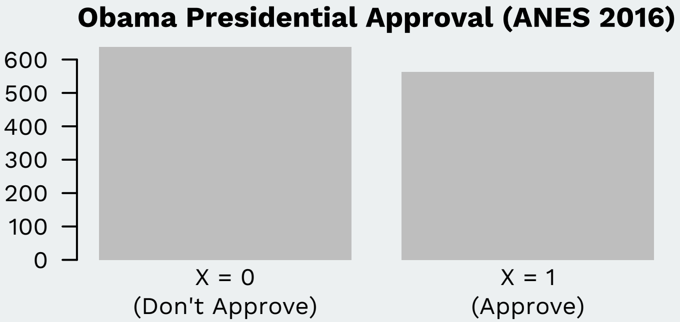
Gov 2000: 2. Random Variables and Probability Distributions

Matthew Blackwell

Fall 2016

Where are we going?



- Long-term goal: inferring the data generating process of this variable.
  - What is the true Obama approval rate in the US?
- Today: given a probability distribution, what data is likely?
  - If we knew the true Obama approval, what samples are likely?

1/ Random Variables

Brief probability review

- Ω is the sample space (set of events that could occur)
- ω is a particular member of the sample space
- Formalize uncertainty over which outcome will occur with probability:
  - $\rightsquigarrow \mathbb{P}(\omega)$ is the probability that a particular outcome will happen.
  - We don't know which outcome will occur, but we know which ones are more likely than others.
- Example: tossing a fair coin twice
  - $\Omega = \{HH, HT, TH, TT\}$
  - Fair coins, independent,

$$\mathbb{P}(HH) = \mathbb{P}(H)\mathbb{P}(H) = 0.5 \times 0.5 = 0.25$$

What are random variables?

Random Variable

A random variable (r.v.) is a function that maps from the sample space of an experiment to the real line or $X : \Omega \to \mathbb{R}$.

- r.v.s are numeric representation of uncertain events $\rightsquigarrow$ we can use math!
- Lower-case letters x are arbitrary values of the r.v.
- Often ω is implicit and we just write the r.v. as X instead of $X(\omega)$.

Examples

- 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
  - $X(HTHTT) = 2$
- Obama approval for a respondent:
  - $\Omega = \{\text{approve}, \text{don't approve}\}$.
  - Random variable converts this into a number:

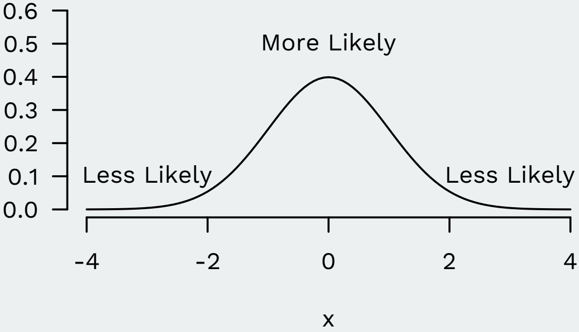
$$X = \begin{cases} 1 \text{ if approve} \\ 0 \text{ if don't approve} \end{cases}$$

  - Called a Bernoulli, binary, or dummy random variable.
- Length of government in a parliamentary system:
  - $\Omega = [0, \infty) \rightsquigarrow$ already numeric so $X(\omega) = \omega$.

2/ Probability Distributions

Randomness and probability distributions

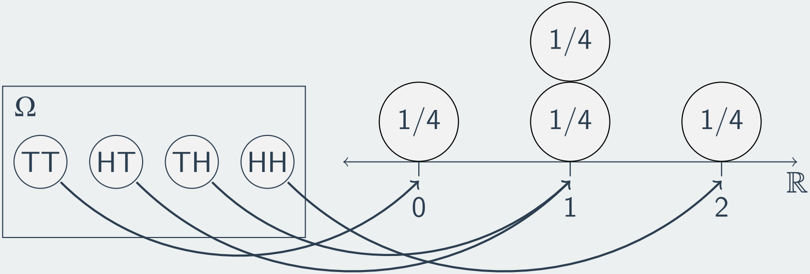
- How are r.v.s random?
  - Uncertainty over $\Omega \rightsquigarrow$ uncertainty over value of X.
  - We'll use probability to formalize this uncertainty.
- The probability distribution of a r.v. gives the probability of all of the possible values of the r.v.



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, $\mathbb{P}(X = 1) = 0.5$
- Data generating process (DGP): assumptions about how the data came to be.
  - Examples: coin flips, randomly selecting a card from a deck, etc.
  - These assumptions imply probabilities over outcomes.
  - 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.

Inducing probabilities



- Let X be the number of heads in two coin flips.

ω	$\mathbb{P}(\{\omega\})$	$X(\omega)$
TT	1/4	0
HT	1/4	1
TH	1/4	1
HH	1/4	2

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

Probability mass function

Discrete Random Variable

A r.v., X, is discrete if its range (the set of values it can take) is finite ($X \in \{x_1, \ldots, x_k\}$) or countably infinite ($X \in \{x_1, x_2, \ldots\}$).

- Probability mass function (p.m.f.) describes the distribution of X when it is discrete:

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

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

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

- Probability of a set of values $S \subset \{x_1, \ldots, x_k\}$:

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

- Examples: Obama approval, number of battle deaths in a conflict, number of parties elected to a legislature.

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.

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

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

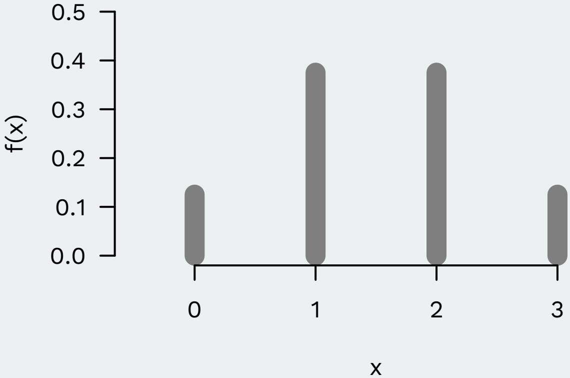
$$f_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(T, T, C) + \mathbb{P}(C, T, T) + \mathbb{P}(T, C, T) = \frac{3}{8}$$

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

- 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?

Real-valued r.v.s

- What if X can take any value on $\mathbb{R}$ or an uncountably infinite subset of the real line?
- Can we just specify $\mathbb{P}(X = x)$?
- No! Proof by counterexample:
    - 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

Probability density functions

Continuous Random Variable

A r.v., X, is continuous if there exists a nonnegative function on $\mathbb{R}$, f_X called the probability density function (p.d.f.) such that for any interval, B:

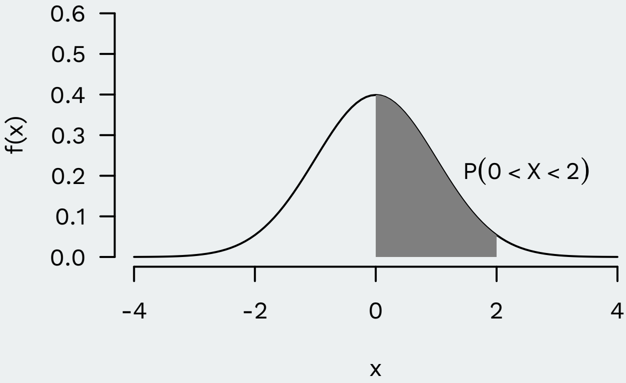
$$\mathbb{P}(X \in B) = \int_B f_X(x)dx$$

- Specifically, for a subset of the real line (a, 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.
- Probability of a point mass: $\mathbb{P}(X = c) = \int_c^c f_X(x)dx = 0$
- Examples: length of time between two governments in a parliamentary system, proportion of voters who turned out, governmental budgets allocations

The p.d.f.



- 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.

3/ Cumulative Distribution Functions

Cumulative distribution functions

- Useful to have a definition of the probability distribution that doesn't depend on discrete vs. continuous:

Cumulative distribution function

The cumulative distribution function (c.d.f.) returns the probability is that a variable is less than a particular value:

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

- Identifies the probability of any interval (including singletons like $X = x$) on the real line.
- For discrete r.v.: $F_X(x) = \sum_{x_j \leq x} f_X(x_j)$
- For continuous r.v.: $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$

Properties of the c.d.f.

1. F_X never decreases: if $x \leq x'$ then $F_X(x) \leq F_X(x')$
   - Proof: the event $X < x$ includes the event $X < x'$ so $\mathbb{P}(X < x')$ can't be smaller than $\mathbb{P}(X < x)$.
2. $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.
3. $F_X(x)$ is right continuous (no jumps when we approach a point from the right)
   - For discrete X, $F_X(x)$ is piecewise constant and staircase-like.
   - For continuous X, $F_X(x)$ is continuous.

Example of discrete c.d.f

- Remember example where X is the number of treated units:

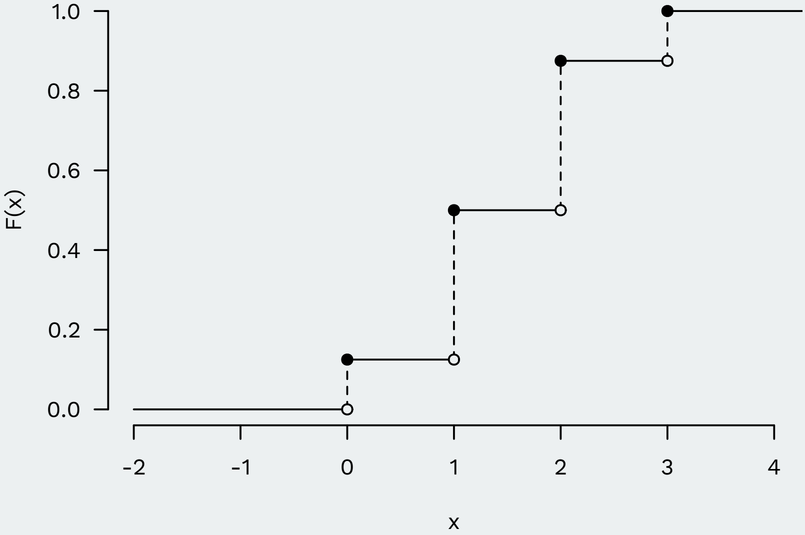
x	$\mathbb{P}(X = x)$
0	1/8
1	3/8
2	3/8
3	1/8

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

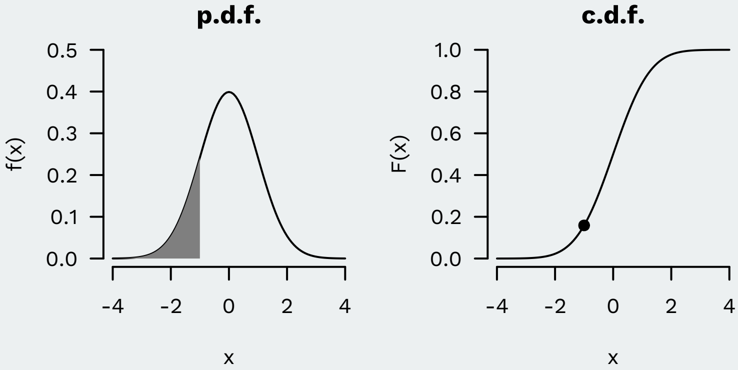
$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \leq x < 1 \\ 1/2 & 1 \leq x < 2 \\ 7/8 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

- What is $F_X(1.4)$ here? 0.5

Graph of discrete c.d.f.



Continuous 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.

Recovering probabilities

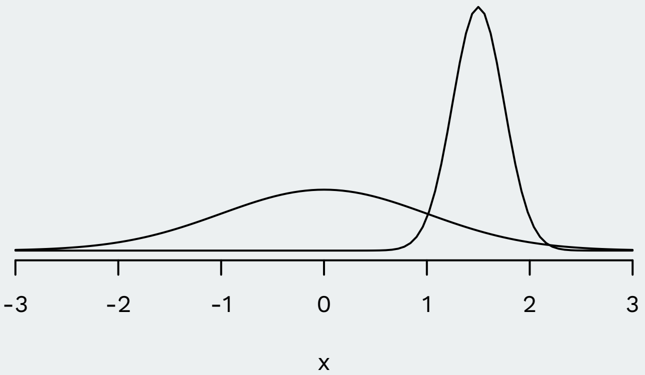
- Let $F_X(x^-) = \lim_{y \uparrow x} F_X(y)$
  - Value of the c.d.f. just below x
  - For continuous r.v., $F_X(x^-) = F_X(x)$
- We can use the c.d.f. to calculate the probability of any interval or value:
  1. $\mathbb{P}(X \leq x) = F_X(x)$
  2. $\mathbb{P}(X > x) = 1 - F_X(x)$
  3. $\mathbb{P}(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$
  4. $\mathbb{P}(X < x) = F_X(x^-)$
  5. $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$
- Example: probability of at least 1 treated and 1 control.
  - $X = 1$ or $X = 2$, so we need the prob. of $0 < X \leq 2$:

$$\mathbb{P}(0 < X \leq 2) = F_X(2) - F_X(0) = 7/8 - 1/8 = 0.75$$

4/ Properties of Distributions

How can we summarize distributions?

- Probability distributions describe the 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

1. Central tendency: where the center of the distribution is.
    - We'll focus on the mean/expectation.
2. Spread: how spread out the distribution is around the center.
    - We'll focus on the variance/standard deviation.

- With real data, we are going to try and infer these values from data on a r.v.

Expectation

- Natural measure of central tendency is the expected value (a/k/a the expectation or mean) of X.
- For discrete $X \in \{x_1, x_2, \dots, x_k\}$ with k levels:

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

  - Weighted average of the values of the r.v. weighted by the probability of each value occurring.

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

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

- Intuition: center of gravity/balance point of the p.m.f./p.d.f.

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:

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

Expectation from a p.d.f.

- Suppose that the p.d.f of a continuous r.v.s is

$$f_X(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

- What is the mean of this variable?

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_0^1 x(2x) dx \\ &= \int_0^1 2x^2 dx \\ &= (2/3)x^3 \Big|_0^1 \\ &= (2/3) \cdot 1^3 - (2/3) \cdot 0^3 = (2/3) \end{aligned}$$

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).

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.:

$$\mathbb{V}[X] = \sum_{j=1}^{k} (x_j - \mathbb{E}[X])^2 f_X(x_j)$$

- Same principle for continuous random variables:

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

- Weighted average of the squared distances from the mean.
  - Larger deviations (+ or −) $\rightsquigarrow$ higher variance
- The standard deviation is the (positive) square root of the variance: $\sigma_X = \sqrt{\mathbb{V}[X]}$.

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:

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

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. $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
4. In general, $\mathbb{V}[X + Y] \neq \mathbb{V}[X] + \mathbb{V}[Y]$ unless X and Y are independent (next week).

Variance from a p.d.f.

- Suppose that the p.d.f of a continuous r.v.s is

$$f_X(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

- We know that $\mathbb{E}[X] = (2/3)$, but what about $\mathbb{V}[X]$?
- We'll exploit $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$:

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\ &= \int_0^1 x^2 (2x) dx \\ &= \int_0^1 2x^3 dx \\ &= (2/4) x^4 \Big|_0^1 = (1/2) \end{aligned}$$

- Plugging back in, $\mathbb{V}[X] = (1/2) - (2/3)^2 = 1/18$

5/ Famous distributions

Families of distributions

- There are several important families of distributions:
  - The p.m.f./p.d.f. within the family has the same form, with parameters that might vary across the family.
  - The parameters determine the shape of the distribution
- Statistical modeling in a nutshell:
  1. Assume the data, $X_1, X_2, \ldots$, are independent draws from a common distribution $f_\theta(x)$ within a family of distributions (normal, poisson, etc)
  2. Use a function of the observed data to estimate the value of the θ: $\hat{\theta}(X_1, X_2, \ldots)$

Bernoulli distribution

- X has a Bernoulli distribution if it is binary and $\mathbb{P}(X = 1) = p$
- Then, for $x \in \{0, 1\}$, the p.m.f. is:

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

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

- Example:
  - $X_1, X_2, \ldots, X_n$ are each a Bernoulli r.v. indicating Obama approval for the ith respondent.
  - p is the Obama approval rate in the population.
  - Sneak peak: how can we learn about p from $X_1, X_2, \ldots, X_n$?

Mean and variance of Bernoulli

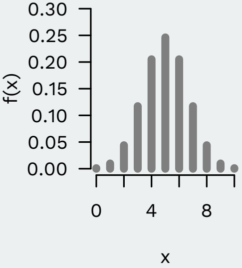
- If X is Bernoulli then the p.m.f. is: $f_X(x) = p^x(1-p)^{1-x}$
- We can calculate $\mathbb{E}[X]$:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{j=1}^{k} x_j f_X(x_j) \\ &= 0 \times f_X(0) + 1 \times f_X(1) \\ &= 0 \times (1-p) + 1 \times p = p\end{aligned}$$

- Note that $X^2 = X$ (why?) so $\mathbb{E}[X^2] = \mathbb{E}[X] = p$
- Variance:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$$

Binomial distribution



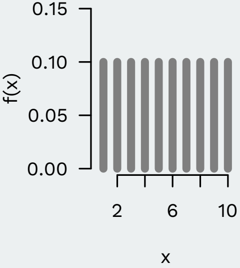
- Let X be the number of heads in n independent coin flips with probability p of heads.
- Then X has a binomial distribution written $X \sim \text{Bin}(n, p)$ which has p.m.f.:

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

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

- Equivalent to the sum of n Bernoulli r.v.s each with probability p.
- $\rightsquigarrow \mathbb{E}[X] = np$ and $\mathbb{V}[X] = np(1-p)$
- Example: number of treated units in the RCT example.

Discrete uniform distribution

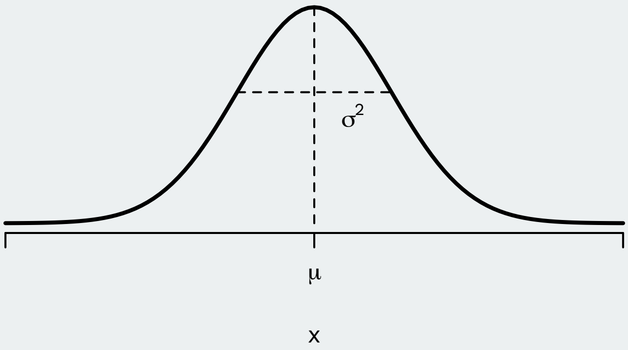


- Equal probability of any value of X:

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

- Justified from the DGP of random sampling.

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 $X \sim N(\mu, \sigma^2)$:
  - $\mathbb{E}[X] = \mu$ and $\mathbb{V}[X] = \sigma^2$ are the parameters of the normal.

Normal distribution

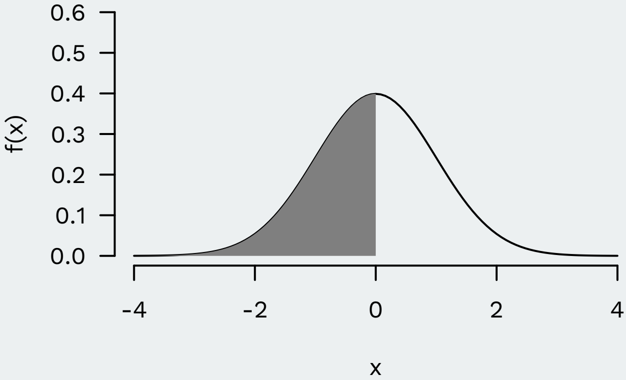
- The p.d.f. for the normal distribution is:

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

- 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:

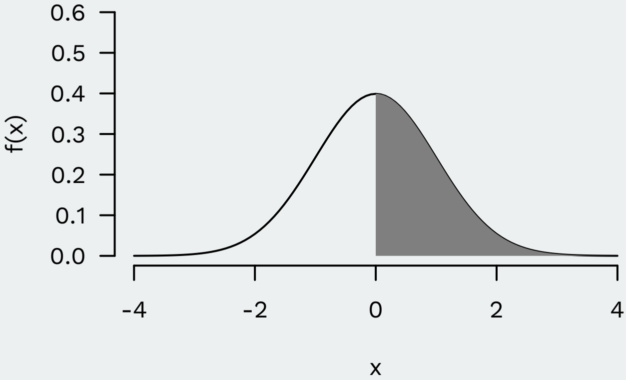



```r
pnorm(q = 0, mean = 0, sd = 1)
```

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

Using pnorm

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

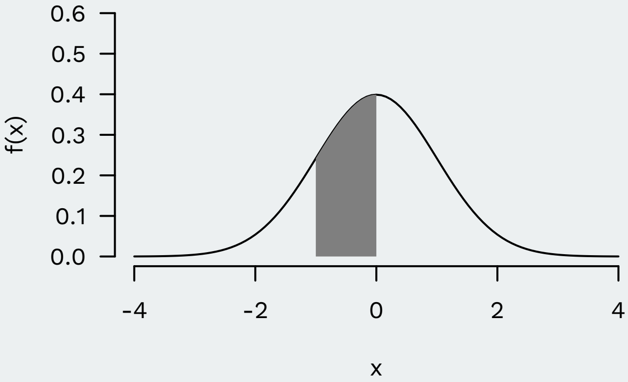



```r
pnorm(q = 0, mean = 0, sd = 1, lower.tail = FALSE)
```

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

Using pnorm

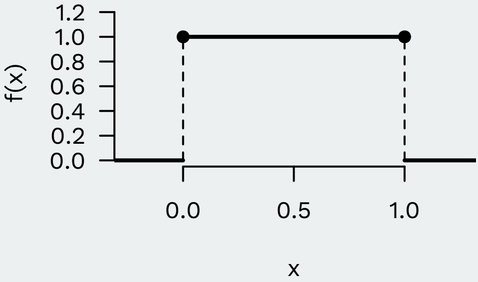
- pnorm() evaluates the c.d.f. of the normal:




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

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

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.

6/ Simulating Random Variables*

Strategies for calculating means/variances

- Do you know the p.m.f./p.d.f.?
  - $\rightsquigarrow$ calculate $\mathbb{E}[X]/\mathbb{V}[X]$ directly using the definitions.
  - Often need calculus/summation tricks.
- Is the X a linear function of another variable(s) whose mean/variance you do know?
  - $\rightsquigarrow$ use linearity of expectations.
  - Ex.: $\mathbb{E}[Y] = 0.2$ and $X = Y + 1 \rightsquigarrow \mathbb{E}[X] = \mathbb{E}[Y] + 1 = 1.2$
- Can you simulate it?
  - draw a large number of realizations of X and calculate the mean/variance of those.
  - useful when using p.d.f./p.m.f. is complicated.

Simulating r.v.s in R

- You can draw multiple realizations of a famous r.v. in R using functions like `runif()` or `rnorm()`.
- One draw from the Unif(0,1) distribution:

```
runif(n = 1, min = 0, max = 1)
```

```
## [1] 0.7265663
```

- Mean of 1000 draws from the same distribution:

```
hold <- runif(n = 1000, min = 0, max = 1)
mean(hold)
```

```
## [1] 0.5134936
```

Simulation of probabilities

- You can also simulate the probabilities of various intervals:

$$\mathbb{P}(X \in B) \approx \frac{\#\text{ of draws in } B}{\text{total number of draws}}$$

- What's the probability of Unif(0,1) being more than 0.7?

```
sum(hold > 0.7)/length(hold)
```

```
## [1] 0.305
```

```
mean(hold > 0.7)
```

```
## [1] 0.305
```

7/ 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.