Gov 2000 - 4. Sums, Means, and Limit Theorems

Matthew Blackwell
Harvard University
mblackwell@gov.harvard.edu

Where are we? Where are we going?

- **Probability**: formal way to quantify uncertain outcomes/random variables.
- Last week: how to work with **multiple r.v.s** at the same time.
- This week: applying those ideas to study **large random samples**

Large random samples

- In real data, we will have a set of n measurements on a variable:

$$X_1, X_2, \dots, X_n$$

- Or we might have a set of n measurements on two variables:

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$$

- Empirical analyses: sums or means of these n measurements
  - Almost all statistical procedures involve a sum/mean.
  - What are the properties of these sums and means?
  - Can they tell us anything about the distribution of X_i?
- **Asymptotics**: what can we learn as n gets big?

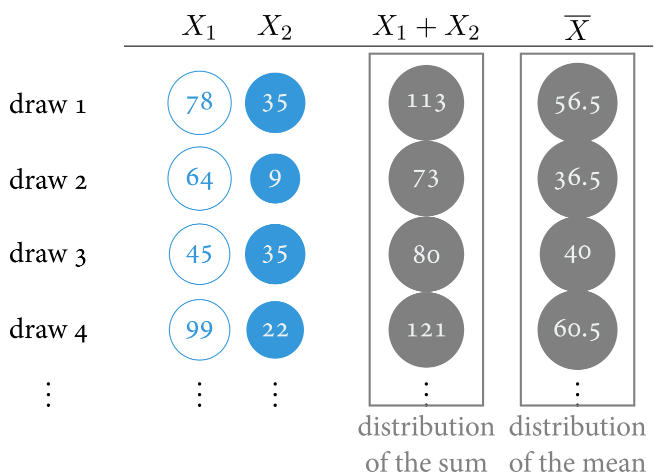
SUMS AND MEANS OF RANDOM VARIABLES

Sums and means are random variables

- If X_1 and X_2 are r.v.s, then $X_1 + X_2$ is a r.v.
  - Has a mean $\mathbb{E}[X_1 + X_2]$ and a variance $\mathbb{V}[X_1 + X_2]$
- The **sample mean** is a function of sums and so it is a r.v. too:

$$\overline{X} = \frac{X_1 + X_2}{2}$$

Distribution of sums/means



Independent and identical r.v.s

- We often will work with **independent and identically distributed** (i.i.d.) r.v.s, $X_1, \ldots, X_n$. The easiest way to justify this assumption is if we are working with a random sample of n respondents from a very large (essentially infinite) population. Remember that "random sample" means that each respondent has equal probability of being picked (think of the R function `sample()`).
- **Example:** X_i is the ith respondent's support for Barack Obama in a sample of n registered voters.
- Mathematically, this implies that each of responses are independent of each other so that $X_i \perp\!\!\!\perp X_j$ for all $i \neq j$. And it implies that the probability distribution (p.m.f./p.d.f.) is exactly the same for all units so that $f_{X_i}(x)$ is the same for

all i. Note that this implies that the expected value and the variance/standard deviation are also the same across units: $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ for all i.

Distribution of the sample mean

- *Sample mean* of i.i.d. r.v.s: $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Hopeful it is clear that $\overline{X}_n$ is a random variable. This leads to a number of questions we might have such as: what is its distribution? What is the expectation of this distribution, $\mathbb{E}[\overline{X}_n]$? What is the variance of this distribution, $\mathbb{V}[\overline{X}_n]$? What is the p.d.f. of the distribution? How do they relate to the expectation, variance of $X_1, \ldots, X_n$?

Example: finite populations and Fulton County

- With finite populations, we can imagine taking a random sample of the population values. Then the population probability distribution is just the distribution of the population data. For instance, suppose our population of interest was all registered voters in Fulton County, GA and we were interested in whether they turned out to vote ($X = 1$) or not ($X = 0$).
- In this case, we want to learn about this specific population, not the general infinite population of potential voters in Fulton County. Conveniently, we have access to data on the entire population of registered voters in Fulton County:

```
## load file of all registered voters in
load("../data/fulton.RData")

## size of the population
nrow(fulton)
```

```
## [1] 339186
```

- The population distribution of turnout can be completely summarized by the expected value, which we can calculate from this data:

```
## calculate the population mean/proportion of
## people turning out this is a little pedantic
## because we are using the defintion of expected
## value
pop.mean <- 0 * sum(fulton$turnout == 0)/nrow(fulton) +
```

```
    1 * sum(fulton$turnout == 1)/nrow(fulton)
pop.mean
```

```
## [1] 0.44
```

- This expected value/mean would represent the (usually unknown) parameter that we want to estimate.
- Let's take a sample of size $n = 10$ from this population, using sampling with replacement (as if we drew names from a hat, putting the names back in after each draw):

```
## set the seed so we can replicate everything!
set.seed(2143)

## quick reminder on how to select certain rows from
## a data frame with a matrix like this, we select
## the rows before the comma, columns after
first.five <- fulton[c(1, 2, 3, 4, 5), "turnout"]
first.five
```

```
## [1] 0 0 0 1 1
```

```
## or more succintly because 1:5 is equivalent to
## c(1,2,3,4,5)
first.five.alt <- fulton[1:5, "turnout"]
first.five.alt
```

```
## [1] 0 0 0 1 1
```

```
## leaving either before or after the comma blank
## gives us all the rows/columns
first.five.allcols <- fulton[1:5, ]
first.five.allcols
```

```
##   turnout black sex age dem rep urban percblk
## 1       0     0   1  19   0   0     0   0.052
## 2       0     0   0  35   0   0     0   0.029
## 3       0     1   0  36   0   0     1   0.992
```

```
## 4       1     0   0  27   0   0     1   0.111
## 5       1     1   1  79   1   0     1   0.992
##   lvbdist school first church
## 1     3.5      0     0      1
## 2     3.3      1     0      0
## 3     2.9      1     0      0
## 4     2.6      0     0      0
## 5     2.8      1     0      0
```

```
## first draw the indices of rows that we want to
## draw: to get a sample of the rows of the data, we
## sample the row indices remember that
## 1:nrow(fulton) is just a list of numbers from 1
## to the number of rows of the data
samp.rows <- sample(1:nrow(fulton), size = 10, replace = TRUE)
samp.rows
```

```
##  [1] 209608 199760 194250 151559 322897   9128
##  [7]  46681  81429 113986 126544
```

```
## now we want to create a new object that is just
## those sampled rows:
f.sample <- fulton[samp.rows, ]
f.sample$turnout
```

```
##  [1] 0 1 1 0 1 0 1 0 1 1
```

- Here we can see that we've sampled 10 observations from the population. Implicitly this also means that we have drawn a random sample of size 10 from the population distribution of turnout.
- The sample mean is just the simple average of all of the values in the sample. Let's calculate this in our Fulton County sample of size 10:

```
mean(f.sample$turnout)
```

```
## [1] 0.6
```

- But here is something to consider: we got a particular sample from the population. What would happen if had gotten a different sample:

```
## first draw the indices of rows that we want to
## draw:
samp.rows2 <- sample(1:nrow(fulton), size = 10, replace = TRUE)
f.sample2 <- fulton[samp.rows2, ]
mean(f.sample2$turnout)
```

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

- These two values are different! Somewhat obviously, this is because the sample mean is a function of the iid random variables. Because these change from sample to sample, then the estimate of the mean (the sample mean) will change from sample to sample.
- Basically, because $\overline{X}$ is a function of random variables, it too is a random variable. So it varies between different random samples of the data and it has its own probability distribution with its own center and spread.

Example: simulating the distribution of the sample mean

- Can we tell what the average of the distribution of the sample mean is? Well, here we can do this by simulation:

1. Take a random sample of size 10 from the population
2. Calculate the sample mean from the sample
3. Repeat 1 and 2 a lot of times (> 10,000) and store the results
4. Take the mean of the resulting distribution of estimates

```
## set up some stuff like a holder for the results
nsims <- 10000
my.means <- rep(NA, times = nsims)

## we're going to do this a bunch of times
for (i in 1:nsims) {
    samp.rows <- sample(1:nrow(fulton), size = 10,
        replace = TRUE)

    ## access the turnout variable for only the sampled
    ## rows and store the results
    my.means[i] <- mean(fulton[samp.rows, "turnout"])
}
```

```r
summary(my.means)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.30    0.40    0.44    0.50    1.00
```

- Here we can see some feature of this sampling distribution. Its mean is actually fairly close to the overall population mean. But sometimes the sample mean was 0 (so there were no voters in the sample) or it was 1 (the sample was all voters).

Mean/variance of the sample mean

- Let's formalize that finding! What is the expectation of the sample mean of the i.i.d. r.v.s?

$$\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i] = \mu$$

- The expectation of the sample mean is just the mean of each observation. This implies that the sample mean $\rightsquigarrow$ right answer on average.
- What about the variance of the sample mean of i.i.d. r.v.s?

$$\mathbb{V}[\overline{X}_n] = \frac{\mathbb{V}[X_i]}{n} = \frac{\sigma^2}{n}$$

- Variance of the sample mean is the variance of each observation divided by the number of observations.
- *Standard error of the sample mean*: $\sqrt{\mathbb{V}[\overline{X}_n]} = \frac{\sigma}{\sqrt{n}}$

USEFUL INEQUALITIES

Why inequalities?

- Behavior of r.v.s depend on their distribution, but we often don't know (or don't want to assume) a distribution. Today, we'll discuss results for r.v. with *any distribution* subject to some restrictions like finite variance.
- In the last few weeks, we have started with a pmf/pdf and then used that to derive the expectation and the variance of some r.v. But what if we don't know the exact form of the distribution but I know the mean and the variance? Can I say anything about the distribution of the r.v. if just know those things? These inequalities are very helpful when thinking about broad classes of r.v.s where we want abstract away from a particular distribution.

- Why study these?
  - Build toward massively important results like LLN
  - Inequalities used regularly throughout statistics
  - Gives us some practice with proofs/analytic reasoning

Markov Inequality

- **Theorem** (Markov Inequality) Suppose that X is r.v. such that $\mathbb{P}(X \geq 0) = 1$. Then, for every real number $t > 0$,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

Chebyshev Inequality

- **Theorem** (Chebyshev Inequality) Suppose that X is r.v. for which $\mathbb{V}[X] < \infty$. Then, for every real number $t > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathbb{V}[X]}{t^2}.$$

Proof of Chebyshev

- Let $Y = (X - \mathbb{E}[X])^2$
  - $\rightsquigarrow \mathbb{P}(Y \geq 0) = 1$ (nonnegative)
  - $\mathbb{E}[Y] = \mathbb{V}[X]$ (definition of variance)
- Note that if $|X - \mathbb{E}[X]| \geq t$ then $Y \geq t^2$ because we just squared both sides.
- Thus, $\mathbb{P}(|X - \mathbb{E}[X]| \geq t) = \mathbb{P}(Y \geq t^2)$
- Apply Markov's inequality:

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) = \mathbb{P}(Y \geq t^2) \leq \frac{\mathbb{E}[Y]}{t^2} = \frac{\mathbb{V}[X]}{t^2}$$

Application: planning a survey

- Suppose we want to estimate the proportion of voters who will vote for Donald Trump, p, from a random sample of size n. We'll say that $X_1, X_2, \ldots, X_n$ indicating voting intention for Trump for each respondent.

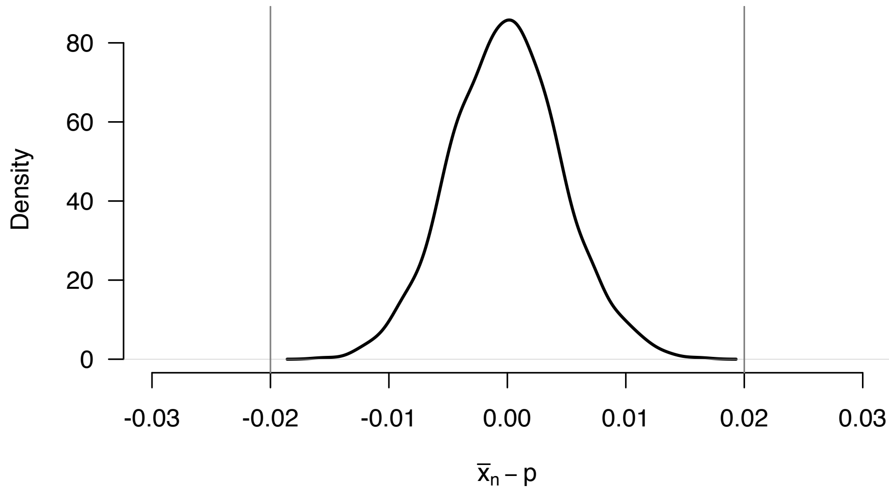
- By our earlier, calculation, $\mathbb{E}[\overline{X}_n] = p$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$ and since this is a Bernoulli r.v., we know that $\sigma^2 = p(1-p)$.
- **Question**: What does n need to be to have at least 0.95 probability that $\overline{X}_n$ is within 0.02 of the true p?
- Noting $\sigma^2 \leq 1/4$ and applying Chebyshev:

$$\mathbb{P}(|\overline{X}_n - p| \geq 0.02) \leq \frac{\mathbb{V}[\overline{X}_n]}{0.02^2} = \frac{p(1-p)}{0.0004n} \leq \frac{1}{0.0016n}$$

- We want this probability to be bounded by 0.05 so we need $(1/0.0016n) \leq 0.05$, which gives us $n \geq 12,500$!!
- Do we really need $n \geq 12,500$ to get a margin of error of ± 2 percentage points?
- **No!** Chebyshev provides a bound that is guaranteed to hold, but actual probabilities are much smaller. We're also using the "worst-case" variance of 0.25.
- Let's simulate 1000 samples of size $n = 12500$ with $p = 0.4$ and show the distribution of the means. What proportion of these are within 0.02 of p?

```
nsims <- 1000
holder <- rep(NA, times = nsims)
for (i in 1:nsims) {
    this.samp <- rbinom(n = 12500, size = 1, prob = 0.4)
    holder[i] <- mean(this.samp)
}
mean(abs(holder - 0.4) > 0.02)
```

```
## [1] 0
```



LAW OF LARGE NUMBERS

Current knowledge

- Recap: we have i.i.d. r.v.s $X_1, \ldots, X_n$
- Current knowledge about the distribution of $\overline{X}_n$:
  - Expectation is $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i] = \mu$
  - Variance is $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$ where $\sigma^2 = \mathbb{V}[X_i]$
  - Some bounds on tail probabilities from Chebyshev.
  - None of these rely on a *specific distribution* for X_i!
- Can we say more about the distribution of the sample mean?
- Yes, but we need to think about how $\overline{X}_n$ changes as n gets big.

Sequence of sample means

- What can we say about the sample mean n gets large?

- Need to think about sequences of sample means with increasing n:

$$\begin{aligned}
\overline{X}_1 &= X_1 \\
\overline{X}_2 &= (1/2) \cdot (X_1 + X_2) \\
\overline{X}_3 &= (1/3) \cdot (X_1 + X_2 + X_3) \\
\overline{X}_4 &= (1/4) \cdot (X_1 + X_2 + X_3 + X_4) \\
\overline{X}_5 &= (1/5) \cdot (X_1 + X_2 + X_3 + X_4 + X_5) \\
&\vdots \\
\overline{X}_n &= (1/n) \cdot (X_1 + X_2 + X_3 + X_4 + X_5 + \cdots + X_n)
\end{aligned}$$

- Note: this is a sequence of random variables!

Convergence in Probability

- **Definition**: A sequence of random variables, $Z_1, Z_2, \ldots$, is said to *converge in probability* to a value b if for every $\varepsilon > 0$,

$$\mathbb{P}(|Z_n - b| > \varepsilon) \to 0,$$

  as $n \to \infty$. We write this $Z_n \xrightarrow{p} b$.

- Basically: probability that Z_n lies in any (teeny, tiny) interval around b approaches 1 as $n \to \infty$. Wooldridge writes $\text{plim}(Z_n) = b$ if $Z_n \xrightarrow{p} b$.

Law of large numbers

- **Theorem** (Weak Law of Large Numbers) Let $X_1, \ldots, X_n$ be a an i.i.d. draws from a distribution with mean μ and finite variance σ^2. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then, $\overline{X}_n \xrightarrow{p} \mu$.

- **Intuition**: The probability of $\overline{X}_n$ being "far away" from μ goes to 0 as n gets big.

LLN proof

- Proof: by Chebyshev, we have

$$\mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

- which implies

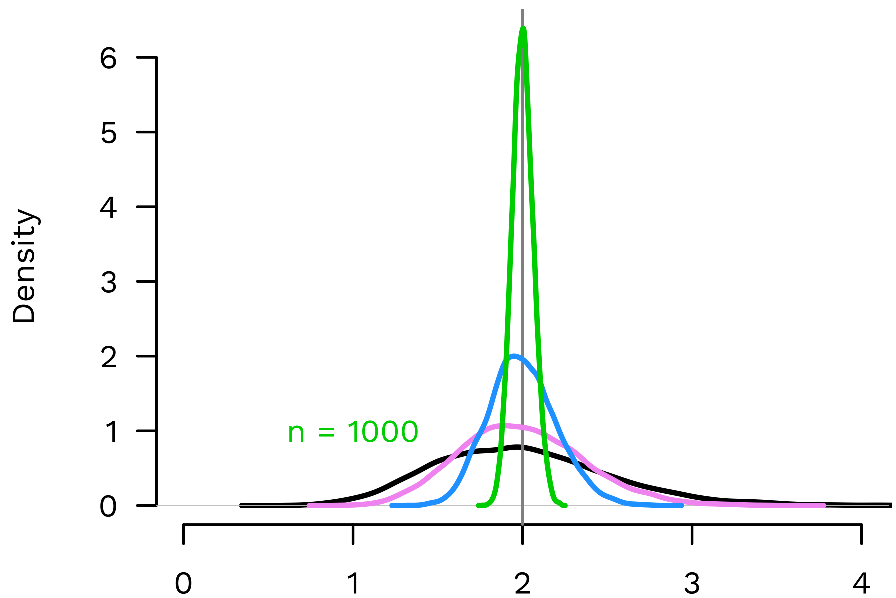
$$\mathbb{P}(|\overline{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$$

- As $n \to \infty$, the right hand side equals 1 which implies

$$\lim_{n \to \infty} \mathbb{P}(|\overline{X}_n - \mu| < \varepsilon) = 1$$

LLN by simulation in R

```
nsims <- 10000
holder <- matrix(NA, nrow = nsims, ncol = 6)
for (i in 1:nsims) {
    s5 <- rexp(n = 5, rate = 0.5)
    s15 <- rexp(n = 15, rate = 0.5)
    s30 <- rexp(n = 30, rate = 0.5)
    s100 <- rexp(n = 100, rate = 0.5)
    s1000 <- rexp(n = 1000, rate = 0.5)
    s10000 <- rexp(n = 10000, rate = 0.5)

    holder[i, 1] <- mean(s5)
    holder[i, 2] <- mean(s15)
    holder[i, 3] <- mean(s30)
    holder[i, 4] <- mean(s100)
    holder[i, 5] <- mean(s1000)
    holder[i, 6] <- mean(s10000)
}
```



Properties of convergence in probability

1. if $X_n \xrightarrow{p} c$, then $g(X_n) \xrightarrow{p} g(c)$ for any continuous function g.
2. if $X_n \xrightarrow{p} a$ and $Z_n \xrightarrow{p} b$, then
   - $X_n + Z_n \xrightarrow{p} a + b$
   - $X_n Z_n \xrightarrow{p} ab$
   - $X_n / Z_n \xrightarrow{p} a/b$ if $b > 0$

- Thus, by LLN:
  - $(\overline{X}_n)^2 \xrightarrow{p} \mu^2$
  - $\log(\overline{X}_n) \xrightarrow{p} \log(\mu)$

CENTRAL LIMIT THEOREM

Recap

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
  - $\overline{X}_n$ converges to μ as n gets big
  - Chebyshev provides some bounds on probabilities.
  - Still no distributional assumptions about X_i!

- Can we say more?
  - Can we approximate $\Pr(a < \overline{X}_n < b)$?
  - What family of distributions (Binomial, Uniform, Gamma, etc)?
- Again, need to analyze when n is large?

Convergence in Distribution

- **Definition** Let $Z_1, Z_2, \ldots$, be a sequence of r.v.s, and for $n = 1, 2, \ldots$ let $F_n(z)$ be the c.d.f. of Z_n. Then it is said that $Z_1, Z_2, \ldots$ **converges in distribution** to r.v. W with c.d.f. F_W if

$$\lim_{n \to \infty} F_n(x) = F_W(x),$$

  which we write as $Z_n \overset{d}{\to} W$.
- Basically: when n is big, the distribution of Z_n is very similar to the distribution of W
- If $X_n \overset{p}{\to} X$, then $X_n \overset{d}{\to} X$

Standardizing an r.v.

- It is very to *standardize* a r.v. by subtracting its expectation and dividing by its standard deviation:

$$Z = \frac{X - \mathbb{E}[X]}{\sqrt{\mathbb{V}[X]}}$$

- It is easy to show that for any X, we have:
  - $\mathbb{E}[Z] = \mathbb{E}[(1/\sqrt{\mathbb{V}[X]})(X - \mathbb{E}[X])] = (1/\sqrt{\mathbb{V}[X]})\,(\mathbb{E}[X] - \mathbb{E}[X]) = 0$
  - $\mathbb{V}[Z] = 1$ (try to prove this for yourself)
- Sometimes called a z-score.

Central Limit Theorem

- **Theorem (Central Limit Theorem)** Let $X_1, \ldots, X_n$ be a an iid draws from a distribution with mean μ and variance $\sigma^2 \leq \infty$. Then

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \overset{d}{\to} N(0, 1).$$

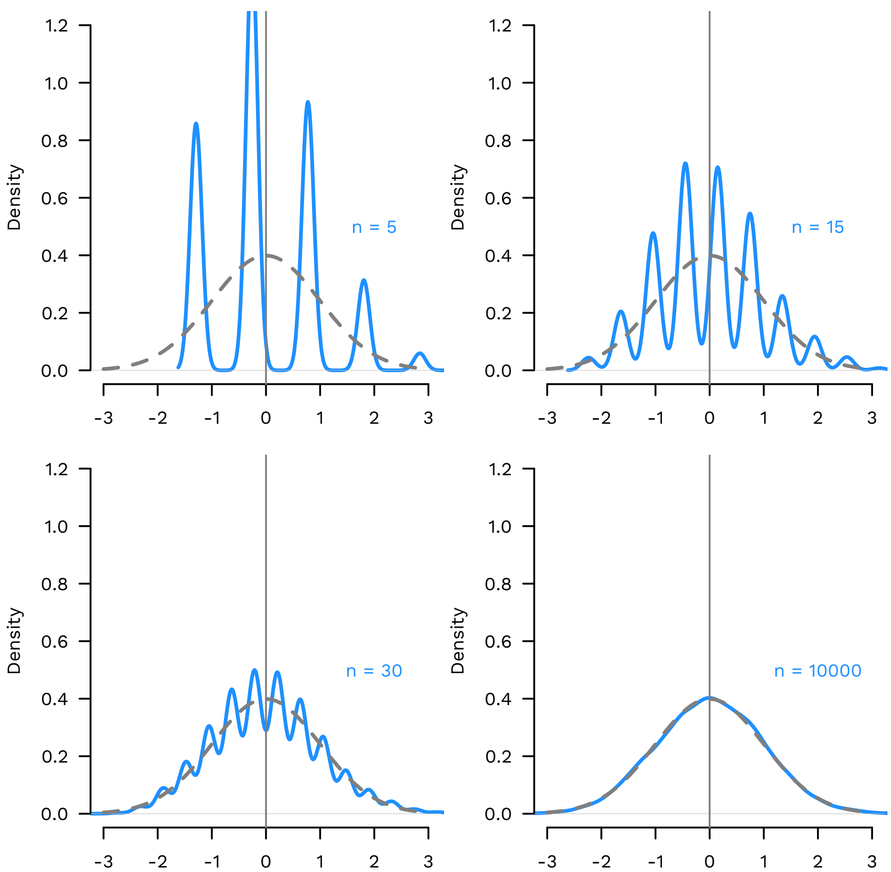
- Distribution free! We don't have to make specific assumptions about the distribution of X_i
- Implies that $\overline{X}_n \sim N(\mu, \sigma^2/n)$
  - $\rightsquigarrow$ easy approximations to probability statements about $\overline{X}_n$ when n is big!

CLT by simulation in R

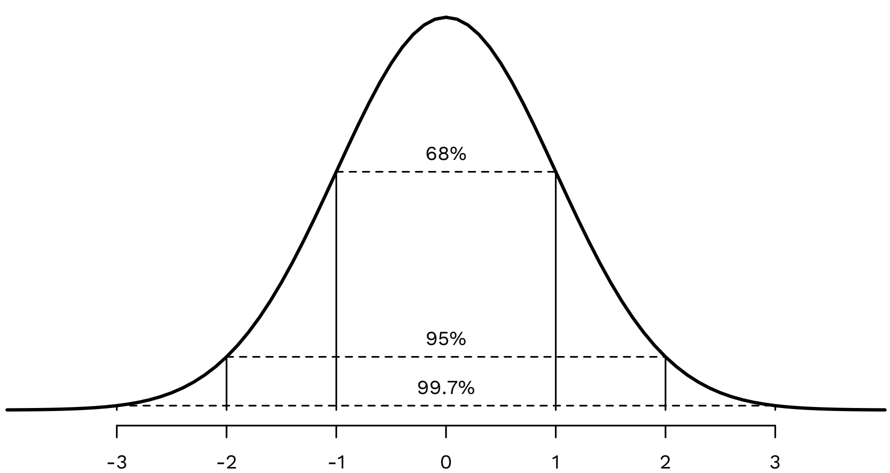
```
nsims <- 10000
holder2 <- matrix(NA, nrow = nsims, ncol = 6)
for (i in 1:nsims) {
    s5 <- rbinom(n = 5, size = 1, prob = 0.25)
    s15 <- rbinom(n = 15, size = 1, prob = 0.25)
    s30 <- rbinom(n = 30, size = 1, prob = 0.25)
    s100 <- rbinom(n = 100, size = 1, prob = 0.25)
    s1000 <- rbinom(n = 1000, size = 1, prob = 0.25)
    s10000 <- rbinom(n = 10000, size = 1, prob = 0.25)

    holder2[i, 1] <- mean(s5)
    holder2[i, 2] <- mean(s15)
    holder2[i, 3] <- mean(s30)
    holder2[i, 4] <- mean(s100)
    holder2[i, 5] <- mean(s1000)
    holder2[i, 6] <- mean(s10000)
}
```

CLT in action

Distribution of $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ at various levels of n:



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.

Simulating the empirical rule

- Actual probability of $Z \sim N(0,1)$ between -2 and 2:

```
pnorm(2) - pnorm(-2)
```

```
## [1] 0.9545
```

- Simulated probability of $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ between -2 and 2:
  - $n = 15 \rightsquigarrow$ 0.97
  - $n = 30 \rightsquigarrow$ 0.97
  - $n = 100 \rightsquigarrow$ 0.95
  - $n = 1000 \rightsquigarrow$ 0.96
  - $n = 10000 \rightsquigarrow$ 0.96
- Quality of the approximation depends on the underlying distribution of the X_i
  - Obviously if $X_i \sim N(0,1)$ it's going to be perfect with $n = 1$

Slutsky's Theorem

- Let $X_1, X_2, \ldots$ converge in distribution to some r.v. X
- Let $Y_1, Y_2, \ldots$ converge in probability to some number, c
- Slutsky's Theorem gives the following result:
  1. $X_n Y_n$ converges in distribution to cX
  2. $X_n + Y_n$ converges in distribution to $X + c$
- Extremely useful when trying to figure out what the large-sample distribution of an estimator is.

Application: planning a survey

- Trump example: we want the the probability of being within 0.02 from the true p to be 95%.
- $\rightsquigarrow$ we want n such that:

$$\mathbb{P}(\overline{X}_n - p < -0.02) + \mathbb{P}(\overline{X}_n - p > 0.02) \leq 0.05$$

- By the CLT, if n is large, then

$$\overline{X}_n - p \approx N\left(0, \frac{\sigma^2}{n}\right)$$

- We don't know σ^2!
  - X_i is Bernoulli so $\sigma^2 \leq 1/4$.
  - $\rightsquigarrow$ conservative to take $\sigma^2 = 1/4$ (prove this to yourself)
  - $\overline{X}_n - p \approx N\left(0, \frac{1}{4n}\right)$
- Standardizing: $\rightsquigarrow Z = 2\sqrt{n}(\overline{X}_n - p) \sim N(0, 1)$

Application: planning a survey

- By symmetry: $\mathbb{P}(\overline{X}_n - p < -0.02) = \mathbb{P}(\overline{X}_n - p > 0.02)$ so:

$$2 \times \mathbb{P}(\overline{X}_n - p < -0.02) \leq 0.05$$

- Or

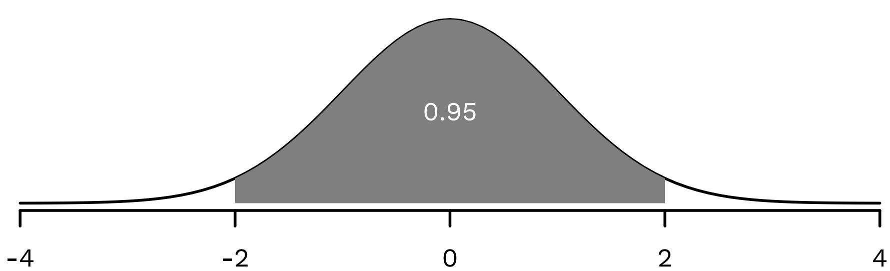
$$\mathbb{P}(\overline{X}_n - p < -0.02) \leq 0.025$$

- Since $\overline{X}_n - p \leq -0.02$ implies $Z \leq -0.02 \times 2\sqrt{n}$,

$$\mathbb{P}(\overline{X}_n - p < -0.02) = \mathbb{P}(Z < -0.02 \times 2\sqrt{n})$$

Application: planning an experiment

- We need: $\mathbb{P}\left(Z < -0.02 \times 2\sqrt{n}\right) \leq 0.025$
- By the empirical rule: $\mathbb{P}(-2 \leq Z \leq 2) = 0.95$ so $\Pr(Z < -2) = 0.025$



- So, we need $-0.04\sqrt{n} \leq -2$ or $n > 20000$
- Solving we get $n \geq$ 2,500—much lower than the 12,500 from Chebyshev.

MORE EXOTIC CLTS*

CLT for non-iid r.v.s

- What if we don't have i.i.d. r.v.s? Does the CLT still apply?
- Let $X_1, X_2, \ldots$ be independent (but not identically distributed) with means $\mathbb{E}[X_i] = \mu_i$ and variances $\mathbb{V}[X_i] = \sigma_i^2$.
- Scaled and centered:

$$Y_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}}$$

  - No need to divide by n because there are n entries in the sum $\sum_{i=1}^n \mu_i$

- Easy to show that $\mathbb{E}[Y_n] = 0$ and $\mathbb{V}[Y_n] = 1$. Does the CLT apply?

Liapounov CLT

- **Theorem (Liapounov CLT)** Suppose that the r.v.s $X_1, X_2, \ldots$ are independent and that $\mathbb{E}[|X_i - \mu_i|^3] < \infty$ for $i = 1, 2, \ldots$. Also, suppose that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n \mathbb{E}\left[|X_i - \mu_i|^3\right]}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} = 0.$$

  Then,

$$Y_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}} \xrightarrow{d} N(0,1)$$

- Key condition: there isn't one r.v.s in the sequence that is "too big" that could dominate the sum

CLT for dependent sequences

- We have shown the CLT for i.i.d. and for independent r.v.s. What about dependent sequences?

- CLT works for a dependent sequence $X_1, X_2, \ldots$.

  - What does dependent sequence mean? Cov$[X_i, X_j] \neq 0$

- *Key condition for dependent CLT*: r.v.s aren't "too correlated"
- Overall conditions for CLT to hold: the sum/mean of many, not too correlated, not too big r.v.s

WRAP-UP

Review

- Sums and means of r.v.s are themselves r.v.s
- Learned about the distribution of the sample mean of i.i.d. r.v.s
  - Expectation $\mathbb{E}[\overline{X}_n] = \mu$
  - Variance $\mathbb{V}[\overline{X}_n] = \sigma^2/n$
  - Converges in probability to true mean (LLN)
  - Converges in distribution to a normal distribution (CLT)
- Ahead: generalizing these ideas to arbitrary estimators of parameters.

TECHNICAL DETAILS

Markov Inequality Proof

- For discrete X:

$$\mathbb{E}[X] = \sum_x x f_X(x) = \sum_{x<t} x f_X(x) + \sum_{x \geq t} x f_X(x)$$

- Because X is nonnegative, $\mathbb{E}[X] \geq \sum_{x \geq t} x f_X(x)$
- Since $x \geq t$, then $\sum_{x \geq t} x f_X(x) \geq \sum_{x \geq t} t f_X(x)$
- But this is just $\sum_{x \geq t} t f_X(x) = t \sum_{x \geq t} f_X(x) = t\mathbb{P}(X \geq t)$
- Implies $\mathbb{E}[X] \geq t\mathbb{P}(X \geq t)$