Gov 2000 - 3. Random Variables and Probability Distributions

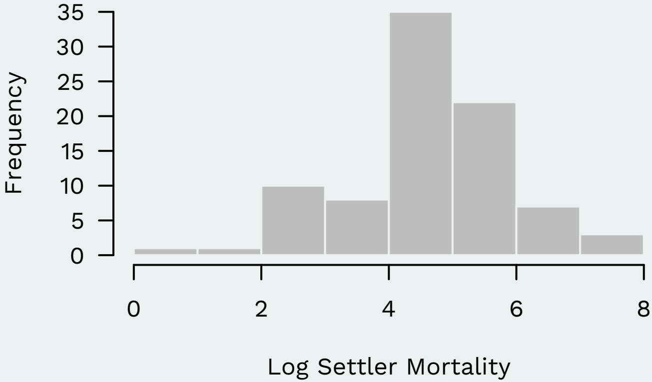
Matthew Blackwell

September 29, 2015

Where are we? Where are we going?

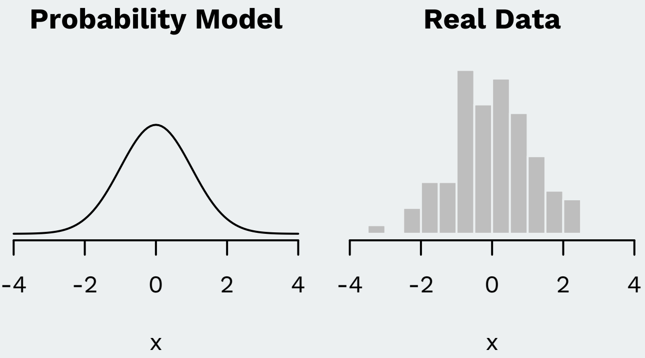
- We've learned about the fundamentals of probability and worked with events.
- Now we are going to focus on random variables and their distributions.
- Variable are roughly columns in a spreadsheet.
- We want to learn about the distribution of these variables.

Motivation



- We want to learn about the distribution of random variables like this one.
- First, we're going to build a probability model for the random variable.
- Once we know how random variables work in theory, we will move onto using real data to learn about them.

Flow



- Today: if we know the probability distribution, what data should we expect?
- Inference: how do we learn about the underlying probability distribution?

Outline for today

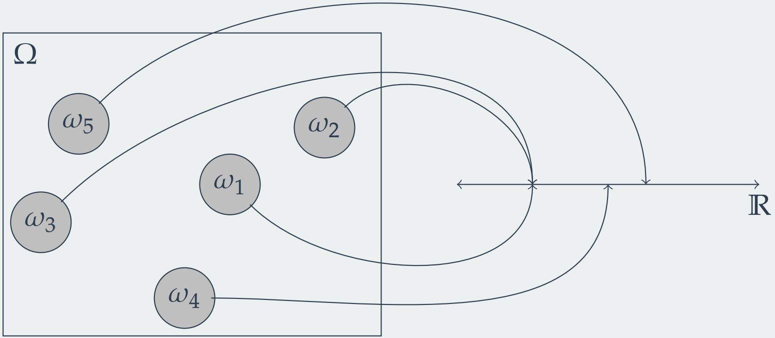
1. How do we define random variables?
2. How do we connect random variables to probabilities?
3. What are some properties of random variables we might want to learn?

1/ Random Variables

What are random variables?

- A random variable (r.v.) is a function that maps from the sample space of an experiment to the real line or $X : \Omega \to \mathbb{R}$.
- Example: tossing a coin 5 times
  - one possible outcome: $\omega = HTHTT$, but not a random variable because it's not numeric.
  - $X(\omega)$ = number of heads in the five tosses
  - $X(HTHTT) = 2$
- Even though X is a function of ω, we will often just write the r.v. as X.
  - Almost always use capital roman letters for r.v.s
  - Lower-case roman x used to refer to arbitrary value that X might take.
- More than 1 r.v. available for each sample space:
  - $Y = 5 - X$ is the number of tails in the five tosses.

Random variables in pictures



Examples

- Turning out to vote for one person:
  - $\Omega = \{\text{voted}, \text{didn't vote}\}$.
  - Random variable converts this into a number:

$$X = \begin{cases} 1 \text{ if voted} \\ 0 \text{ if didn't vote} \end{cases}$$

  - Called a **Bernoulli** or **binary** random variable.
- Length of government in a parliamentary system:
  - $\Omega = [0, \infty) \rightsquigarrow$ already numeric
  - X might be equal to the outcome: $X(\omega) = \omega$.

Outcomes to random variables

- Suppose that two people decide independently to vote (V) or abstain (A).
- Each person has a probability 0.5 of voting.
- Let X be the number of these two that turn out to vote.

ω	$\mathbb{P}(\{\omega\})$	$X(\omega)$
AA	$1/4$	0
AV	$1/4$	1
VA	$1/4$	1
VV	$1/4$	2

x	$\mathbb{P}(X = x)$
0	$1/4$
1	$1/2$
2	$1/4$

Why random variables?

- Statistics is about data and data is numeric.
  - Need r.v.s to bridge the gap between sample spaces and statistics.
- How are r.v.s random?
  - Numerical summaries of uncertain events.
  - Uncertainty over which of Ω will occur induces uncertainty over what X will equal.

Types of random variables

Two types of random variables:

1. **Discrete random variable:**
   - A r.v. that only takes on a finite or countably infinite number of values.
   - Countably infinite $\rightsquigarrow$ takes on integer values, but no upper bound.
2. **Continuous random variable:**
   - A r.v. that can take on any real value.
   - $\rightsquigarrow$ uncountably infinite number of possible realizations.

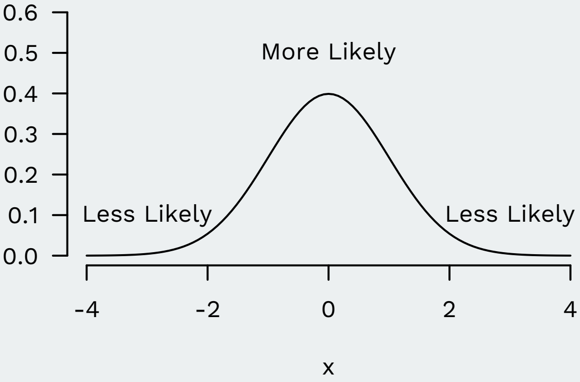
Examples

- Discrete r.v.s:
  - Number of Democrats who win election in the Senate
  - An indicator of whether two countries go to war
  - The number of times a particular word is used in a document
- Continuous r.v.s:
  - The length of time between two governments in a parliamentary system
  - The proportion of voters who turned out
  - Budgets allocations to various government programs

2/ Probability Distributions

What is a probability distribution?

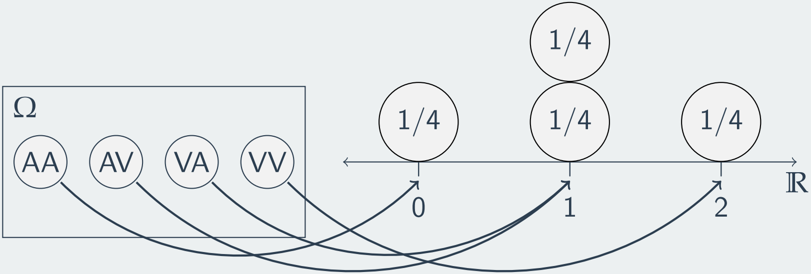
- Uncertainty over the values of the r.v.s.
- Probability over outcomes $\rightsquigarrow$ some values more likely than others
- The probability distribution of a r.v. describes the likelihood of all of the possible values.



Where do the probability distributions come from?

- Probabilities on Ω induce probabilities for X
  - Independent fair coin flips so that $\mathbb{P}(H) = 0.5$
  - Then if $X = 1$ for heads, $\mathbb{P}(X = 1) = 0.5$
- Remember that the DGP induces the probabilities for Ω
- Often we'll skip the definition of Ω and directly connect the DGP and a r.v.
- Goal of statistics is often to learn about the distribution of X.
- Easier to talk about discrete r.v.s and continuous r.v.s separately.

Inducing probabilities



- Let X be the number of these two that turn out to vote.

ω	$\mathbb{P}(\{\omega\})$	$X(\omega)$
AA	1/4	0
AV	1/4	1
VA	1/4	1
VV	1/4	2

x	$\mathbb{P}(X = x)$
0	1/4
1	1/2
2	1/4

Probability mass function

- For a discrete r.v., each possible outcome has an associated probability of occurring.
- Need to summarize the probability distribution.
- Probability mass function: for a discrete random variable, X, we can define the probability mass function (p.m.f.) as:

$$f_X(x) = \mathbb{P}(X = x)$$

- Some properties of the p.m.f. (from probability):

$$0 \leq f_X(x) \leq 1 \qquad \sum_{i=1}^{k} f_X(x_j) = 1$$

- Can't use this definition for a continuous r.v.

Example – random assignment to treatment

- You want to run a randomized control trial on 3 people.
- Use the following procedure:
  - Flip independent fair coins for each unit
  - Heads assigned to Control (C), tails to Treatment (T)
- Let X be the number of treated units:

$$X = \begin{cases} 0 & \text{if } (C, C, C) \\ 1 & \text{if } (T, C, C) \text{ or } (C, T, C) \text{ or } (C, C, T) \\ 2 & \text{if } (T, T, C) \text{ or } (C, T, T) \text{ or } (T, C, T) \\ 3 & \text{if } (T, T, T) \end{cases}$$

- Use independence and fair coins:

$$\mathbb{P}(C, T, C) = \mathbb{P}(C)\mathbb{P}(T)\mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

Calculating the p.m.f.

$$f_X(0) = \mathbb{P}(X = 0) = \mathbb{P}(C, C, C) = \frac{1}{8}$$

$$f_X(1) = \mathbb{P}(X = 1) = \mathbb{P}(T, C, C) + \mathbb{P}(C, T, C) + \mathbb{P}(C, C, T) = \frac{3}{8}$$

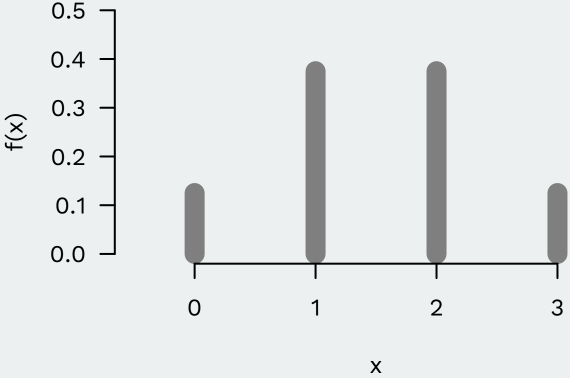
$$f_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(T, T, C) + \mathbb{P}(C, T, T) + \mathbb{P}(T, C, T) = \frac{3}{8}$$

$$f_X(3) = \mathbb{P}(X = 3) = \mathbb{P}(T, T, T) = \frac{1}{8}$$

- What's $\mathbb{P}(X = 4)$? 0!

Plotting the p.m.f.

- We could plot this p.m.f. using R:



- **Question**: Does this seem like a good way to assign treatment? What is one major problem with it?

Named distributions

- Some distributions are so common that we give them names.
  - Uniform, exponential, normal, poisson, etc.
- Usually these are a family of distributions.
  - The p.m.f. within the family has the same form, with parameters that might vary across the family.
  - The parameters determine the shape of the distribution
- Most modeling in statistics looks like this:
  1. Assume the data come from a particular family of distributions (say, normal or poisson)
  2. Assume that the particular parameters are unknown
  3. Use the data to infer what the parameters are.

Bernoulli distribution

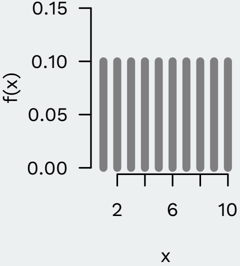
- Some distributions are so common we give them names.
- Bernoulli distribution describes binary r.v.s
- Suppose X is binary with $\mathbb{P}(X = 1) = p$
- Then, for $x \in \{0, 1\}$, the p.m.f. is:

$$f_X(x) = p^x(1 - p)^{1-x}$$

- A family of distributions with parameter p.

$$f_X(1) = p^1(1 - p)^0 = p$$
$$f_X(0) = p^0(1 - p)^1 = 1 - p$$

Discrete uniform distribution



- Equal probability of any value of X:

$$f_X(x) = \begin{cases} 1/k & \text{for } x = 1, \dots, k \\ 0 & \text{otherwise} \end{cases}$$

- Justified from the DGP of random sampling.

Cumulative distribution functions

- Other ways to write down the distribution? Yes!
- The cumulative distribution function (c.d.f.) returns the probability is that a variable is less than a particular realization:

$$F_X(x) \equiv \mathbb{P}(X \leq x).$$

- If r.v.s X and Y have the same c.d.f., then we say they have the same distribution.
    - $\rightsquigarrow$ the p.m.f. for X is the same as for Y.
- We can write the c.d.f. of a discrete r.v. is:

$$F_X(x) = \sum_{x_k \leq x} f_X(x_k)$$

Example of c.d.f

- Remember example where X is the number turning out (of 2) from earlier:

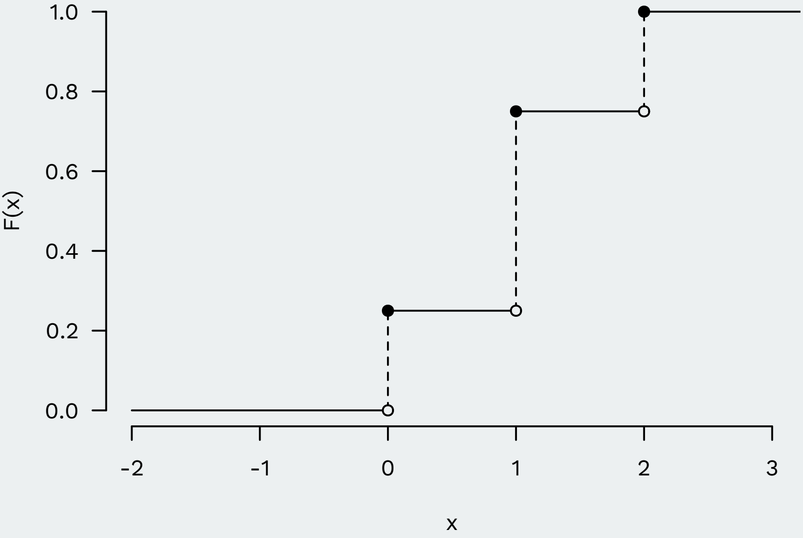
x	$\mathbb{P}(X = x)$
0	1/4
1	1/2
2	1/4

- Let's calculate the c.d.f., $F_X(x) = \mathbb{P}(X \leq x)$ for this:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

- What is $F_X(1.4)$ here? 0.75

Graph of example c.d.f.



Properties of the c.d.f.

- A couple of properties:
  1. never decreases,
  2. limits to 0 toward negative infinity, limits to 1 toward positive infinity,
  3. right-continuous (no jumps when we approach a point from the right)
- We can use the c.d.f. to calculate the probability of other events, such as:

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$$

- Example with 2 voter election:

$$\mathbb{P}(0 < X \leq 2) = F_X(2) - F_X(0) = 1 - 0.25 = 0.75$$

Continuous random variables

- What about continuous r.v.s?
- Can we just specify $\mathbb{P}(X = x)$?
- No! Continuous r.v.s $\rightsquigarrow$ uncountably infinite number of possible values.
  - Suppose $\mathbb{P}(X = x) = \varepsilon$ for $x \in (0, 1)$ where ε is a very small number.
  - What's the probability of being between 0 and 1?
  - There are an infinite number of real numbers between 0 and 1:

$$0.232879873\ldots \qquad 0.57263048743\ldots \qquad 0.9823612984\ldots$$

  - Each one has probability $\varepsilon \rightsquigarrow \mathbb{P}(X \in (0, 1)) = \infty \times \varepsilon = \infty$
- But $\mathbb{P}(X \in (0, 1))$ must be less than 1!
- $\rightsquigarrow \mathbb{P}(X = x)$ must be 0.

Thought experiment: draw a random real value between 0 and 10. What's the probability that we draw a value that is exact equal to π?

3.1415926535 8979323846 2643383279 5028841971 6939937510 5820974944
5923078164 0628620899 8628034825 3421170679 8214808651 3282306647 0938446095
5058223172 5359408128 4811174502 8410270193 8521105559 6446229489 5493038196
4428810975 6659334461 2847564823 3786783165 2712019091 4564856692 3460348610
4543266482 1339360726 0249141273 7245870066 0631558817 4881520920 9628292540
9171536436 7892590360 0113305305 4882046652 1384146951 9415116094 3305727036
5759591953 0921861173 8193261179 3105118548 0744623799 6274956735 1885752724
8912279381 8301194912 9833673362 4406566430 8602139494 6395224737 1907021798
6094370277 0539217176 2931767523 8467481846 7669405132 0005681271 4526356082
7785771342 7577896091 7363717872 1468440901 2249534301 4654958537 1050792279
6892589235 4201995611 2129021960 8640344181 5981362977 4771309960 5187072113
4999999837 2978049951 0597317328 1609631859 5024459455 3469083026 4252230825
3344685035 2619311881 7101000313 7838752886 5875332083 8142061717 7669147303
5982534904 2875546873 1159562863 8823537875 9375195778 1857780532 1712268066
1300192787 6611195909 2164201989 3809525720 1065485863 2788659361 5338182796
8230301952 0353018529 6899577362 2599413891 2497217752 8347913151 5574857242

Probability density functions

- By this logic, $\mathbb{P}(X = x) = 0$ for a continuous r.v.
  - Take-away: can't use a p.m.f. for these variables.
- Will use an alternative function that can answer many of the same questions.
- The probability density function (p.d.f.) for a continuous random variable, X as the function $f_X(x) \geq 0$ for all x and $a \leq b$:

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx.$$

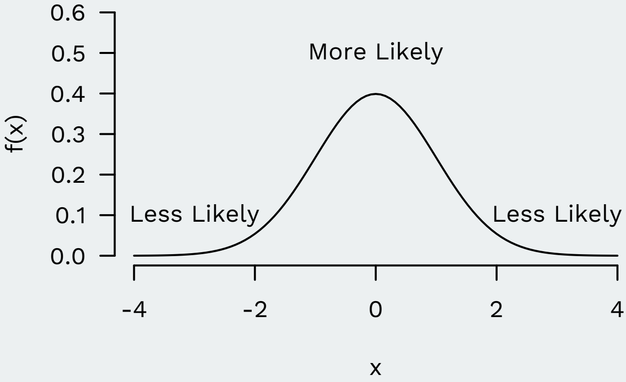
- $\rightsquigarrow$ the probability of a region is the area under the p.d.f. for that region.

Quick calculus note

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx$$

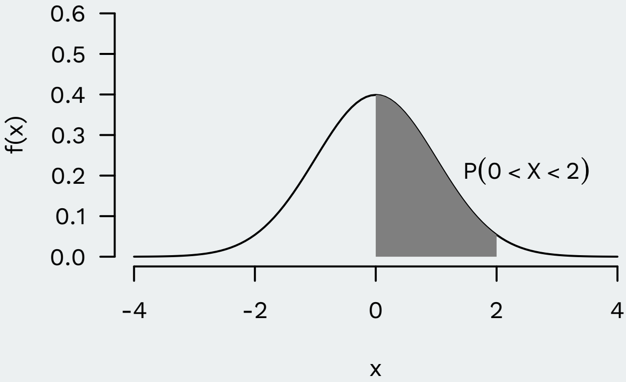
- To get the probability of a particular region of x, we integrate the p.d.f. from a to b.
- Integral: area under the curve.
- Knowing how to integrate: not crucial for this class.
- We'll mostly use R to help us calculate these values.

The p.d.f.



- The p.d.f. gives us information about how likely various outcomes are.
- Regions with higher values of the p.d.f. are areas where we are more likely to see a realization of X.

But be careful!

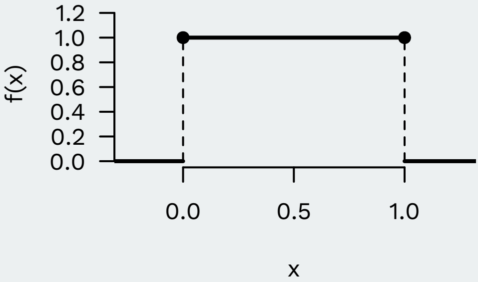


- The height of the curve is not the probability of x:

$$f_X(x) \neq \mathbb{P}(X = x)$$

- We can use the integral to get the probability of falling in a particular region.

Continuous uniform distribution

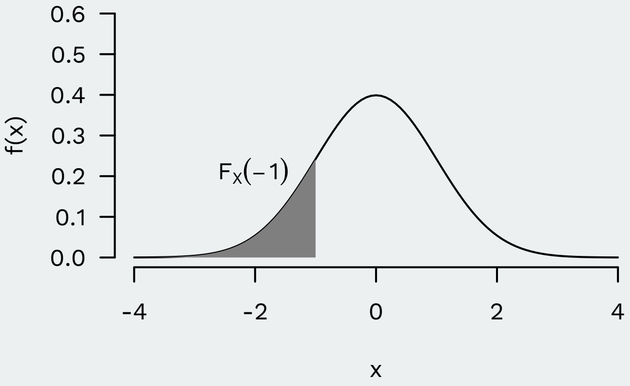


- Continuous uniform distribution on the (a, b) interval.
- We write $X \sim \text{Unif}(a, b)$ and it has the p.d.f.:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

- Every equal-sized region has the same probability of containing X.

p.d.f. -> c.d.f.



- We can write the c.d.f. of a continuous r.v. as:

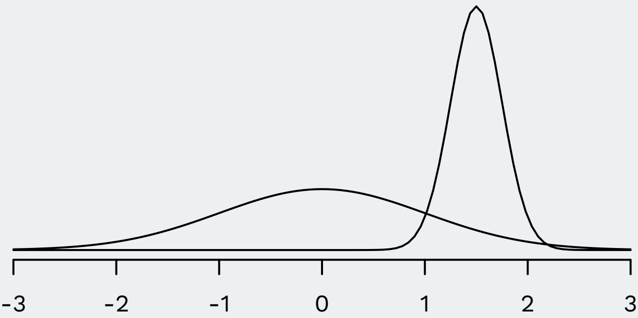
$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

- c.d.f. for continuous r.v. = integral of p.d.f. up to a certain value.

3/ Properties of Distributions

How can we summarize distributions?

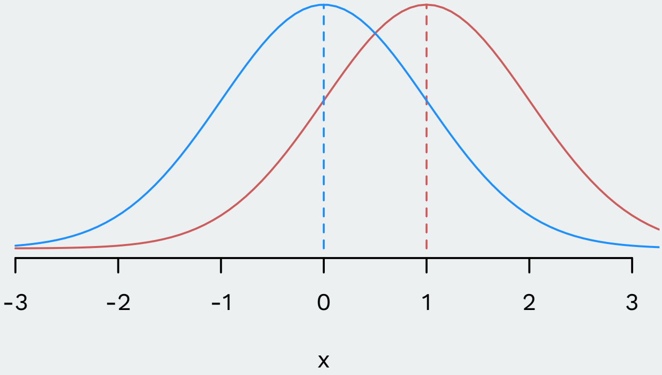
- Probability distributions summarize our uncertainty about r.v.s.
- Can we summarize probability distributions?
- **Question**: What is the difference between these two density curves? How might we summarize this difference?



Goals for summarizing

- One-number summaries of the following:
    1. Central tendency: where the center of the distribution is.
    2. Spread: how spread out the distribution is around the center.
- With real data, we are going to want to estimate these values for a given r.v.

Central tendency



- Where is the middle of the distribution?
- Sometimes called the location of the distribution.

Expectation

- Natural measure of central tendency is the expected value of X.
  - Also called the expectation or mean of the distribution.
  - Written as $\mathbb{E}[X]$ or μ_X.
- For discrete X with k levels, $j = 1, \ldots, k$:

$$\mathbb{E}[X] = \sum_{j=1}^{k} x_j f(x_j)$$

- For continuous X, we have to use the integral:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

- Expected values depend on the distributions of the r.v. through the p.m.f./p.d.f.

Expectation

$$\mathbb{E}[X] = \sum_{j=1}^{k} x_j f(x_j)$$

- Weighted average of the values of the r.v. weighted by the probability of each value occurring.
- Best guess as to where X will be.
- Example: Bernoulli X with $\mathbb{P}(X = 1) = p$:

$$\mathbb{E}[X] = 0 \times (1 - p) + 1 \times p = p$$

- Sometimes, we'll calculate $\mathbb{E}[X]$ directly.
- For named distributions, there is often a formula that is a function of the parameters.

Example – number of treated units

- Randomized experiment with 3 units. X is number of treated units.

x	$f_X(x)$	$xf_X(x)$
0	1/8	0
1	3/8	3/8
2	3/8	6/8
3	1/8	3/8

- Calculate the expectation of X:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{j=1}^{k} x_j f(x_j) \\ &= 0 \times f_X(0) + 1 \times f_X(1) + 2 \times f_X(2) + 3 \times f_X(3) \\ &= 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{12}{8} = 1.5\end{aligned}$$

Properties of the expected value

- Can we figure out the expectation of transformations of X?
  1. **Additivity:** (expectation of sums are sums of expectations)
  $$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$
  2. **Homoegeneity:** Suppose that a and c are constants. Then,
  $$\mathbb{E}[aX + c] = a\mathbb{E}[X] + c$$
  3. **Law of the Unconscious Statistician**, or LOTUS, if $g(X)$ is a function of a discrete random variable, then
  $$\mathbb{E}[g(X)] = \sum_x g(x) f_X(x),$$
- But, in general, the following are also true:
  - $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$ unless $g(\cdot)$ is a linear function.
  - $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$ unless X and Y are independent (next week).

List experiments for sensitive questions

- A list experiment attempts to get around the problem of social desirability bias.
- What proportion of being would be upset by a black family moving in next door to them?
- Randomly split the survey into two halves. Ask both how many of the following upset you:

Control Group:

- the federal government increasing the tax on gasoline;
- professional athletes getting million-dollar salaries;
- large corporations polluting the environment.

Treated Group:

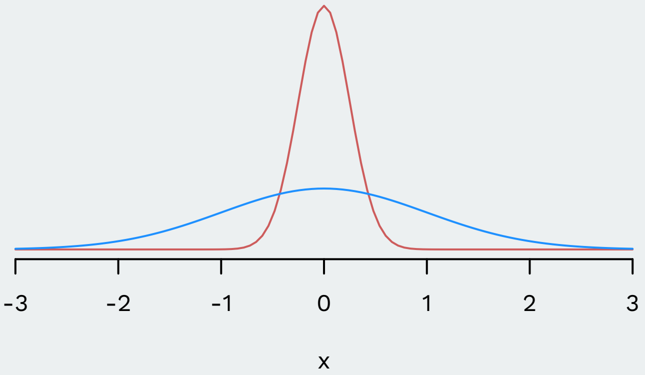
- the federal government increasing the tax on gasoline;
- professional athletes getting million-dollar salaries;
- large corporations polluting the environment.
- a black family moving in next door.

Example: list experiments

- Can define 3 r.v.s:
  - X = number of items under the control group (3 items possible)
  - Y = number of items under the treated group (4 items possible)
  - $A = 1$ if the black neighbors question upset them.
- We want to learn about $\mathbb{E}[A] = \mathbb{P}(A = 1)$, but we don't observe that, only X or Y.
- But note, if randomly assigned, then $Y = X + A$.
  - Y includes all of the control items plus an addition item if the person would be upset by a black family moving in.
- Then, we can write this as:

$$\mathbb{E}[Y] = \mathbb{E}[X] + \mathbb{E}[A]$$
$$\rightsquigarrow \mathbb{E}[A] = \mathbb{E}[Y] - \mathbb{E}[X]$$

Measures of Spread



- How spread out the distribution is around the middle?
- Also called the dispersion of the distribution.
- We'll talk about two of these that are closely related: the variance and the standard deviation.

Variance

- The variance measures the spread of the distribution:

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

- Use LOTUS to calculate the variance for a discrete r.v.:

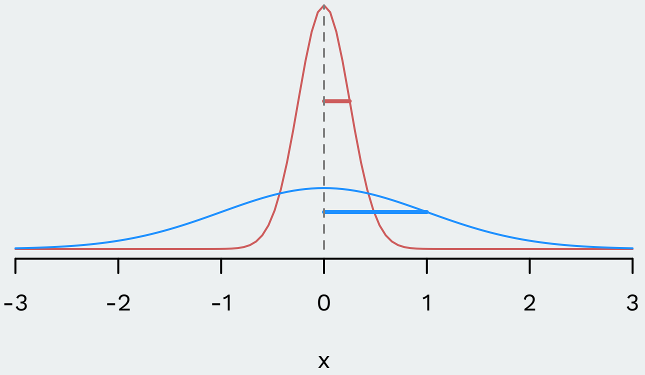
$$\mathbb{V}[X] = \sum_x (x - \mathbb{E}[X])^2 f_X(x)$$

- Same principle for continuous random variables:

$$\mathbb{V}[X] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x) dx$$

- Again, the variance depends on the p.m.f./p.d.f.
- The standard deviation is the (positive) square root of the variance: $\sigma_X = \sqrt{\mathbb{V}[X]}$.

Variance intuition



$$\mathbb{V}[X] = \sum_x (x - \mathbb{E}[X])^2 f_X(x)$$

- Weighted average of the squared distances from the mean.

Example - number of treated units

x	$f_X(x)$	$x - \mathbb{E}[X]$	$(x - \mathbb{E}[X])^2$
0	1/8	-1.5	2.25
1	3/8	-0.5	0.25
2	3/8	0.5	0.25
3	1/8	1.5	2.25

- Let's go back to the number of treated units to figure out the variance of the number of treated units:

$$
\begin{aligned}
\mathbb{V}[X] &= \sum_{j=1}^{k} (x_j - \mathbb{E}[X])^2 f_X(x_j) \\
&= (-1.5)^2 \times \frac{1}{8} + (-0.5)^2 \times \frac{3}{8} + 0.5^2 \times \frac{3}{8} + 1.5^2 \times 18 \\
&= 2.25 \times \frac{1}{8} + 0.25 \times \frac{3}{8} + 0.25 \times \frac{3}{8} + 2.25 \times \frac{1}{8} = 0.75
\end{aligned}
$$

Properties of variances

1. If b is a constant, then

$$\mathbb{V}[b] = 0.$$

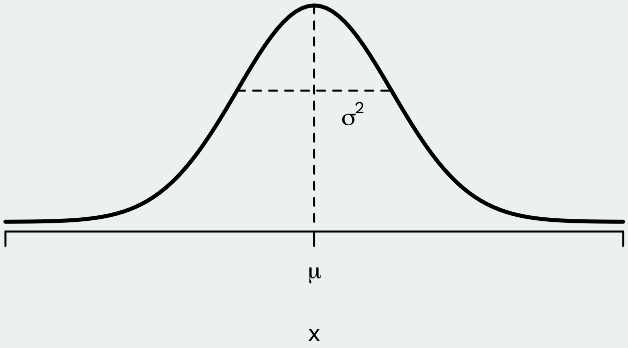
2. If a and b are constants,

$$\mathbb{V}[aX + b] = a^2\mathbb{V}[X].$$

3. In general, $\mathbb{V}[X + Y] \neq \mathbb{V}[X] + \mathbb{V}[Y]$ unless X and Y are independent (next week).

4/ Normal distribution

The normal distribution



- The normal distribution is the classic "bell-shaped" curve.
  - It is extremely useful and ubiquitous in statistics.
- If X has a normal distribution, we write $X \sim N(\mu, \sigma^2)$:
  - μ and σ^2 are the parameters of the normal.
  - μ is the expected value and σ^2 is the variance.

Normal distribution

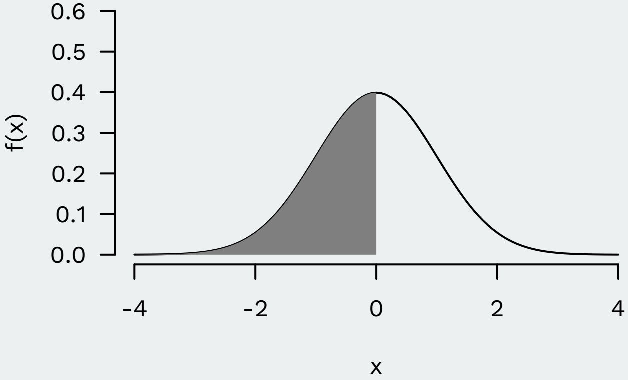
- The p.d.f. for the normal distribution is:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}.$$

- A special member of this family is the standard normal distribution with $N(0,1)$.

Using pnorm

- `pnorm()` evaluates the c.d.f. of the normal:

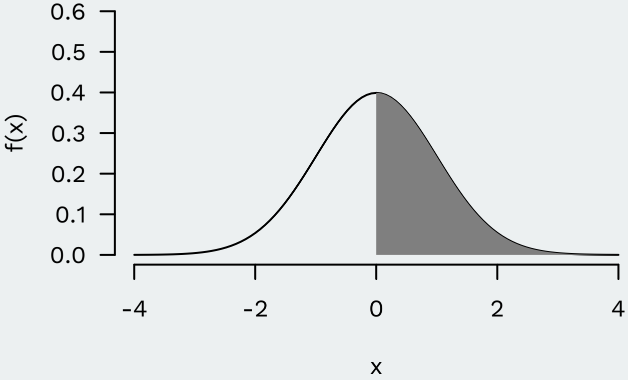



```
pnorm(q = 0, mean = 0, sd = 1)
```

```
## [1] 0.5
```

Using pnorm

- pnorm() evaluates the c.d.f. of the normal:

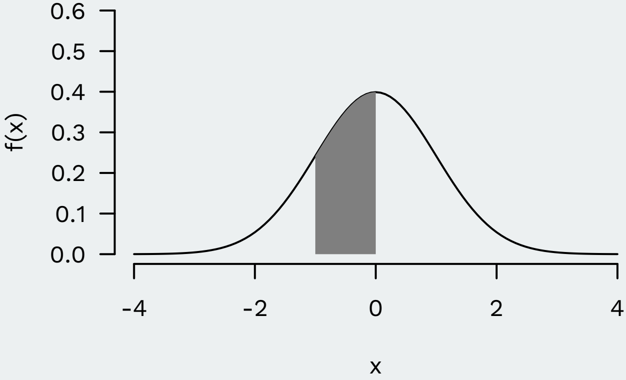



```r
pnorm(q = 0, mean = 0, sd = 1, lower.tail = FALSE)
```

```
## [1] 0.5
```

Using pnorm

- pnorm() evaluates the c.d.f. of the normal:




```
pnorm(q = 0, mean = 0, sd = 1) - pnorm(q = -1, mean = 0,
    sd = 1)
```

```
## [1] 0.3413447
```

5/ Wrap-up

Take-home points

1. Random variables are theoretical constructs that represent our data.
2. Random variables have distributions that summarize the uncertainty in their outcomes.
3. We can summarize these distribution using expectations and variances.

A peek ahead

- Next week: thinking about the distribution of more than r.v.
- How do we evaluate $\mathbb{P}(X = x, Y = y)$?
- Going to define a hugely important concept: conditional expectation.