Gov 2000 - 5. Univariate Inference

Matthew Blackwell

October 6, 2015

1/ Evaluation

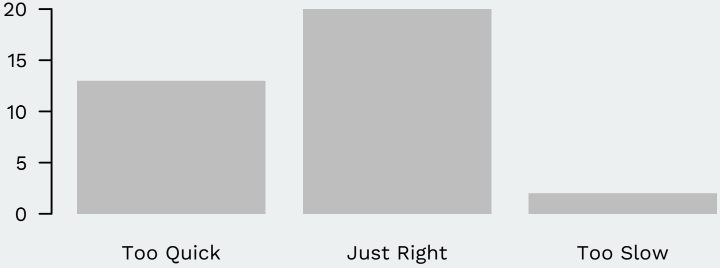
Evaluation overview

- 35 responses
- Good mix between the different course numbers
- Overall, things are going OK, but definite room for improvement

Pace

- How is the pace of the class? Are lectures too quick, too slow, or just right?

```
barplot(table(evals$pace), border = NA)
```




```
mean(evals$pace.numeric)
```

```
## [1] 1.7
```

Lecture Notes

- How have you found the lecture notes in the course so far?

```r
barplot(table(evals$lecture.notes), border = NA)
```


```r
mean(evals$notes.numeric)
```

```
## [1] 4.1
```

How's Matt doing?

- What overall rating would you give this instructor?

```r
barplot(table(evals$instructor), border = NA)
```

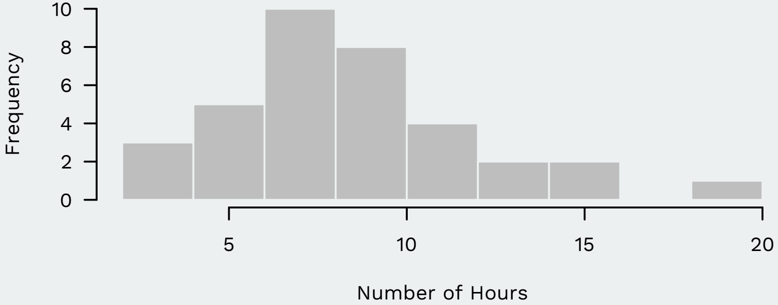

```r
mean(evals$instructor.numeric)
```

```
## [1] 4.4
```

HW Hours

- How difficult do you find the problem sets in the course so far?

```
hist(evals$hw.hours, border = "#ECF0F1", col = "grey",
    main = "", xlab = "Number of Hours")
```




```
mean(evals$hw.hours)
```

```
## [1] 9
```

Course Overall

- What overall rating would you give the course?

```
barplot(table(evals$overall), border = NA)
```


```
mean(evals$overall.numeric)
```

```
## [1] 3.9
```

Best things about the course

- “Mayya and Anton are THE BEST. So helpful, always make time to help”
- “The diversity of materials and supplemental material”
- “The quality of instruction/explanation in both the lecture and the section is superb.”

Ways to improve

- “very heavy workload”
- “The problem sets are taking far, far too long”
- “go through the examples a little faster so we don’t speed through the last 15 mins of slides and concepts”
- “there’s quite a substantial gap in difficulty between the material covered in the lectures and the problem sets.”
- “we got into the weeds a bit and Matt gave fewer examples of how these things (all the weird distributions and PMFs) tie back into actual research”
- “As I am an absolute beginner in R, I find the coding part most challenging.”
- “It would help if we could see more examples in lecture/sections that are similar to what we will see on problem sets”

Our Plan

- More R!
- More connection between the lectures and problem sets
- More examples in lecture.
- No easing off the gas.
- But things should get easier with time.

Where are we? Where are we going?

- Last few weeks: probability, learning how to think about r.v.s
- Now: how to estimate features of underlying distributions with real data.
- Have to think about how much our estimate is due to chance variation.

2/ Populations and Samples

Motivating example

- Gerber, Green, and Larimer (APSR, 2008)

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	______
9995 JENNIFER KAY SMITH		Voted	______
9997 RICHARD B JACKSON		Voted	______
9999 KATHY MARIE JACKSON		Voted	______

Motivating Example

```
load("gerber_green_larimer.RData")
## turn turnout variable into a numeric
social$voted <- 1 * (social$voted == "Yes")
neigh.mean <- mean(social$voted[social$treatment ==
    "Neighbors"])
neigh.mean
```

```
## [1] 0.38
```

```
contr.mean <- mean(social$voted[social$treatment ==
    "Civic Duty"])
contr.mean
```

```
## [1] 0.31
```

```
neigh.mean - contr.mean
```

```
## [1] 0.063
```

- Is this a “real”? Is it big?

Why study estimators?

- **Goal 1: Inference**
  - How is this estimate related to the true/population difference in means?
  - How much uncertainty do we have in this estimate?
- **Goal 2: Compare estimators**
  - In an experiment, use simple difference in sample means $(\bar{Y} - \bar{X})$?
  - Or the **post-stratification estimator**, where we estimate the estimate the difference among two subsets of the data (male and female, for instance) and then take the weighted average of the two ($\bar{Z}$ is the share of women):

$$\hat{\theta}_{ps} = (\bar{Y}_f - \bar{X}_f)\bar{Z} + (\bar{Y}_m - \bar{X}_m)(1 - \bar{Z})$$

  - Which (if either) is better? How would we know?

Samples from the DGP

- $Y_i = 1$ if citizen i votes, $Y_i = 0$ otherwise.
- $Y_1, \ldots, Y_n$ represent the **sample**
- Our goal: learn the **data generating process** that generated the sample.
- DGP: $Y_1, \ldots, Y_n$ are i.i.d. draws from the **population distribution** with p.m.f. or p.d.f. $f_Y(y)$.
    - $Y_i \sim_{iid} \text{Bern}(\mu)$
    - "Y_i are distributed i.i.d. Bernoulli with probability μ":

$$f_Y(y_i) = \mu^{y_i}(1-\mu)^{1-y_i}$$

- μ is an unknown **parameter** of the population distribution.

DGP

- Data generating process (DGP) $\rightsquigarrow$ population distribution of the data
- Where do DGPs come from?
- **First principles**:
  - Waiting time between random events $\rightsquigarrow$ Exponential distribution
  - Number of events in a fixed interval $\rightsquigarrow$ Poisson distribution
- **Random sampling** from a population.
  - Y_i randomly selected from the population with mean μ is an r.v. with expectation μ

Partial probability models

- We don't always need to specify a full probability model, $f_Y(y)$, just the mean/variance.
  - $Y_i, \dots, Y_n$ i.i.d. with mean $\mathbb{E}[Y_i] = \mu$ and variance $\mathbb{V}[Y_i] = \sigma^2$
  - Could be justified from random sampling
- Don't have to pick a distribution from Bernoulli, uniform, normal, etc.
- Why is this useful?
  - We can learn about part of the distribution ($\mathbb{E}[Y_i]$) without having to know about other parts of the distribution.

3/ Point Estimation

Parameters

- Trying to estimate **parameters** of population distributions:
  - $\mu = \mathbb{E}[Y]$: the mean
  - $\sigma^2 = \mathbb{V}[X]$: the variance
  - σ: the standard deviation
  - $\mu_y - \mu_x = \mathbb{E}[Y] - \mathbb{E}[X]$: the difference in means between two groups
  - $\mathbb{E}[Y|X] = \alpha + \beta X$: intercept ($\alpha$) and slope ($\beta$) of the regression line
- We'll generically refer to the parameter we're trying to estimate as θ.
- These are the things we want to learn about.

Estimators

- **Definition**: An **estimator** , $\hat{\theta}$ of some parameter θ, is a function of the sample: $\hat{\theta} = h(Y_1, \dots, Y_n)$.
  - Note: $\hat{\theta}$ is a r.v. because it is a function of r.v.s.
- **Question** Why is the following statement wrong: "My estimate was the sample mean and my estimator was 0.38"?

Examples of Estimators

- For the population mean, μ, we have many different possible estimators:
  - $\hat{\theta} = \bar{Y}$ the sample mean
  - $\hat{\theta} = Y_1$ just use the first observation
  - $\hat{\theta} = \max(Y_1, \dots, Y_n)$
  - $\hat{\theta} = 3$ always guess 3

```r
# mean
mean(social$voted[social$treatment == "Neighbors"])
```

```
## [1] 0.38
```

```r
# first observation
social$voted[social$treatment == "Neighbors"][1]
```

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

```r
# maximum
max(social$voted[social$treatment == "Neighbors"])
```

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

```r
# always choose 3
3
```

```
## [1] 3
```

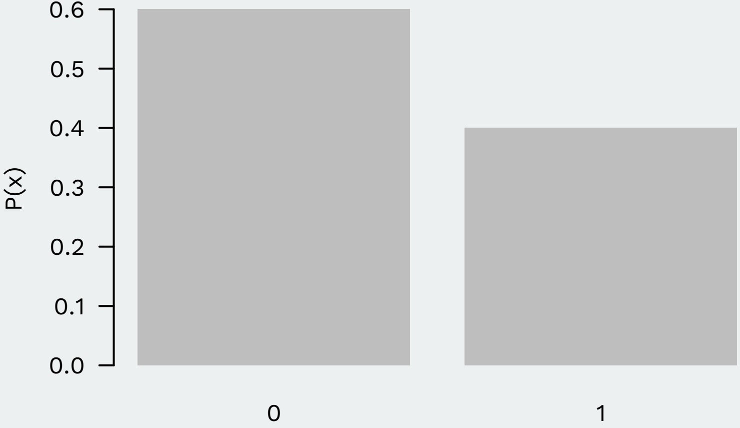
The Three Distributions

- **Population Distribution**: the data-generating process
  - Bernoulli in the case of the social pressure/voter turnout example)
- **Sample distribution**: $Y_1, \ldots, Y_n$
  - series of 1s and 0s in the sample
- **Sampling distribution**: distribution of the estimator over repeated samples from the population distribution
  - the 0.38 sample mean in the "Neighbors" group is one draw from this distribution

- **Question**: If $Y_1, \ldots, Y_n$ is a random sample from a (population) Bernoulli distribution with mean/probability μ, will sampling distribution of the sample mean ($\bar{Y}$) be Bernoulli as well?

Population Distribution

```r
## population distribution is Bernoulli with
## probability of success 0.4
barplot(dbinom(0:1, size = 1, prob = 0.4), names.arg = c(0,
    1), border = NA, ylab = "P(x)", xlab = "x")
```



Sample Distribution

```
my.samp <- rbinom(n = 10, size = 1, prob = 0.4)
table(my.samp)
```

```
## my.samp
## 0 1
## 8 2
```

```
plot(table(my.samp), type = "h", lwd = 5, bty = "n")
```

Sampling Distribution

```
## now we take the mean of the this sample, which is
## one draw from the **sampling distribution**
mean(my.samp)
```

```
## [1] 0.2
```

```
## let's take another draw from the population dist
my.samp.2 <- rbinom(n = 10, size = 1, prob = 0.4)
```

```
## Let's feed this sample to the sample mean
## estimator to get another estimate, which is
## another draw from the sampling distribution
mean(my.samp.2)
```

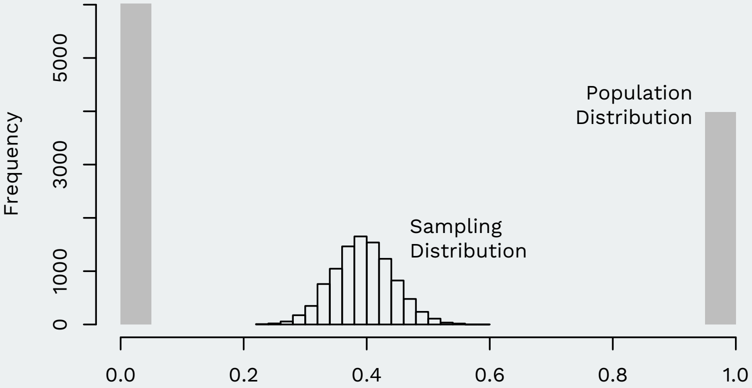
```
## [1] 0.5
```

Sampling distribution by simulation

- Let's generate 10,000 draws from the sampling distribution of the sample mean here when $n = 100$.

```
nsims <- 10000
mean.holder <- rep(NA, times = nsims)
first.holder <- rep(NA, times = nsims)
for (i in 1:nsims) {
    my.samp <- rbinom(n = 100, size = 1, prob = 0.4)
    mean.holder[i] <- mean(my.samp)  ## sample mean
    first.holder[i] <- my.samp[1]  ## first obs
}
```

Sampling distribution versus population distribution



- We only get one draw from the sampling distribution, $\hat{\theta}$.
- Ideally, the sampling distribution would put most of its mass close to the true value of θ.

- **Question** The sampling distribution refers to the distribution of θ, true or false.

Properties of estimators

- We want to learn about true difference in means, but we only have the sample difference in means.
- Is our estimator good? It is better than some other estimators?
- How do we evaluation these little machines that take in samples and output estimates?
- There are two ways we evaluate estimators:
  - **Finite sample**: the properties of its sampling distribution for a fixed sample size n.
  - **Large sample**: the properties of the sampling distribution as we let $n \to \infty$.

Running example

- $Y_1, \dots, Y_n$ are i.i.d. with mean μ_y and variance σ_y^2
- $X_1, \dots, X_n$ are i.i.d. with mean μ_x and variance σ_x^2
- Assume samples are independent.
- Differences in sample means:

$$\widehat{D}_n = \bar{Y}_n - \bar{X}_n$$

- $\widehat{D}_n$ is a random variable!
- What we'll try to figure out: the sampling distribution of $\widehat{D}_n$:

$$\widehat{D}_n \sim ?(\mathbb{E}[\hat{D}_n], \mathbb{V}[\hat{D}_n])$$

4/ Finite-Sample Properties of Estimators

Unbiasedness

- **Definition**: The **bias** of an estimator $\hat{\theta}$ for population parameter θ is
$$\text{bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$$
- **Definition**: An estimator $\hat{\theta}$ of θ is **unbiased** if
$$\mathbb{E}[\hat{\theta}] = \theta$$

Difference in means

- We want to estimate the population difference in means, $\mu_y - \mu_x$
  - True difference between those who received mailers and those who didn't.

- We know from last week and the reading that $\mathbb{E}[\bar{X}_n] = \mu_x$ and $\mathbb{E}[\bar{Y}_n] = \mu_y$.
- What's the expected value of the difference in means?

$$\begin{aligned}\mathbb{E}[\widehat{D}_n] &= \mathbb{E}[\bar{Y}_n - \bar{X}_n] \\ &= \mathbb{E}[\bar{Y}_n] - \mathbb{E}[\bar{X}_n] \\ &= \mu_y - \mu_x\end{aligned}$$

- $\rightsquigarrow$ sample difference in means is **unbiased** for the population difference in means.

- We can also check the bias of an estimator using simulation:

```
nsims <- 10000
mean.holder <- rep(NA, times = nsims)
first.holder <- rep(NA, times = nsims)
for (i in 1:nsims) {
    my.samp.y <- rbinom(n = 100, size = 1, prob = 0.5)
    my.samp.x <- rbinom(n = 100, size = 1, prob = 0.3)
    mean.holder[i] <- mean(my.samp.y) - mean(my.samp.x)
    first.holder[i] <- my.samp.y[1] - my.samp.x[1]
}
```

```
mean(mean.holder) - 0.2
```

```
## [1] 0.0012
```

```
mean(first.holder) - 0.2
```

```
## [1] 0.0027
```

- Both are pretty close to 0!

Sampling variance

- Unbiasedness is about the center/expectation of the sampling distribution of $\hat{\theta}$.
- What about the spread of the sampling distribution?

- **Definition**: The **sampling variance** of an estimator is simply its variance over repeated samples, $\mathbb{V}[\hat{\theta}]$.
- **Definition**: The **standard error** of an estimator is the standard deviation of the sampling distribution, $SE[\hat{\theta}] = \sqrt{\mathbb{V}[\hat{\theta}]}$

Sampling variance of the difference in means

- Can we get a measure of uncertainty in our difference in sample means?
- Variance of the difference in means:

$$\mathbb{V}[\widehat{D}_n] = \mathbb{V}[\bar{Y}_n - \bar{X}_n] = \mathbb{V}[\bar{Y}_n] + \mathbb{V}[\bar{X}_n]$$

- What is the sampling variance of the sample mean?

$$\mathbb{V}[\bar{Y}_n] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[Y_i] = \frac{1}{n^2} n\sigma_y^2 = \frac{\sigma_y^2}{n}$$

- Variance of the sample mean is the variance of each observation divided by the number of observations.
- Difference, redux:

$$\mathbb{V}[\widehat{D}_n] = \frac{\sigma_y^2}{n} + \frac{\sigma_x^2}{n}$$

- We can also investigate the sampling variance by simulation. Take the above code and let's look at the variance of the two sampling distributions:

```r
var(mean.holder)
```

```
## [1] 0.0047
```

```r
var(first.holder)
```

```
## [1] 0.46
```

- Obviously, the sample mean has a much lower variance than just using the first observation. And since they are both unbiased, this means that our estimates will be closer to the truth on average.

Sampling distribution

- Putting things together, we can see that:

$$\widehat{D}_n \sim ?\left(\mu_y - \mu_x, \frac{\sigma_y^2}{n} + \frac{\sigma_x^2}{n}\right)$$

- Sample difference in means is an r.v. with expectation $\mu_y - \mu_x$ and variance $\frac{\sigma_y^2}{n} + \frac{\sigma_x^2}{n}$
- Awesome, we know it gets the right answer on average.
- Two problems:
    1. We don't get to observe σ_y^2 or σ_x^2
    2. We don't know what distribution it follows (yet)

Estimating the Sampling Variance/Standard Error

- To estimate $\mathbb{V}[\bar{Y}_n] = \sigma_y^2/n$ we need and estimate for σ_y^2.
- Use the sample variance of Y_i:

$$S_{yn}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2$$

- Thus, our estimate of the sampling variance of the sample mean is $\widehat{\mathbb{V}}[\bar{Y}_n] = S_{yn}^2/n$ and the standard error of the sample mean is the $\hat{\sigma}_y = S_{yn}/\sqrt{n}$.
- These are estimates of how uncertain our estimates are.
- (S_{yn}^2 is unbiased for σ_y^2)

Estimated sampling distribution

- Our best guess about the sampling distribution of $\widehat{D}_n$:

$$\widehat{D}_n \sim {\color{red}?}\left(\mu_y - \mu_x, \frac{S_{yn}^2}{n} + \frac{S_{xn}^2}{n}\right)$$

- What is the distribution???????

5/ Large-Sample Properties of Estimators

Sequences of estimators

- So far: fixed sample size n.
- But how does the estimator perform as we give it more data (increase n)?
- Need to think about sequences of estimators with increasing n:

$$
\begin{aligned}
\bar{Y}_1 &= Y_1 \\
\bar{Y}_2 &= (1/2) \cdot (Y_1 + Y_2) \\
\bar{Y}_3 &= (1/3) \cdot (Y_1 + Y_2 + Y_3) \\
\bar{Y}_4 &= (1/4) \cdot (Y_1 + Y_2 + Y_3 + Y_4) \\
\bar{Y}_5 &= (1/5) \cdot (Y_1 + Y_2 + Y_3 + Y_4 + Y_5) \\
&\vdots \\
\bar{Y}_n &= (1/n) \cdot (Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + \cdots + Y_n)
\end{aligned}
$$

- Note: this is a sequence of random variables!

Question

- **Question** From what we know, how does the distribution of the sample mean change as n increases?

Convergence in Probability

- **Definition**: A sequence of random variables, $X_1, X_2, \ldots$, is said to **converge in probability** to a value c if for every $\varepsilon > 0$,

$$\mathbb{P}(|X_n - c| > \varepsilon) \to 0,$$

  as $n \to \infty$. We write this $X_n \xrightarrow{p} c$.

- Wooldridge writes $\text{plim}(X_n) = c$ if $X_n \xrightarrow{p} c$.

Consistency

- **Definition** An estimator $\hat{\theta}_n$ is **consistent** for θ if $\hat{\theta}_n \xrightarrow{p} \theta$.
- An unbiased estimator is consistent if the sampling variance goes to 0 as $n \to \infty$ or $\lim_{n\to\infty} \mathbb{V}[\hat{\theta}_n] = 0$.

- **Theorem** (Weak Law of Large Numbers) Let $Y_1, \dots, Y_n$ be a an iid draws from a distribution with mean μ and let $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Then, $\bar{Y}_n \xrightarrow{p} \mu$.

Consistency vs. unbiasedness

- The "first observation" estimator, $\hat{\theta}_n^f = Y_1$ is unbiased, $\mathbb{E}[Y_1] = \mu$
- But it is inconsistent.
- The sampling distribution never collapses to any value. As we add more observations, only use the first:

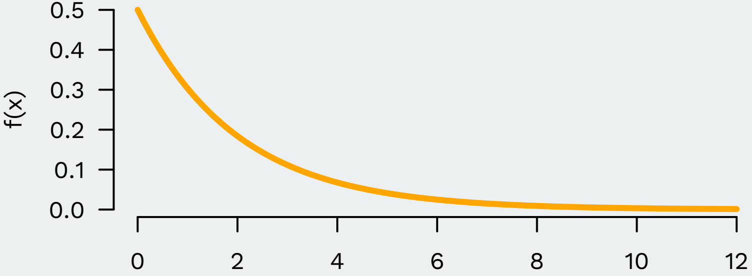
$$\begin{aligned} \hat{\theta}_1^f &= Y_1 \\ \hat{\theta}_2^f &= Y_1 \\ \hat{\theta}_3^f &= Y_1 \\ &\vdots \\ \hat{\theta}_n^f &= Y_1 \end{aligned}$$

- Distribution of $\hat{\theta}_n^f$ never changes so $\mathbb{P}(|\hat{\theta}_n^f - \mu| > \varepsilon)$ never changes $\rightsquigarrow$ never converges.

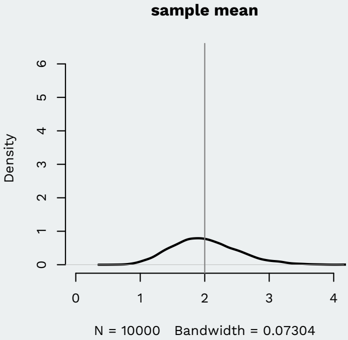
Consistency example

- Suppose that we are interested in how long a government lasts in parliamentary democracies.
- Now, suppose that the distribution of X is Exponential with rate 0.5.

```r
curve(dexp(x, rate = 0.5), from = 0, to = 12, col = "orange",
    lwd = 3, bty = "n", ylab = "f(x)", xlab = "x")
```

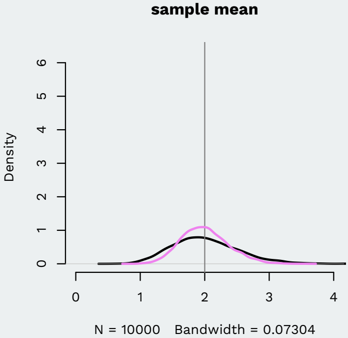


- Now imagine two data collection schemes:
  1. wait until we've collected n observations from the process (this is just an iid sample)
  2. stop collecting after 3 years, which we call a **censored sample**.
- What are the properties of these two approaches?



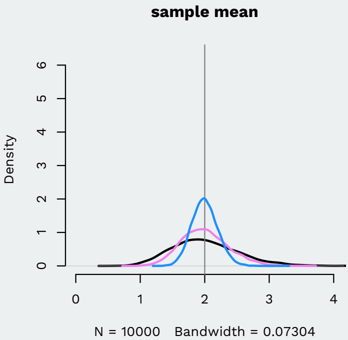
- Now imagine two data collection schemes:
  1. wait until we've collected n observations from the process (this is just an iid sample)
  2. stop collecting after 3 years, which we call a **censored sample**.

- What are the properties of these two approaches?



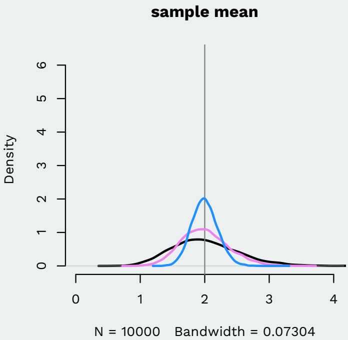
- Now imagine two data collection schemes:
    1. wait until we've collected n observations from the process (this is just an iid sample)
    2. stop collecting after 3 years, which we call a **censored sample**.

- What are the properties of these two approaches?



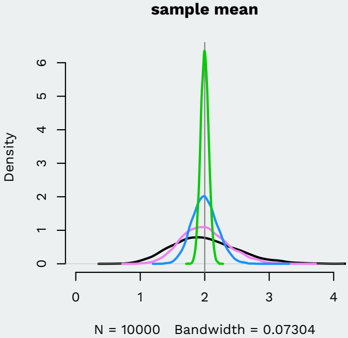
- Now imagine two data collection schemes:
  1. wait until we've collected n observations from the process (this is just an iid sample)
  2. stop collecting after 3 years, which we call a **censored sample**.

- What are the properties of these two approaches?



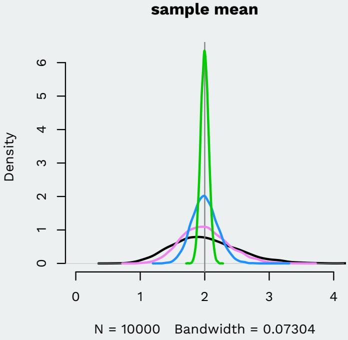
- Now imagine two data collection schemes:
    1. wait until we've collected n observations from the process (this is just an iid sample)
    2. stop collecting after 3 years, which we call a **censored sample**.

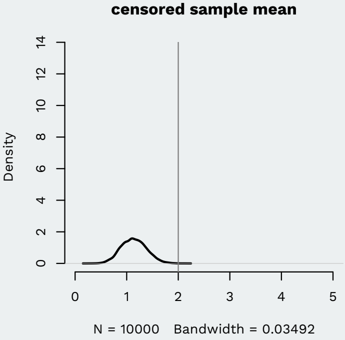
- What are the properties of these two approaches?



- Now imagine two data collection schemes:
  1. wait until we've collected n observations from the process (this is just an iid sample)
  2. stop collecting after 3 years, which we call a **censored sample**.

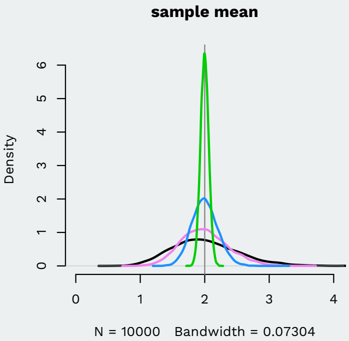
- What are the properties of these two approaches?

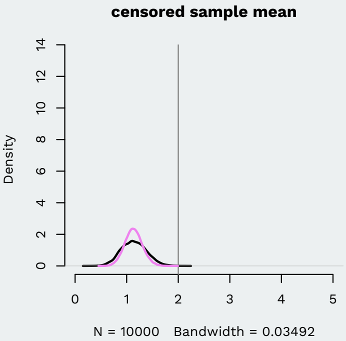




- Now imagine two data collection schemes:
  1. wait until we've collected n observations from the process (this is just an iid sample)
  2. stop collecting after 3 years, which we call a **censored sample**.

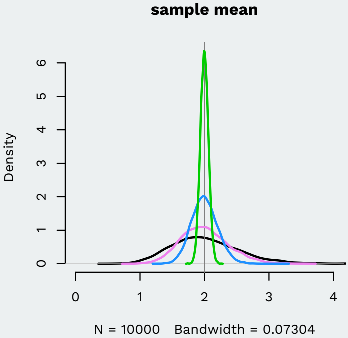
- What are the properties of these two approaches?

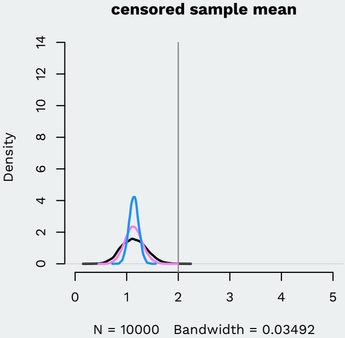




- Now imagine two data collection schemes:
    1. wait until we've collected n observations from the process (this is just an iid sample)
    2. stop collecting after 3 years, which we call a **censored sample**.

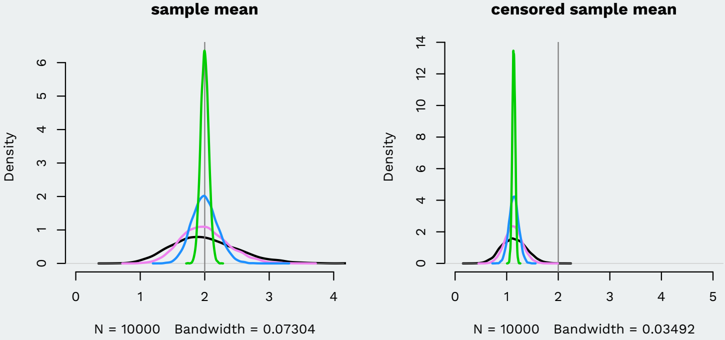
- What are the properties of these two approaches?





- Now imagine two data collection schemes:
    1. wait until we've collected n observations from the process (this is just an iid sample)
    2. stop collecting after 3 years, which we call a **censored sample**.

- What are the properties of these two approaches?



Convergence in Distribution

- **Definition**: A sequence of random variables, $X_1, X_2, \ldots$, is said to **converge in distribution** to Z if

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(Z \leq x),$$

which we write as $X_n \xrightarrow{d} Z$.

Asymptotic Normality

- **Definition**: An estimator is said to be **asymptotically normal** if

$$\frac{\hat{\theta} - \theta}{\sqrt{\mathbb{V}(\hat{\theta})}} \xrightarrow{d} N(0, 1).$$

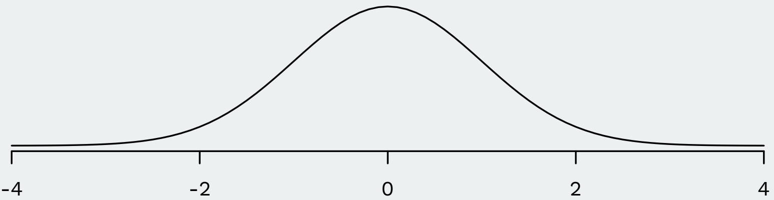
- Facts about normal r.v.s: $\sim N(\mu, \sigma^2)$, then:
  - $(X - \mu) \sim N(0, \sigma^2)$
  - $(X - \mu)/\sigma \sim N(0, 1)$

Central Limit Theorem

- **Theorem** (Central Limit Theorem) Let $Y_1, \ldots, Y_n$ be a an iid draws from a distribution with mean μ and variance σ^2. Then

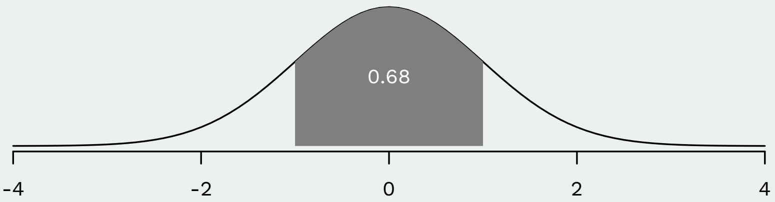
$$Z_n = \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$

Empirical Rule for the Normal Distribution



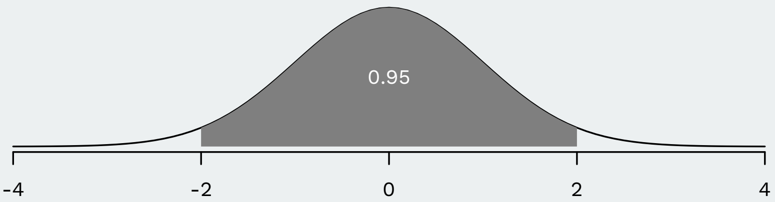
- If $Z \sim N(0,1)$, then the following are roughly true:

Empirical Rule for the Normal Distribution



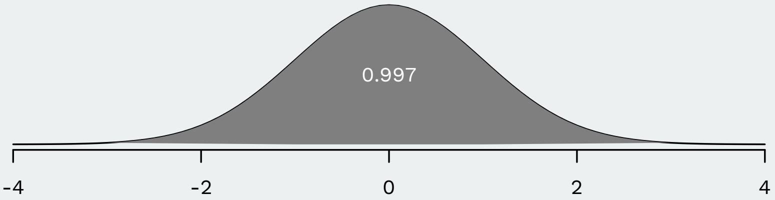
- If $Z \sim N(0,1)$, then the following are roughly true:
- Roughly 68% of the distribution of Z is between -1 and 1.

Empirical Rule for the Normal Distribution



- If $Z \sim N(0, 1)$, then the following are roughly true:
- Roughly 68% of the distribution of Z is between -1 and 1.
- Roughly 95% of the distribution of Z is between -2 and 2.

Empirical Rule for the Normal Distribution



- If $Z \sim N(0, 1)$, then the following are roughly true:
- Roughly 68% of the distribution of Z is between -1 and 1.
- Roughly 95% of the distribution of Z is between -2 and 2.
- Roughly 99.7% of the distribution of Z is between -3 and 3.

- You can use the `pnorm()` function in R to figure out any probability questions about the Normal distribution.

Final sampling distribution

- Putting everything together, we know that as n **gets large**, the distribution of $\widehat{D}_n$ will be approximately:

$$\widehat{D}_n \sim N\left(\mu_y - \mu_x, \frac{\sigma_y^2}{n} + \frac{\sigma_x^2}{n}\right)$$

- Using the properties of normals, we also know that:

$$\frac{\widehat{D}_n - (\mu_y - \mu_x)}{\sqrt{\sigma_y^2/n + \sigma_x^2/n}} \sim N\left(0, 1\right)$$

Application: planning an experiment

- Suppose we know that $\sigma_x^2 = \sigma_y^2 = 0.25$
- What if we wanted the sample difference in means within 0.01 of the true difference in means with high probability?
    - $\rightsquigarrow \mathbb{P}\left(-0.01 \leq \widehat{D}_n - (\mu_y - \mu_x) \leq 0.01\right) = 0.95$
- How big does n need to be to ensure this?
- What is the variance:

$$\mathbb{V}[\widehat{D}_n] = \frac{\sigma_y^2}{n} + \frac{\sigma_x^2}{n} = \frac{0.25}{n} + \frac{0.25}{n} = \frac{1}{2n}$$

Application: planning an experiment

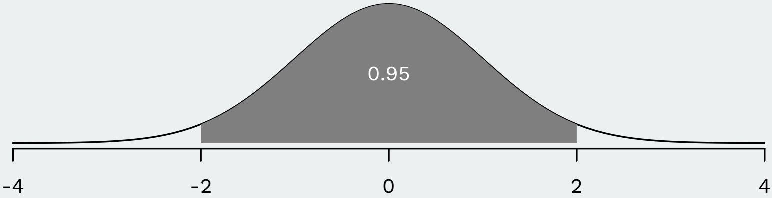
- In large samples, $\widehat{D}_n - (\mu_y - \mu_x)$ will be $N(0, 1/2n)$
- $\rightsquigarrow Z = \sqrt{2n}\left(\widehat{D}_n - (\mu_y - \mu_x)\right) \sim N(0, 1)$
- Then, we want n such that:

$$\mathbb{P}\left(-0.01 \cdot \sqrt{2n} \leq Z \leq 0.01 \cdot \sqrt{2n}\right) = 0.95$$

Application: planning an experiment

$$\mathbb{P}\left(-0.01 \cdot \sqrt{2n} \leq Z \leq 0.01 \cdot \sqrt{2n}\right) = 0.95$$

- But, notice by the empirical rule: $\mathbb{P}(-2 \leq Z \leq 2) = 0.95$



- So, we need $0.01 \cdot \sqrt{2n} > 2$ or $n > 20000$
- 0.95 is convenient with the empirical rule, use `qnorm()` for other values.

6/ Comparing estimators

Efficiency

- **Definition**: If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of θ, then $\hat{\theta}_1$ is **efficient** relative to $\hat{\theta}_2$ when $\mathbb{V}[\hat{\theta}_1] \leq \mathbb{V}[\hat{\theta}_2]$ for any possible value of θ with strict inequality for at least one value of θ.

Bias-Variance Tradeoff

- In many situations, there is tradeoff between bias and variance.
- Extreme example:
    - Sample mean, $\hat{\theta}_1 = \bar{Y}_n$ vs. always-choose-3 $\hat{\theta}_2 = 3$.
    - Variance of always choose 3 is 0, which is better than $\bar{Y}_n$
    - But bias is high if $\mu \neq 3$

Mean Squared Error

- **Definition**: The **mean squared error (MSE)** of an estimator $\hat{\theta}$ for θ is $\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$. We can write this as:

$$\text{MSE}(\hat{\theta}) = \mathbb{V}[\hat{\theta}] + [\text{bias}(\hat{\theta})]^2$$

- **Problem** What is the mean squared error of an unbiased estimator?