independent as a group:

- Consider two flips of a fair coin. $\{HH, HT, TH, TT\}$. Let
  1. $A_1 = H\Box$
  2. $A_2 = \Box H$
  3. $A_3 =$ exactly one H
- These are pairwise independent, but not independent as a group:

$$p(A_1 \cap A_2 \cap A_3) = 0$$

- Consider two flips of a fair coin. $\{HH, HT, TH, TT\}$. Let
  1. $A_1 = H\square$
  2. $A_2 = \square H$
  3. $A_3 =$ exactly one H
- These are pairwise independent, but not independent as a group:

$$
\begin{aligned}
p(A_1 \cap A_2 \cap A_3) &= 0 \\
p(A_1)p(A_2)p(A_3) &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}
\end{aligned}
$$

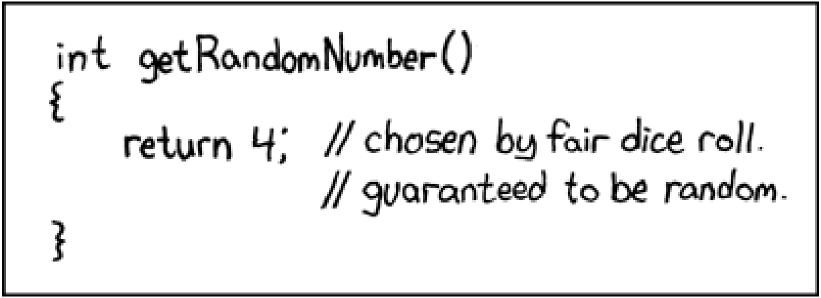
Conditional independence

- If the occurrence of A or B conveys no information about the occurrence of the other, once you know the occurrence of a third event C, then A and B are **conditionally independent** (conditional on C)

Conditional independence

- If the occurrence of A or B conveys no information about the occurrence of the other, once you know the occurrence of a third event C, then A and B are **conditionally independent** (conditional on C):
  1. $\Pr(A|B \cap C) = \Pr(A|C)$
  2. $\Pr(B|A \cap C) = \Pr(B|C)$
  3. $\Pr(A \cap B|C) = \Pr(A|C)\Pr(B|C)$
- Conditional independence is one of the fundamental assumptions deployed for most statistical estimation techniques. It is a *very* strong assumption.

Random Variables


int getRandomNumber()
{
return 4; // chosen by fair dice roll.
// guaranteed to be random.
}

Getting oriented

- The intellectual beginnings of probability began in gambling, and this is still the easiest way to teach it.

Getting oriented

- The intellectual beginnings of probability began in gambling, and this is still the easiest way to teach it.
- In probability theory, random variables are something abstract. A random variable is a yet-to-be observed value.

Getting oriented

- The intellectual beginnings of probability began in gambling, and this is still the easiest way to teach it.
- In probability theory, random variables are something abstract. A random variable is a yet-to-be observed value.
- What is the probability that a coin will turn up heads? What is the probability the next card will be an ace?

- Depending on the kinds of events we are talking about, we have identified several “types” of random variables.
- These variables have known functional forms, several of which we will discuss today.

- Depending on the kinds of events we are talking about, we have identified several “types” of random variables.
- These variables have known functional forms, several of which we will discuss today.
- Moreover, these functions have been extensively studied and their properties are well understood.

- Depending on the kinds of events we are talking about, we have identified several "types" of random variables.
- These variables have known functional forms, several of which we will discuss today.
- Moreover, these functions have been extensively studied and their properties are well understood.
- The focus of the rest of this lecture is to get you famillair with these "kinds" of variables.

Levels measurement

- In empirical research, data can be classified along several dimensions. We have already distinguished between discrete (countable) and continuous (uncountable) data.

Levels measurement

- In empirical research, data can be classified along several dimensions. We have already distinguished between discrete (countable) and continuous (uncountable) data.
- We can also look at the precision with which the underlying quantities are measured.

Levels measurement

- In empirical research, data can be classified along several dimensions. We have already distinguished between discrete (countable) and continuous (uncountable) data.
- We can also look at the precision with which the underlying quantities are measured.
- If you do not already understand the difference, please review:
  - Nominal
  - Ordinal
  - Interval
  - Ratio

Types of distribution functions

You will work primarily with three types of distribution functions:

1. Probability mass functions
2. Probability density functions
3. Cumulative distribution functions

Probability Mass Functions

- The PMF of a discrete distribution maps $p : X \to [0, 1]$

Probability Mass Functions

- The PMF of a discrete distribution maps $p : X \to [0, 1]$
- Domain: possible discrete values of X, often $\subseteq \mathbb{N}$

Probability Mass Functions

- The PMF of a discrete distribution maps $p : X \to [0, 1]$
- Domain: possible discrete values of X, often $\subseteq \mathbb{N}$
- Range: $0 \leq p(x) \leq 1$

Probability Mass Functions

- The PMF of a discrete distribution maps $p : X \to [0, 1]$
- Domain: possible discrete values of X, often $\subseteq \mathbb{N}$
- Range: $0 \leq p(x) \leq 1$
- From Axioms:

$$\sum p(x) = 1$$

Probability Mass Functions

- The PMF of a discrete distribution maps $p : X \to [0, 1]$
- Domain: possible discrete values of X, often $\subseteq \mathbb{N}$
- Range: $0 \leq p(x) \leq 1$
- From Axioms:
$$\sum p(x) = 1$$
- Write
$$p(X = x|\text{parameters})$$

Probability Density Functions

- PDF of a continuous(/discrete) distribution maps
  $p : X \to [0, 1]$

Probability Density Functions

- PDF of a continuous(/discrete) distribution maps $p : X \to [0, 1]$
- PDF also called the "marginal dist'n of x"

Probability Density Functions

- PDF of a continuous(/discrete) distribution maps $p : X \to [0, 1]$
- PDF also called the "marginal dist'n of x"
- Domain: possible values of X, $\subseteq \mathbb{R}^1$

Probability Density Functions

- PDF of a continuous(/discrete) distribution maps $p : X \to [0, 1]$
- PDF also called the "marginal dist'n of x"
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq p(x) \leq 1$

Probability Density Functions

- PDF of a continuous(/discrete) distribution maps $p : X \to [0, 1]$
- PDF also called the "marginal dist'n of x"
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq p(x) \leq 1$
- From Axioms:

$$\int_{-\infty}^{\infty} p(x) = 1$$

Probability Density Functions

- PDF of a continuous(/discrete) distribution maps $p : X \to [0, 1]$
- PDF also called the "marginal dist'n of x"
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq p(x) \leq 1$
- From Axioms:

$$\int_{-\infty}^{\infty} p(x) = 1$$

- Write

$$p(X = x|\text{parameters})$$

Probability Density Functions

- PDF of a continuous(/discrete) distribution maps $p : X \to [0, 1]$
- PDF also called the "marginal dist'n of x"
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq p(x) \leq 1$
- From Axioms:

$$\int_{-\infty}^{\infty} p(x) = 1$$

- Write

$$p(X = x|\text{parameters})$$

- For joint distn's, $p(x, y) = p(X = x, Y = y)$

Probability Density Functions

- PDF of a continuous(/discrete) distribution maps $p : X \to [0, 1]$
- PDF also called the "marginal dist'n of x"
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq p(x) \leq 1$
- From Axioms:

$$\int_{-\infty}^{\infty} p(x) = 1$$

- Write

$$p(X = x|\mathrm{parameters})$$

- For joint distn's, $p(x, y) = p(X = x, Y = y)$
- Generally, the joint dist'n is **not** the product of the marginals

(Cumulative) Distribution Functions

- The CDF of a distribution maps $F : X \to [0, 1]$

(Cumulative) Distribution Functions

- The CDF of a distribution maps $F : X \rightarrow [0, 1]$
- Domain: possible values of X, $\subseteq \mathbb{R}^1$

(Cumulative) Distribution Functions

- The CDF of a distribution maps $F : X \to [0, 1]$
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq F(x) \leq 1$

(Cumulative) Distribution Functions

- The CDF of a distribution maps $F : X \to [0, 1]$
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq F(x) \leq 1$
- Discrete dist'n: $F(x) = p(X \leq x) = \sum_{x \leq z} p(z) = \sum_{x=0}^{z} p(z)$

(Cumulative) Distribution Functions

- The CDF of a distribution maps $F : X \rightarrow [0, 1]$
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq F(x) \leq 1$
- Discrete dist'n: $F(x) = p(X \leq x) = \sum_{x \leq z} p(z) = \sum_{x=0}^{z} p(z)$
- Continuous dist'n: $F(x) = p(X \leq x) = \int_{-\infty}^{x} p(z) dz$

(Cumulative) Distribution Functions

- The CDF of a distribution maps $F : X \rightarrow [0, 1]$
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq F(x) \leq 1$
- Discrete dist'n: $F(x) = p(X \leq x) = \sum_{x \leq z} p(z) = \sum_{x=0}^{z} p(z)$
- Continuous dist'n: $F(x) = p(X \leq x) = \int_{-\infty}^{x} p(z)dz$
- Note: In both cases, $F(max(x)) = 1$

(Cumulative) Distribution Functions

- The CDF of a distribution maps $F : X \to [0, 1]$
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq F(x) \leq 1$
- Discrete dist'n: $F(x) = p(X \leq x) = \sum_{x \leq z} p(z) = \sum_{x=0}^{z} p(z)$
- Continuous dist'n: $F(x) = p(X \leq x) = \int_{-\infty}^{x} p(z) dz$
- Note: In both cases, $F(max(x)) = 1$
- $F^{-1}(x)$ is the *quantile function* of X

(Cumulative) Distribution Functions

- The CDF of a distribution maps $F : X \rightarrow [0, 1]$
- Domain: possible values of X, $\subseteq \mathbb{R}^1$
- Range: $0 \leq F(x) \leq 1$
- Discrete dist'n: $F(x) = p(X \leq x) = \sum_{x \leq z} p(z) = \sum_{x=0}^{z} p(z)$
- Continuous dist'n: $F(x) = p(X \leq x) = \int_{-\infty}^{x} p(z)dz$
- Note: In both cases, $F(max(x)) = 1$
- $F^{-1}(x)$ is the *quantile function* of X

Discret distributions

- **Random Variable**: A random variable is a real-valued function defined on the sample space S.
- It assigns a real number to every outcome $s \in S$.

Discret distributions

- **Random Variable**: A random variable is a real-valued function defined on the sample space S.
- It assigns a real number to every outcome $s \in S$.
- **Discrete Random Variable**: Y is a discrete random variable if it can assume only a finite or countably infinite number of distinct values.
- Examples: number of wars per year, heads or tails, voting Republican or Democrat, number on a rolled die.

Probability Mass Function

- For a discrete random variable Y, the probability mass function (pmf) $p(x) = \Pr(X = x)$ assigns probabilities to a countable number of distinct x values such that

1. $0 \leq p(x) \leq 1$
2. $\sum_x p(x) = 1$

Example

- For one fair six-sided die, there is an equal probability of rolling any number.

Example

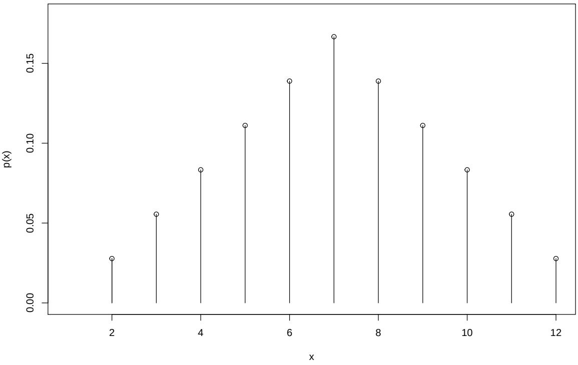
- For one fair six-sided die, there is an equal probability of rolling any number.
- Since there are six sides, the probability mass function is then $p(y) = 1/6$ for $y = 1, \ldots, 6$.

Example

- For one fair six-sided die, there is an equal probability of rolling any number.
- Since there are six sides, the probability mass function is then $p(y) = 1/6$ for $y = 1, \ldots, 6$.
- Each $p(y)$ is between 0 and 1.
- And, the sum of the $p(y)$'s is 1.

- If there are two six-sided dice, the probability mass funciton is shown below.

```
y<-c(1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1)/36; x<-c(2:12)
plot(x, y, xlim=c(1, 12), ylim=c(0, .18),
     xlab="x", ylab="p(x)")
segments(x0=x, y0=rep(0,12), x1=x, y1=y)
```



Cumulative distribution

- The cumulative distribution $F(x)$ or $\Pr(X \leq x)$ is the probability that Y is less than or equal to some value y, or

$$\Pr(X \leq x) = \sum_{i \leq x} p(i)$$

.

Cumulative distribution

- The cumulative distribution $F(x)$ or $\Pr(X \leq x)$ is the probability that Y is less than or equal to some value y, or

$$\Pr(X \leq x) = \sum_{i \leq x} p(i)$$

. The CDF must satisfy these properties:

1. $F(x)$ is non-decreasing in x.
2. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
3. $F(x)$ is right-continuous.

Cumulative distribution

- The cumulative distribution $F(x)$ or $\Pr(X \leq x)$ is the probability that Y is less than or equal to some value y, or

$$\Pr(X \leq x) = \sum_{i \leq x} p(i)$$

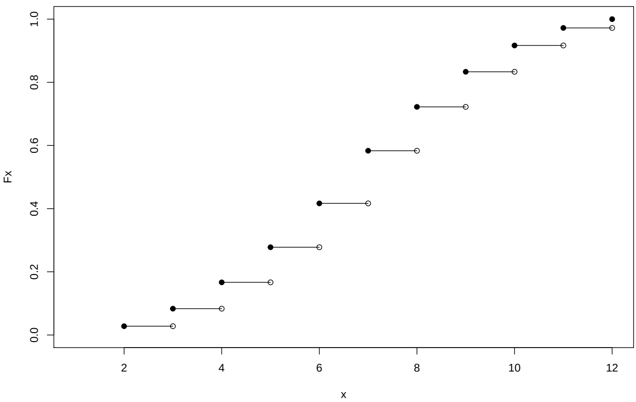
. The CDF must satisfy these properties:

1. $F(x)$ is non-decreasing in x.
2. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
3. $F(x)$ is right-continuous.

- Example: For a fair die, $\Pr(Y \leq 1) = 1/6$, $\Pr(Y \leq 3) = 1/2$, and $\Pr(Y \leq 6) = 1$.

Example: Two Fair die

```
fx<-c(1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1)/36; x<-c(2:12)
Fx<-sapply(1:11, function(i, fx) sum(fx[1:i]), fx=fx)
plot(x, Fx, xlim=c(1, 12), ylim=c(0, 1), pch=19)
points(3:12, Fx[-11], xlim=c(1, 12), ylim=c(0, 1), pch=1)
segments(x0=2:11, x1=3:12, y0=Fx[-11], y1=Fx[-11])
```



Bernoulli Distribution

- Single trial, binary outcome $0, 1$.

Bernoulli Distribution

- Single trial, binary outcome $0, 1$.
- "Prob of success in 1 trial?"

Bernoulli Distribution

- Single trial, binary outcome $0, 1$.
- "Prob of success in 1 trial?", $x \in \{0, 1\}$
- $X \sim Bern(p)$

Bernoulli Distribution

- Single trial, binary outcome $0, 1$.
- "Prob of success in 1 trial?", $x \in \{0, 1\}$
- $X \sim Bern(p)$
- $p(X = x|p) = p^x(1 - p)^{1-x}$

Bernoulli Distribution

- Single trial, binary outcome $0, 1$.
- "Prob of success in 1 trial?", $x \in \{0, 1\}$
- $X \sim Bern(p)$
- $p(X = x|p) = p^x(1-p)^{1-x}$

Political science examples

- Let $X = \begin{cases} 1 & \text{if you turnout} \\ 0 & \text{if you abstain} \end{cases}$.
  - Then, $p(X = 1|p = .4) = .4$ prob of you turning out to vote in next election, given underlying true prob $p = .4$
  - $p(X = 0|p = .4) = .6$ prob of you abstaining in next election.
- What is the probability of a of US-NKorea conflict in 2018?

Binomial distribution

- *n* $\{$*independent*$\}$, $\{$*identically distributed*$\}$ *(iid) trials, binary outcome* $0, 1$.

Binomial distribution

- *n* $\{$*independent*$\}$, $\{$*identically distributed*$\}$ *(iid) trials, binary outcome* $0, 1$.
- *"Prob of k successes in n trials?"*

Binomial distribution

- n $\{$*independent*$\}$, $\{$*identically distributed*$\}$ *(iid) trials, binary outcome* $0, 1$.
- *"Prob of* k *successes in* n *trials?"*, $k \in \{0, \ldots, n\}$

Binomial distribution

- *n* $\{$*independent*$\}$, $\{$*identically distributed*$\}$ *(iid) trials, binary outcome* $0, 1$.
- *"Prob of* k *successes in* n *trials?"*, $k \in \{0, \ldots, n\}$
- $X \sim Bin(n, p)$

Binomial distribution

- *n* $\{$*independent*$\}$, $\{$*identically distributed*$\}$ *(iid) trials, binary outcome* $0, 1$.
- *"Prob of k successes in n trials?"*, $k \in \{0, \ldots, n\}$
- $X \sim Bin(n, p)$
- $p(X = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$

Binomial distribution

- *n* {*independent*}, {*identically distributed*} *(iid) trials, binary outcome* $0, 1$.
- *"Prob of k successes in n trials?"*, $k \in \{0, \ldots, n\}$
- $X \sim Bin(n, p)$
- $p(X = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$
- $X \sim Bin(1, p) \sim p(X = k|1, p) \sim Bern(p)$

Binomial distribution

- n {*independent*}, {*identically distributed*} *(iid) trials, binary outcome* $0, 1$.
- *"Prob of k successes in n trials?"*, $k \in \{0, \ldots, n\}$
- $X \sim Bin(n, p)$
- $p(X = k|n, p) = \binom{n}{k} p^k (1 - p)^{n-k}$
- $X \sim Bin(1, p) \sim p(X = k|1, p) \sim Bern(p)$
- $X_1 \sim Bin(n_1, p)$, $X_2 \sim Bin(n_2, p)$, $X_1 \perp\!\!\!\perp X_2$, *then* $X_1 + X_2 \sim Bin(n_1 + n_2, p)$

Binomial distribution

- *n {independent}, {identically distributed} (iid) trials, binary outcome* $0, 1$.
- *"Prob of k successes in n trials?"*, $k \in \{0, \ldots, n\}$
- $X \sim Bin(n, p)$
- $p(X = k | n, p) = \binom{n}{k} p^k (1-p)^{n-k}$
- $X \sim Bin(1, p) \sim p(X = k | 1, p) \sim Bern(p)$
- $X_1 \sim Bin(n_1, p)$, $X_2 \sim Bin(n_2, p)$, $X_1 \perp\!\!\!\perp X_2$, *then* $X_1 + X_2 \sim Bin(n_1 + n_2, p)$
- $F(x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$

Binomial distribution

- n {*independent*}, {*identically distributed*} (*iid*) *trials, binary outcome* $0, 1$.
- *"Prob of k successes in n trials?"*, $k \in \{0, \dots, n\}$
- $X \sim Bin(n, p)$
- $p(X = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$
- $X \sim Bin(1, p) \sim p(X = k|1, p) \sim Bern(p)$
- $X_1 \sim Bin(n_1, p)$, $X_2 \sim Bin(n_2, p)$, $X_1 \perp\!\!\!\perp X_2$, *then* $X_1 + X_2 \sim Bin(n_1 + n_2, p)$
- $F(x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$
- *Sum of n Bernoullis*

Political examples:

- Prob 3 of 6 opposing Senators support an amendment:

Political examples:

- Prob 3 of 6 opposing Senators support an amendment: $p(X = 3|n = 6, p = .3) =$

Political examples:

- Prob 3 of 6 opposing Senators support an amendment:
  $p(X = 3|n = 6, p = .3) =$ `dbinom(3,6,prob=.3)` $\approx .19$

Political examples:

- Prob 3 of 6 opposing Senators support an amendment: $p(X = 3|n = 6, p = .3) =$ `dbinom(3,6,prob=.3)` $\approx .19$
- Prob ≥ 3 of 6 opposing Senators support an amendment:

Political examples:

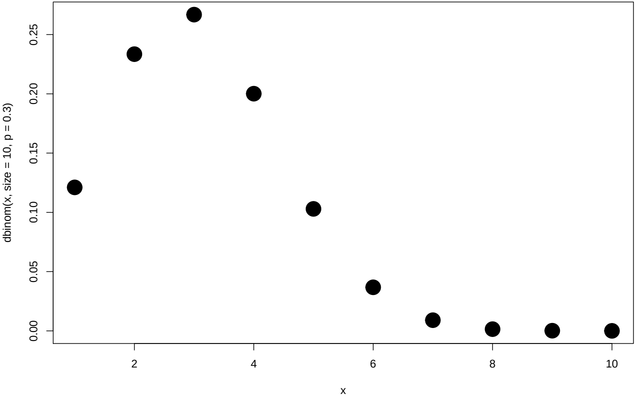
- Prob 3 of 6 opposing Senators support an amendment: $p(X = 3|n = 6, p = .3) =$ `dbinom(3,6,prob=.3)` $\approx .19$
- Prob ≥ 3 of 6 opposing Senators support an amendment: $p(X \geq 3|n = 6, p = .3) =$

Political examples:

- Prob 3 of 6 opposing Senators support an amendment: $p(X = 3|n = 6, p = .3) =$ `dbinom(3,6,prob=.3)` $\approx .19$
- Prob ≥ 3 of 6 opposing Senators support an amendment: $p(X \geq 3|n = 6, p = .3) =$ `1-pbinom(2,6,prob=.3)` $=$ `pbinom(2,6,prob=.3,lower.tail=FALSE)` $\approx .26$

Binomial PMF: N=10, p=0.3

```
x<-1:10
plot(x, dbinom(x, size=10, p=.3), cex=3, pch=19)
```



Poisson Distribution

- iid Bernoulli trial at every instant

Poisson Distribution

- iid Bernoulli trial at every instant
- "Prob of k successes in unit time?"

Poisson Distribution

- iid Bernoulli trial at every instant
- "Prob of k successes in unit time?", $k \in \mathbb{N}$

Poisson Distribution

- iid Bernoulli trial at every instant
- "Prob of k successes in unit time?", $k \in \mathbb{N}$
- $\lambda > 0$, expected number of successes/events in unit time

Poisson Distribution

- iid Bernoulli trial at every instant
- "Prob of k successes in unit time?", $k \in \mathbb{N}$
- $\lambda > 0$, expected number of successes/events in unit time
- $X \sim \mathit{Pois}(\lambda)$

Poisson Distribution

- iid Bernoulli trial at every instant
- "Prob of k successes in unit time?", $k \in \mathbb{N}$
- $\lambda > 0$, expected number of successes/events in unit time
- $X \sim Pois(\lambda)$
- $p(X = k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$

Poisson Distribution

- iid Bernoulli trial at every instant
- "Prob of k successes in unit time?", $k \in \mathbb{N}$
- $\lambda > 0$, expected number of successes/events in unit time
- $X \sim Pois(\lambda)$
- $p(X = k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$
- $X_1 \sim Pois(\lambda_1)$, $X_2 \sim Pois(\lambda_2)$, $X_1 \perp\!\!\!\perp X_2$, then $X_1 + X_2 \sim Pois(\lambda_1 + \lambda_2)$

Poisson Distribution

- iid Bernoulli trial at every instant
- "Prob of k successes in unit time?", $k \in \mathbb{N}$
- $\lambda > 0$, expected number of successes/events in unit time
- $X \sim Pois(\lambda)$
- $p(X = k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$
- $X_1 \sim Pois(\lambda_1)$, $X_2 \sim Pois(\lambda_2)$, $X_1 \perp\!\!\!\perp X_2$, then $X_1 + X_2 \sim Pois(\lambda_1 + \lambda_2)$
- $F(x|\lambda) = \sum_{z=0}^{x} \frac{e^{-\lambda}\lambda^z}{z!}$

Poisson Distribution

- iid Bernoulli trial at every instant
- "Prob of k successes in unit time?", $k \in \mathbb{N}$
- $\lambda > 0$, expected number of successes/events in unit time
- $X \sim Pois(\lambda)$
- $p(X = k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$
- $X_1 \sim Pois(\lambda_1)$, $X_2 \sim Pois(\lambda_2)$, $X_1 \perp\!\!\!\perp X_2$, then $X_1 + X_2 \sim Pois(\lambda_1 + \lambda_2)$
- $F(x|\lambda) = \sum_{z=0}^{x} \frac{e^{-\lambda}\lambda^z}{z!}$

Political examples

- Political examples:
  - prob of 2 HoR censures this Congress, when usually 4:

Political examples

- Political examples:
  - prob of 2 HoR censures this Congress, when usually 4:
    $p(X = 2|\lambda = 4) =$

Political examples

- Political examples:
  - prob of 2 HoR censures this Congress, when usually 4: $p(X = 2|\lambda = 4) =$ `dpois(2,4)`

Political examples

- Political examples:
  - prob of 2 HoR censures this Congress, when usually 4: $p(X = 2|\lambda = 4) =$ `dpois(2,4)` $\approx .15$

Political examples

- Political examples:
  - prob of 2 HoR censures this Congress, when usually 4: $p(X = 2|\lambda = 4) = \texttt{dpois(2,4)} \approx .15$
  - prob of $\geq$ 2 HoR censures this Congress, when usually 4:

Political examples

- Political examples:
  - prob of 2 HoR censures this Congress, when usually 4: $p(X = 2|\lambda = 4) = \texttt{dpois(2,4)} \approx .15$
  - prob of $\geq$ 2 HoR censures this Congress, when usually 4: $p(X \geq 2|\lambda = 4) =$

Political examples

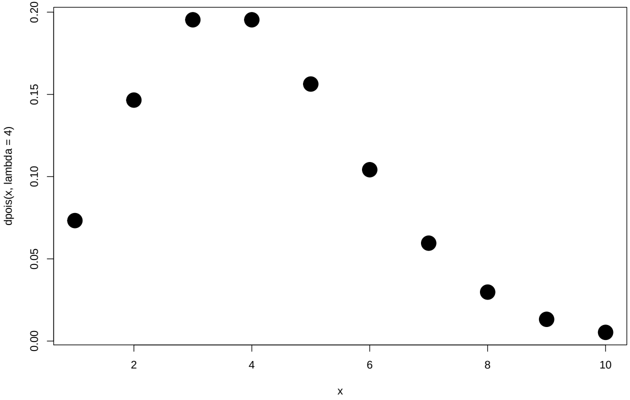
- Political examples:
  - prob of 2 HoR censures this Congress, when usually 4: $p(X = 2|\lambda = 4) = \texttt{dpois(2,4)} \approx .15$
  - prob of $\geq$ 2 HoR censures this Congress, when usually 4: $p(X \geq 2|\lambda = 4) = \texttt{1-ppois(1,4)}$

Political examples

- Political examples:
  - prob of 2 HoR censures this Congress, when usually 4: $p(X = 2|\lambda = 4) = \texttt{dpois(2,4)} \approx .15$
  - prob of ≥ 2 HoR censures this Congress, when usually 4: $p(X \geq 2|\lambda = 4) = \texttt{1-ppois(1,4)} \approx .91$

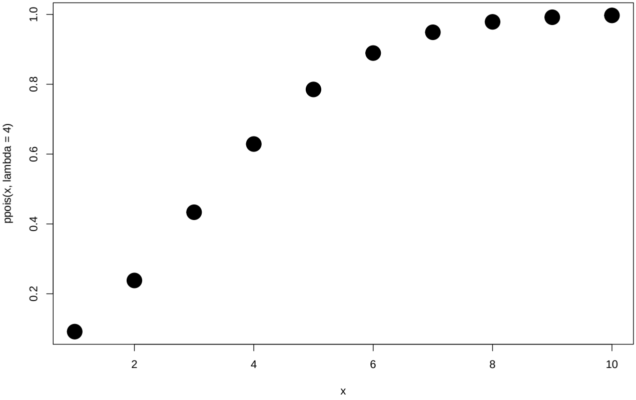
PMF for Poisson with $\lambda = 4$

```r
x<-1:10
plot(x, dpois(x, lambda=4), cex=3, pch=19)
```



CDF for Poisson with $\lambda = 4$

```
x<-1:10
plot(x, ppois(x, lambda=4), cex=3, pch=19)
```



Other distributions you might encounter

- Negative Binomial
- Geometric
- HyperGeometric
- Multinomial (The dice examples)

Continuous Random Variables:

- X is a continuous random variable if there exists a nonnegative function $f(x)$ defined for all real $y \in (-\infty, \infty)$, such that for any interval A,
$$\Pr(x \in A) = \int_A f(x)dx$$
- Examples: income, GNP, temperature

Probability Density function

- The function f above is called the probability density function (pdf) of x and must satisfy
    1. $f(x) \geq 0$
    2. $\int_{-\infty}^{\infty} f(x)dx = 1$

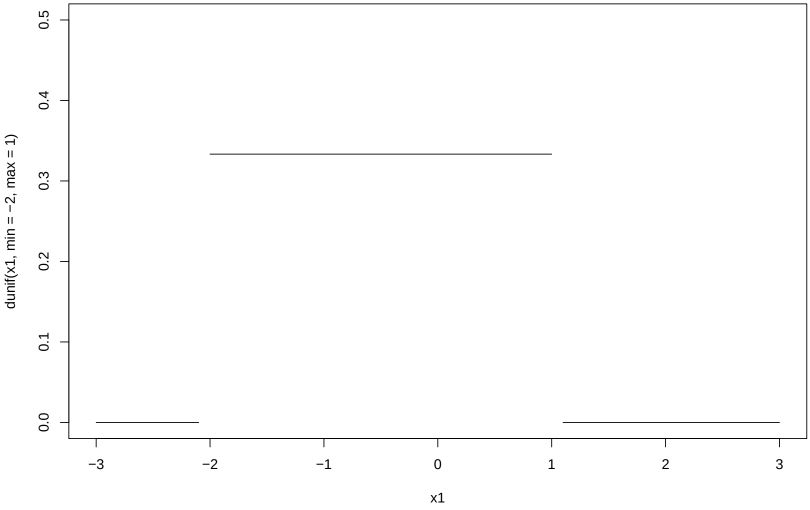
Probability Density function

- The function f above is called the probability density function (pdf) of x and must satisfy
  1. $f(x) \geq 0$
  2. $\int\limits_{-\infty}^{\infty} f(x)dx = 1$
- Note also that $\Pr(X = x) = 0$ — i.e., the probability of any point x is zero.

Example: Uniform distribution

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

PDF for Uniform(-2, 1)



Cumulative Distribution

- Because the probability that a continuous random variable will assume any particular value is zero, we can only make statements about the probability of a continuous random variable being within an interval.

Cumulative Distribution

- Because the probability that a continuous random variable will assume any particular value is zero, we can only make statements about the probability of a continuous random variable being within an interval.
- The cumulative distribution gives the probability that X lies on the interval $(-\infty, x)$ and is defined as

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(s)ds$$

Cumulative Distribution

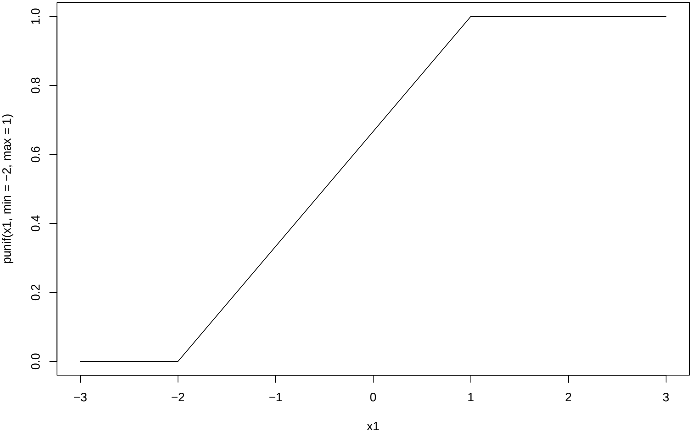
- Because the probability that a continuous random variable will assume any particular value is zero, we can only make statements about the probability of a continuous random variable being within an interval.
- The cumulative distribution gives the probability that X lies on the interval $(-\infty, x)$ and is defined as

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(s)ds$$

- Note that $F(x)$ has similar properties with continuous distributions as it does with discrete
  - non-decreasing, continuous (not just right-continuous),
  - and $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

CDF for Uniform(-2, 1)

```
x1<-seq(-3, 3, by=.1)
plot(x1, punif(x1, min=-2, max=1), cex=2, pch=19,
     type="l", xlim=c(-3,3), ylim=c(0, 1))
```



- Similarly, we can also make probability statements about X falling in an interval $a \leq x \leq b$.

$$\Pr(a \leq x \leq b) = \int_{a}^{b} f(x) dy$$

- Similarly, we can also make probability statements about X falling in an interval $a \leq x \leq b$.

$$\Pr(a \leq x \leq b) = \int_a^b f(x)dy$$

Example: $f(x) = 1, \quad 0 < x < 1$. Find $F(x)$ and $\Pr(.5 < y < .75)$.

- Similarly, we can also make probability statements about X falling in an interval $a \leq x \leq b$.

$$\Pr(a \leq x \leq b) = \int\limits_a^b f(x)dy$$

Example: $f(x) = 1, \quad 0 < x < 1$. Find $F(x)$ and $\Pr(.5 < y < .75)$.

$$F(y) = \int\limits_0^y f(s)ds = \int_0^y 1ds = s|_0^y = y$$

- Similarly, we can also make probability statements about X falling in an interval $a \leq x \leq b$.

$$\Pr(a \leq x \leq b) = \int\limits_{a}^{b} f(x)dy$$

Example: $f(x) = 1, \quad 0 < x < 1$. Find $F(x)$ and $\Pr(.5 < y < .75)$.

$$F(y) = \int\limits_{0}^{y} f(s)ds = \int_0^y 1ds = s|_0^y = y$$

$$\Pr(.5 < y < .75) = \int_{.5}^{.75} 1ds = s|_{.5}^{.75} = .25$$

- Finally, note that:

$$F'(y) = \frac{dF(y)}{dy} = f(y)$$

Uniform Distribution

- $X \sim Unif(a, b)$

Uniform Distribution

- $X \sim Unif(a, b)$
- $x \in [a, b]$

Uniform Distribution

- $X \sim Unif(a, b)$
- $x \in [a, b]$
- $p(x) = \frac{1}{b-a}$

Uniform Distribution

- $X \sim Unif(a, b)$
- $x \in [a, b]$
- $p(x) = \frac{1}{b-a}$
- $F(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \frac{x-a}{b-a} & \text{for} \quad a \leq x < b \\ 1 & \text{for} \quad x \geq b \end{cases}$

Uniform Distribution

- $X \sim Unif(a, b)$
- $x \in [a, b]$
- $p(x) = \frac{1}{b-a}$
- $F(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \frac{x-a}{b-a} & \text{for} \quad a \leq x < b \\ 1 & \text{for} \quad x \geq b \end{cases}$
- Common: $X \sim Unif(0, 1)$

Uniform Distribution

- $X \sim Unif(a, b)$
- $x \in [a, b]$
- $p(x) = \frac{1}{b-a}$
- $F(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \frac{x-a}{b-a} & \text{for} \quad a \leq x < b \\ 1 & \text{for} \quad x \geq b \end{cases}$
- Common: $X \sim Unif(0, 1)$

Political examples:

- “Suppose voter’s probability of turnout is draw from uniform”
- Suppose the ideology of an agent is drawn from the uniform distribution

Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2)$

Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2)$
- $x \in \mathbb{R}$

Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $\mu \in \mathbb{R}$

Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $\mu \in \mathbb{R}$
- $\sigma > 0$

Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $\mu \in \mathbb{R}$
- $\sigma > 0$
- $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $\mu \in \mathbb{R}$
- $\sigma > 0$
- $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Common: $X \sim N(0, 1)$, "standard normal"

Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $\mu \in \mathbb{R}$
- $\sigma > 0$
- $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Common: $X \sim N(0, 1)$, "standard normal"
- $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $\mu \in \mathbb{R}$
- $\sigma > 0$
- $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Common: $X \sim N(0, 1)$, "standard normal"
- $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- $\Phi(x) =$ standard normal CDF – that no one knows

Normal (Gaussian) Distribution

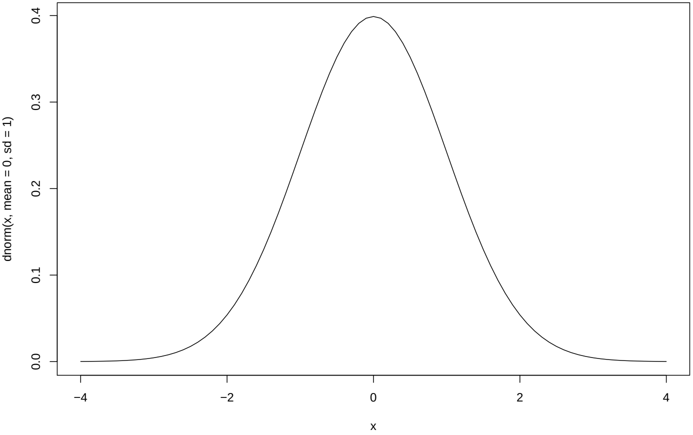
- $X \sim N(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $\mu \in \mathbb{R}$
- $\sigma > 0$
- $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Common: $X \sim N(0, 1)$, "standard normal"
- $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- $\Phi(x) =$ standard normal CDF – that no one knows

Political examples:

- population quantities, asymptotic/known variance sampling distributions
- $\Phi(x) = p(X = 1)$ is the basic *probit model*

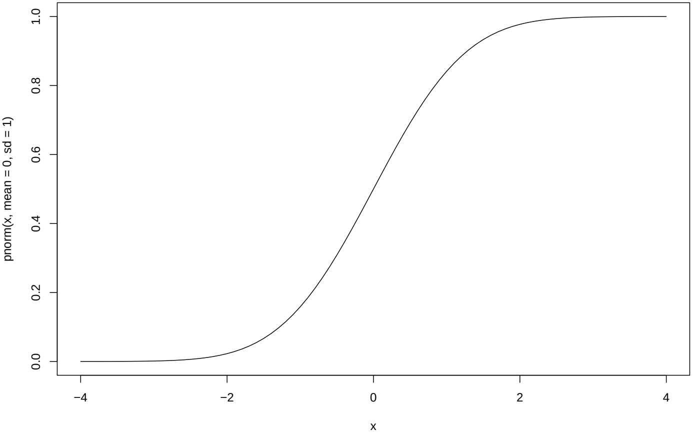
PDF for standard normal distribution

```
x<-seq(-4, 4, by=.1)
plot(x, dnorm(x, mean=0, sd=1),  type="l")
```



CDF for standard normal distribution

```
x<-seq(-4, 4, by=.1)
plot(x, pnorm(x, mean=0, sd=1),  type="l")
```



Student's t Distribution

- $X \sim t_n(\mu, \sigma^2)$

Student's t Distribution

- $X \sim t_n(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $n \in \mathbb{Z}_+$, often called degrees of freedom
- $p(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sigma\sqrt{n\pi}} \left(1 + \frac{1}{n}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{n+1}{2}}$
- Common: $t_n \sim \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$

Student's t Distribution

- $X \sim t_n(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $n \in \mathbb{Z}_+$, often called degrees of freedom
- $p(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sigma\sqrt{n\pi}} \left(1 + \frac{1}{n}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{n+1}{2}}$
- Common: $t_n \sim \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$
- $t_1 \sim$ *Cauchy* (a very dangerous distribution)

Student's t Distribution

- $X \sim t_n(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $n \in \mathbb{Z}_+$, often called degrees of freedom
- $p(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sigma\sqrt{n\pi}} \left(1 + \frac{1}{n}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{n+1}{2}}$
- Common: $t_n \sim \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$
- $t_1 \sim$ *Cauchy* (a very dangerous distribution)
- $t_{large} \sim N(0,1)$

Student's t Distribution

- $X \sim t_n(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $n \in \mathbb{Z}_+$, often called degrees of freedom
- $p(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sigma\sqrt{n\pi}} \left(1 + \frac{1}{n}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{n+1}{2}}$
- Common: $t_n \sim \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$
- $t_1 \sim$ *Cauchy* (a very dangerous distribution)
- $t_{large} \sim N(0,1)$
- If $X \sim N(0,1)$, $Y \sim \chi^2_n$, $X \perp\!\!\!\perp Y$, then $\frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n$

Student's t Distribution

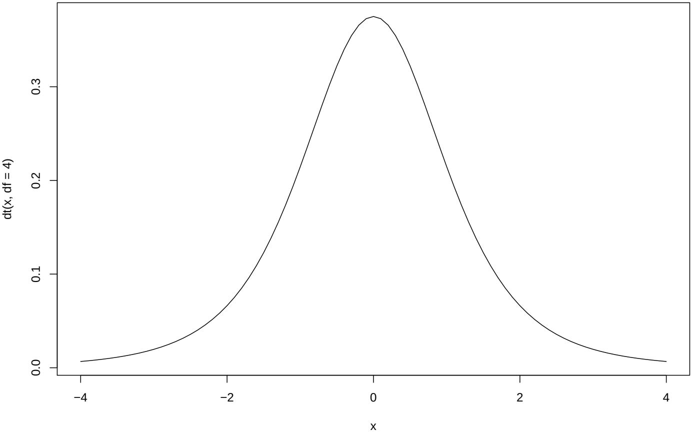
- $X \sim t_n(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $n \in \mathbb{Z}_+$, often called degrees of freedom
- $p(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sigma\sqrt{n\pi}} \left(1 + \frac{1}{n}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{n+1}{2}}$
- Common: $t_n \sim \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$
- $t_1 \sim$ *Cauchy* (a very dangerous distribution)
- $t_{large} \sim N(0, 1)$
- If $X \sim N(0,1)$, $Y \sim \chi^2_n$, $X \perp\!\!\!\perp Y$, then $\frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n$

Political examples

- Finite sample/unknown variance distributions
- robust estimation

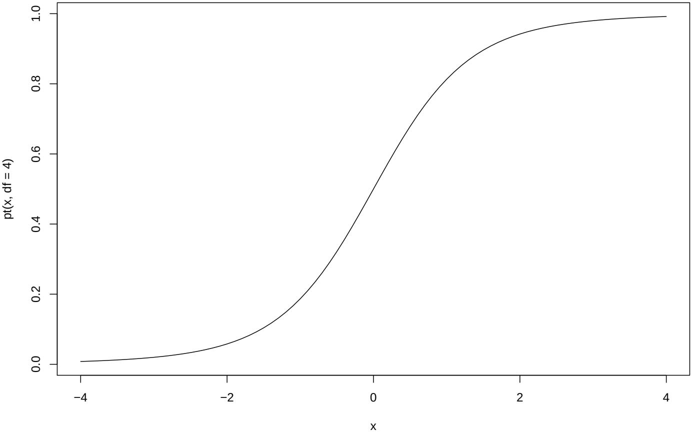
PDF for Student-t with df=4

```r
x<-seq(-4, 4, by=.1)
plot(x, dt(x, df=4),  type="l")
```



CDF for Student-t with df=4

```
x<-seq(-4, 4, by=.1)
plot(x, pt(x, df=4),  type="l")
```



Other distributions you may encounter

- Exponential distribution
- χ^2 distribution
- Logistic distribution
- Beta distribution
- F(isher's) distribution
- Gamma distribution
- Laplace distribution
- Weibull disribution
- Log-normal distribution
- Pareto distribution
- Dirichlet distribution

Joint distributions

- Often, we are interested in two or more random variables defined on the same sample space.
- The distribution of these variables is called a **joint distribution**.

Joint distributions

- Often, we are interested in two or more random variables defined on the same sample space.
- The distribution of these variables is called a **joint distribution**.
- Joint distributions can be made up of any combination of discrete and continuous random variables.

Example

- Suppose we are interested in the outcomes of flipping a coin and rolling a 6-sided die at the same time.

Example

- Suppose we are interested in the outcomes of flipping a coin and rolling a 6-sided die at the same time.
- The sample space for this process contains 12 elements:

$$\{h1, h2, h3, h4, h5, h6, t1, t2, t3, t4, t5, t6\}$$

Example

- Suppose we are interested in the outcomes of flipping a coin and rolling a 6-sided die at the same time.
- The sample space for this process contains 12 elements:

$$\{h1, h2, h3, h4, h5, h6, t1, t2, t3, t4, t5, t6\}$$

- We can define two random variables X and Y such that $X = 1$ if heads and $X = 0$ if tails, while Y equals the number on the die.

Example

- Suppose we are interested in the outcomes of flipping a coin and rolling a 6-sided die at the same time.
- The sample space for this process contains 12 elements:

$$\{h1, h2, h3, h4, h5, h6, t1, t2, t3, t4, t5, t6\}$$

- We can define two random variables X and Y such that $X = 1$ if heads and $X = 0$ if tails, while Y equals the number on the die.
- We can then make statements about the joint distribution of X and Y.

Joint discret random variables

- If both X and Y are discrete, their joint probability mass function assigns probabilities to each pair of outcomes

$$p(x, y) = \Pr(X = x, Y = y)$$

- Again, $p(x, y) \in [0, 1]$ and $\sum \sum p(x, y) = 1$.

Marginal pmf

- If we are interested in the marginal probability of one of the two variables (ignoring information about the other variable), we can obtain the marginal pmf by summing across the variable that we don't care about:

$$p_X(x) = \sum_i p(x, y_i)$$

Conditional pmf

- We can also calculate the conditional pmf for one variable, holding the other variable fixed.
- Recalling from the previous lecture that $\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$, we can write the conditional pmf as

$$p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)}, \quad p_X(x) > 0$$

Joint continuous random variables

- If both X and Y are continuous, their joint probability density function defines their distribution:

$$\Pr((X, Y) \in A) = \iint_A f(x, y)dxdy$$

- Likewise, $f(x, y) \geq 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$.

Marginal pdf

- Instead of summing, we obtain the marginal probability density function by integrating out one of the variables:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

Conditional pdf

- Finally, we can write the conditional pdf as

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, \quad f_X(x) > 0$$

Class business

- Problem set 1 will be given out on Thursday. You will have one week.
- It should not be hard, it will be long.
- PROOFS
- Next class will cover Wasserman Chpts 3 and 5 (but carefully read the very short Chapter 4.)