

450A Political Methodology I

Lecture 2: Univariate Statistical Inference

1. Point Estimation

Stanford University

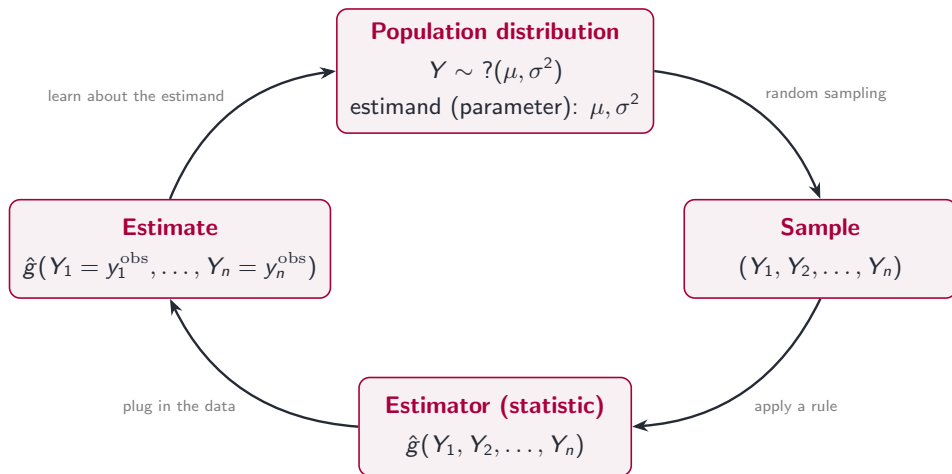
What is Point Estimation?

Finite-sample Properties of Estimators

Asymptotic Properties of Estimators

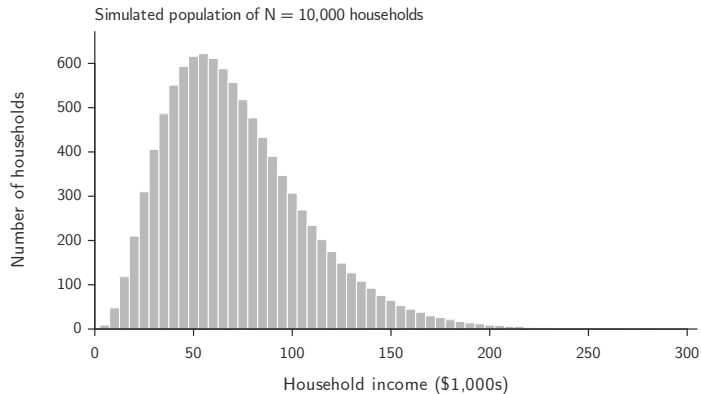
Estimands, Estimators, and Estimates

- The typical goal of statistical inference is to use sample data to learn about an unobserved population distribution.
- The unobserved population distribution is typically characterized by **parameters**. Often denoted with Greek letters (e.g. μ, σ^2, θ).
- **Estimands** are the parameters of the unobserved population distribution that we aim to estimate from the sample data.
- **Estimators** are functions of sample data (i.e. **statistics**) which we use to learn about the estimands. Often denoted with a “hat” (e.g. $\hat{\mu}, \hat{\theta}$)
- **Estimates** are particular values of estimators that are realized in a given sample.



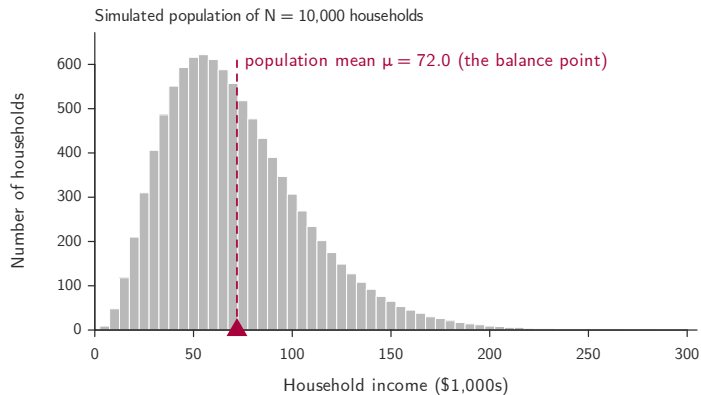
Example: Household Incomes

Assume a population of N units indexed by i with incomes $y_1, y_2, \dots, y_N$ (fixed numbers).



Let $\mu = \frac{1}{N} \sum_{i=1}^N y_i$ be the population mean and $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2$ the population variance.

Defining the Estimand



Let's consider the problem of using sample data from this population to learn about its mean (i.e. the **true mean**), μ .

Using Random Samples to Estimate Population Parameters

Let's look at possible estimators $\hat{\mu}$ for the (unobserved) mean income in the population μ .

How can we use sample data to estimate μ ?

If we think of a sample of size n as **randomly sampled** with replacement from the population, then $Y_1, \dots, Y_n$ are **independently and identically distributed** (i.i.d.) random variables with $E[Y_i] = \mu$ and $V[Y_i] = \sigma^2$ for all $i \in \{1, \dots, n\}$ (each Y_i is one of $y_1, \dots, y_N$, picked at random).

In R, we can draw an i.i.d. random sample of size n from population using

```
sample(population, size=n, replace=TRUE)
```

where `population` is a vector that contains the y_i values for all N units in the population.

Our estimators, $\hat{\mu}$, are functions of $Y_1, \dots, Y_n$ and will therefore be random variables with their own probability distributions.

Candidate estimators:

- ① $\hat{\mu}_1 = Y_1$ (the first data observation)
- ② $\hat{\mu}_2 = 4$ (April, the first full month of the baseball season)
- ③ $\hat{\mu}_3 = \bar{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n)$ (the sample average)
- ④ $\hat{\mu}_4 = \tilde{Y}_n \equiv \frac{1}{n+5}(Y_1 + \cdots + Y_n)$ (an “adjusted” sample average)

How do we choose between these estimators?

Properties of Estimators

Estimators are random variables, for which randomness comes from **repeated sampling** from the population.

The distribution of an estimator due to repeated sampling is called the **sampling distribution**.

The properties of an estimator refer to the characteristics of its sampling distribution.

Finite-sample Properties (apply for any sample size):

- **Unbiasedness**: Is the sampling distribution of our estimator centered at the true parameter value? $E[\hat{\theta}] = \theta$
- **Efficiency**: Among unbiased estimators, which has the smaller variance? $V[\hat{\theta}_A] < V[\hat{\theta}_B]$

Asymptotic Properties (kick in when n is large):

- **Consistency**: As our sample size grows to infinity, does the sampling distribution of our estimator converge to the true parameter value?
- **Asymptotic Normality**: As our sample size grows large, does the sampling distribution of our estimator approach a normal distribution?

Evaluating Sampling Distributions

Two ways to evaluate the sampling distributions of estimators:

① Analytic:

- Use formal rules of mathematics and probability to derive properties
- Guaranteed to be correct, no ambiguity (unless you screw up the math!)
- Might be complicated or even infeasible

② Simulation:

- Do “repeated experiments” on your computer
- Never exactly correct (cannot repeat ∞ times)
- Only works for specific scenarios (e.g. you have to assume a specific population)
- Easy to implement and often precise enough

Both should yield essentially the same result.

Use one as a check on the other.

Simulating Sampling Distributions

Let's simulate the sampling distribution of our four candidate estimators for the population mean:

① $\hat{\mu}_1 = Y_1$

② $\hat{\mu}_2 = 4$

③ $\hat{\mu}_3 = \bar{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n)$

④ $\hat{\mu}_4 = \tilde{Y}_n \equiv \frac{1}{n+5}(Y_1 + \cdots + Y_n)$

Algorithm to simulate the sampling distribution of $\hat{\mu}$

- ① Randomly sample m datasets of size n with replacement from the population: (m should be large, choose n as desired)
- ② Code a function that implements the estimator and use it to calculate the estimate $\hat{\mu}$ for each sampled dataset
- ③ Plot the distribution of estimates $\hat{\mu}$ with a histogram

Simulating Sampling Distributions

R Code

```
> # population: simulated incomes of 10,000 households ($1,000s)
> length(incomes)
[1] 10000
> incomes[1:30]
[1] 71.4 125.3 38.2 108.8 53.7 65.0 49.0 66.8 95.8 90.4 51.6
[12] 159.9 90.6 45.4 52.8 115.2 63.6 49.2 46.3 108.1 52.8 66.0
[23] 49.0 115.8 50.5 37.4 33.2 24.7 85.3 62.7
> set.seed(2026)
> # take one random sample
> income.sample <- sample(x=incomes, size=20, replace=TRUE)
> income.sample
[1] 46.5 35.5 123.1 29.7 67.9 91.2 103.6 40.4 106.1 25.3 129.0
[12] 85.8 66.6 32.0 143.0 63.9 54.0 140.9 12.8 57.9
> # estimator for sample mean
> mean.estimator <- function(x){
+   out <- sum(x)/length(x)
+   return(out)
+ }
> # apply estimator to sample
> mean.estimator(income.sample)
[1] 72.76
```

Simulating Sampling Distributions: For Loop

R Code

```
> # set parameters
> m <- 100000; n <- 20
> # set up storage matrix for means
> store.means <- matrix(NA, m, 1)
> store.means[1:6, 1]
[1] NA NA NA NA NA NA
> # simulation starts here
> for(i in 1:m){
+   income.sample <- sample(x=incomes, size=n, replace=TRUE)
+   store.means[i, 1] <- mean.estimator(income.sample)
+ }
> store.means[1:6]
[1] 67.315 68.875 70.730 81.975 69.175 66.700
> hist(store.means)
```

Simulating Sampling Distributions: Using Matrix Functions

R Code

```
> # take m samples of size n
> income.samples <- replicate(m,
+                             sample(x=incomes, size=n, replace=TRUE)
+                             )
> dim(income.samples)
[1]      20 100000
> income.samples[1:3, 1:4]
      [,1] [,2] [,3] [,4]
[1,] 132.6 147.9 123.2 41.7
[2,]  18.1  77.9  17.4 52.8
[3,]  98.0  30.8  97.4 37.5
> # compute estimator for each sample
> store.means <- apply(income.samples, 2, mean.estimator)
> store.means[1:6]
[1] 67.495 76.795 63.770 69.725 69.620 80.045
```

Sampling Distributions ($m = 100,000$; $n = 20$)

100,000 samples of $n = 20$

$$\hat{\mu}_1 = Y_1$$



$$\hat{\mu}_2 = 4$$

always 4



$$\hat{\mu}_3 = \bar{Y}_n$$



$$\hat{\mu}_4 = \tilde{Y}_n$$



0 50 100 150 200 250 300
Estimate of mean income (\$1,000s)

What is Point Estimation?

Finite-sample Properties of Estimators

Unbiasedness

Efficiency

Asymptotic Properties of Estimators

Definition (Bias)

Bias is the expected difference between an **estimator** $\hat{\theta}$ and a **parameter** θ :

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta} - \theta] = E[\hat{\theta}] - \theta$$

Note that bias is **not** the difference between an estimate (= a particular realized value of an estimator) and a parameter.

Definition (Unbiasedness)

An estimator is **unbiased** iff:

$$\text{Bias}(\hat{\theta}) = 0 \iff E[\hat{\theta}] = \theta$$

To check unbiasedness, we need to compute the expected value of the estimator.

Deriving the Expected Values of Estimators

Our candidate estimators:

① $\hat{\mu}_1 = Y_1$

② $\hat{\mu}_2 = 4$

③ $\hat{\mu}_3 = \overline{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n)$

④ $\hat{\mu}_4 = \tilde{Y}_n \equiv \frac{1}{n+5}(Y_1 + \cdots + Y_n)$

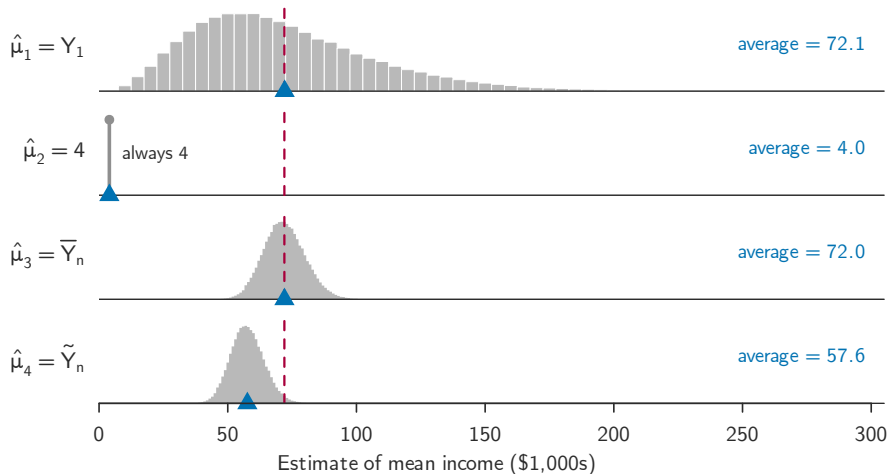
Worksheet Q1: Which of $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$ are unbiased?

Simulating the Expected Values of Estimators

100,000 samples of $n = 20$

dashed line: true mean

triangle: average of the estimates



What is Point Estimation?

Finite-sample Properties of Estimators

Unbiasedness

Efficiency

Asymptotic Properties of Estimators

- How should we choose among unbiased estimators?
- All else equal, we prefer estimators that have a sampling distribution with smaller variance.

Definition (Efficiency)

If $\hat{\theta}_A$ and $\hat{\theta}_B$ are two **unbiased** estimators of θ , then $\hat{\theta}_A$ is **more efficient** relative to $\hat{\theta}_B$ iff

$$V[\hat{\theta}_A] < V[\hat{\theta}_B]$$

- Under repeated sampling, estimates based on $\hat{\theta}_A$ are closer to θ on average (smaller expected squared distance)
- Note that this does **not** imply that a particular estimate is always close to the true parameter value
- The standard deviation of the sampling distribution of an estimator, $\sqrt{V[\hat{\theta}]}$, is often called the **standard error** of the estimator

Deriving the Variances of Estimators

① $\hat{\mu}_1 = Y_1$

② $\hat{\mu}_2 = 4$

③ $\hat{\mu}_3 = \bar{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n)$

④ $\hat{\mu}_4 = \tilde{Y}_n \equiv \frac{1}{n+5}(Y_1 + \cdots + Y_n)$

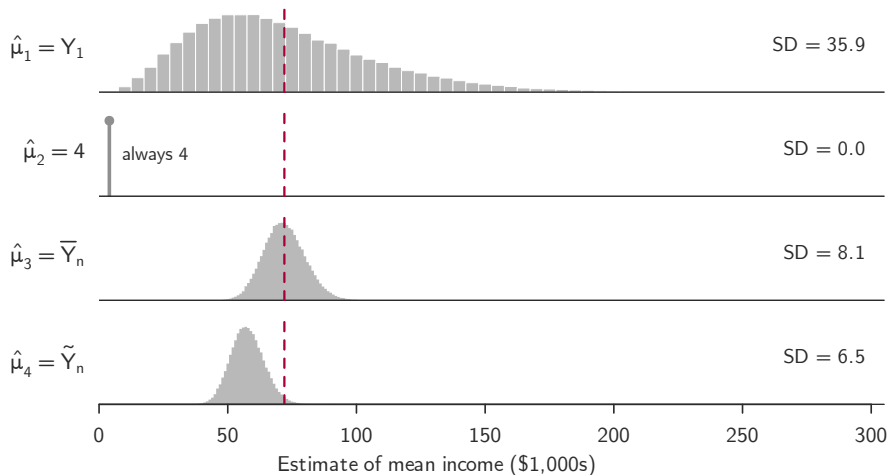
Worksheet Q2: Derive the variances of $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$. What do you conclude?

Simulating the Variances of Estimators

100,000 samples of $n = 20$

dashed line: true mean

SD: standard deviation of the estimates



Mean Squared Error

How can we choose between an unbiased estimator and a biased estimator with smaller variance?

Definition (Mean Squared Error)

To compare estimators in terms of both efficiency and unbiasedness we can use the **Mean Squared Error** (MSE), the expected squared difference between $\hat{\theta}$ and θ :

$$MSE(\hat{\theta}) = E \left[(\hat{\theta} - \theta)^2 \right] = Bias(\hat{\theta})^2 + V(\hat{\theta})$$

Worksheet Q3–Q4:

- Q3: Prove the decomposition above.
- Q4: Show: $MSE(\hat{\mu}_1) = \sigma^2$, $MSE(\hat{\mu}_2) = (4 - \mu)^2$, $MSE(\hat{\mu}_3) = \frac{\sigma^2}{n}$, and $MSE(\hat{\mu}_4) = \frac{25\mu^2 + n\sigma^2}{(n+5)^2}$.

What is Point Estimation?

Finite-sample Properties of Estimators

Asymptotic Properties of Estimators

- Consistency

- Asymptotic Normality

- Unbiasedness and efficiency are **finite-sample properties** of estimators, which hold regardless of sample size
- Estimators also have **asymptotic properties**, i.e., the characteristics of sampling distributions when sample size becomes infinitely large
- To define asymptotic properties, consider a **sequence** of estimators at increasing sample sizes:

$$\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n$$

- For example, for i.i.d. draws $X_1, X_2, \dots$ (such as the incomes Y_i), the sequence of sample means ($\bar{X}_n$) is defined as:

$$\bar{X}_1, \bar{X}_2, \dots, \bar{X}_n = X_1, \frac{X_1 + X_2}{2}, \dots, \frac{X_1 + \dots + X_n}{n}$$

- Asymptotic properties of an estimator are defined by the behavior of $\hat{\theta}_1, \dots, \hat{\theta}_n$ when n goes to infinity.

Stochastic Convergence

- When a sequence of random variables stabilizes to a certain probabilistic behavior as $n \rightarrow \infty$, the sequence is said to show **stochastic convergence**.
- Two types of stochastic convergence are of particular importance:
 - ① **Convergence in probability**: the probability of an “unusual” value becomes increasingly smaller as the sequence progresses such that values eventually stabilize in taking on a constant value (i.e. the **limiting distribution** is a constant)
 - ② **Convergence in distribution**: values in the sequence continue to vary, but the variation eventually comes to follow an unchanging distribution as the sequence progresses (i.e. the limiting distribution is a well characterized distribution)

Definition (Convergence in Probability)

A sequence $W_1, W_2, \dots$ of random variables **converges in probability** towards a real number a if, for all accuracy levels $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \Pr(|W_n - a| \geq \varepsilon) = 0$$

We write this as

$$W_n \xrightarrow{p} a \quad \text{or} \quad \text{plim}_{n \rightarrow \infty} W_n = a.$$

- As n increases, almost all of the PDF of W_n will be concentrated in the ε -interval around a , $[a - \varepsilon, a + \varepsilon]$
- A sufficient condition for convergence in probability:

$$E[W_n] \rightarrow a \quad \text{and} \quad V[W_n] \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty$$

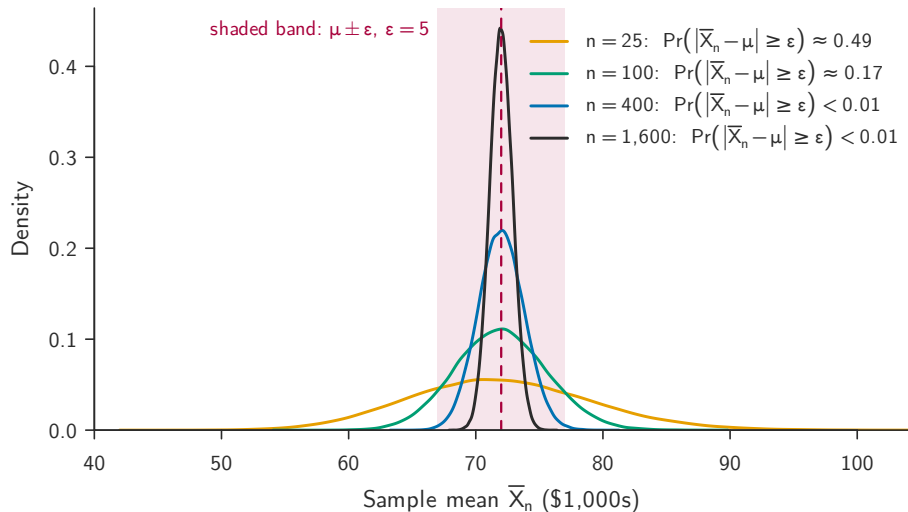
- For example, if the X_i are i.i.d. with finite variance σ^2 , the sample mean $\bar{X}_n$ converges to the population mean μ in probability because

$$E[\bar{X}_n] \rightarrow \mu \quad \text{and} \quad V[\bar{X}_n] = \sigma^2/n \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty$$

- The condition is sufficient, not necessary. Take $W_n = n$ with probability $1/n$ and $W_n = 0$ otherwise:

$$\Pr(|W_n - 0| \geq \varepsilon) \leq 1/n \rightarrow 0, \text{ so } W_n \xrightarrow{p} 0, \quad \text{but} \quad E[W_n] = 1 \quad \text{and} \quad V[W_n] = n - 1 \rightarrow \infty$$

Sample Mean Converges to Population Mean in Probability



Definition

An estimator $\hat{\theta}_n$ is **consistent** if the sequence $\hat{\theta}_1, \dots, \hat{\theta}_n$ converges in probability to the true parameter value θ as the sample size n grows to infinity:

$$\hat{\theta}_n \xrightarrow{P} \theta \quad \text{or} \quad \text{plim}_{n \rightarrow \infty} \hat{\theta}_n = \theta$$

- Often seen as a minimal requirement for estimators
- A consistent estimator may still perform badly in small samples
- Two ways to verify consistency:
 - ① Analytic: Often easier to check if $E[\hat{\theta}_n] \rightarrow \theta$ and $V[\hat{\theta}_n] \rightarrow 0$
 - ② Simulation: Increase n and see how the sampling distribution changes
- Does unbiasedness imply consistency?
- Does consistency imply unbiasedness?

Our candidate estimators:

- ① $\hat{\mu}_1 = Y_1$
- ② $\hat{\mu}_2 = 4$
- ③ $\hat{\mu}_3 = \bar{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n)$
- ④ $\hat{\mu}_4 = \tilde{Y}_n \equiv \frac{1}{n+5}(Y_1 + \cdots + Y_n)$

Which of these estimators are consistent for μ ? (The subscript on $\hat{\mu}_k$ labels the estimator; each is computed at every sample size n .)

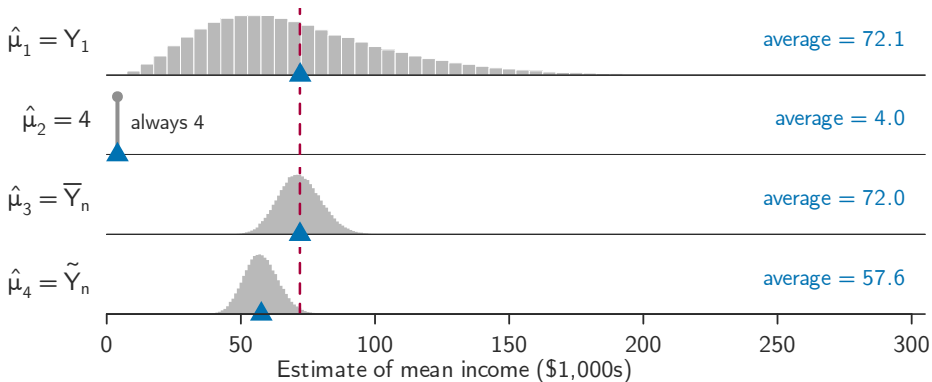
- ① $E[\hat{\mu}_1] = \mu$ and $V[\hat{\mu}_1] = \sigma^2$
- ② $E[\hat{\mu}_2] = 4$ and $V[\hat{\mu}_2] = 0$
- ③ $E[\hat{\mu}_3] = \mu$ and $V[\hat{\mu}_3] = \frac{1}{n}\sigma^2$
- ④ $E[\hat{\mu}_4] = \frac{n}{n+5}\mu$ and $V[\hat{\mu}_4] = \frac{n}{(n+5)^2}\sigma^2$

Sampling Distributions of Estimators ($n = 20$)

100,000 samples of $n = 20$

dashed line: true mean

triangle: average of the estimates

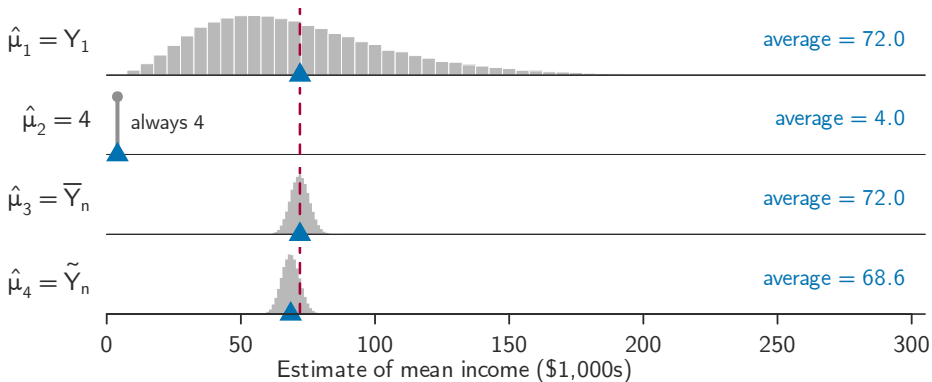


Sampling Distributions of Estimators ($n = 100$)

100,000 samples of $n = 100$

dashed line: true mean

triangle: average of the estimates

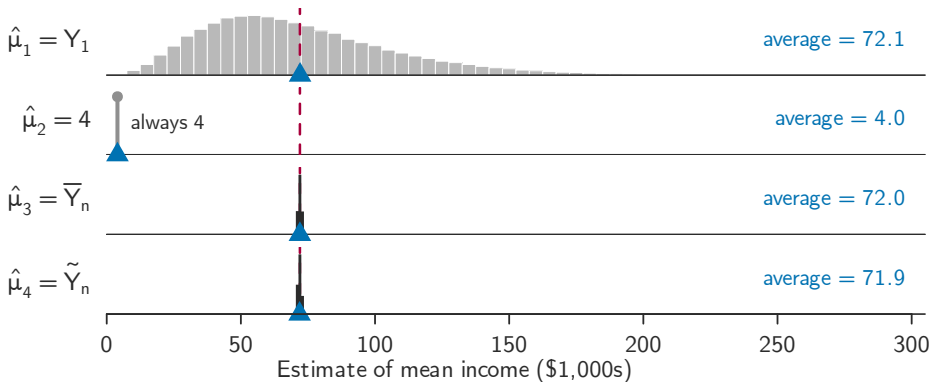


Sampling Distributions of Estimators ($n = 3000$)

100,000 samples of $n = 3,000$

dashed line: true mean

triangle: average of the estimates



Therefore,

$\hat{\mu}_1$: Unbiased but inconsistent

$\hat{\mu}_2$: Biased and inconsistent

$\hat{\mu}_3$: Unbiased and consistent

$\hat{\mu}_4$: Biased but consistent

Weak Law of Large Numbers

Theorem (Weak Law of Large Numbers)

Let $X_1, X_2, \dots, X_n$ be a sequence of i.i.d. random variables, each with finite mean μ . Then for all $\varepsilon > 0$,

$$\bar{X}_n \xrightarrow{P} \mu \text{ as } n \rightarrow \infty$$

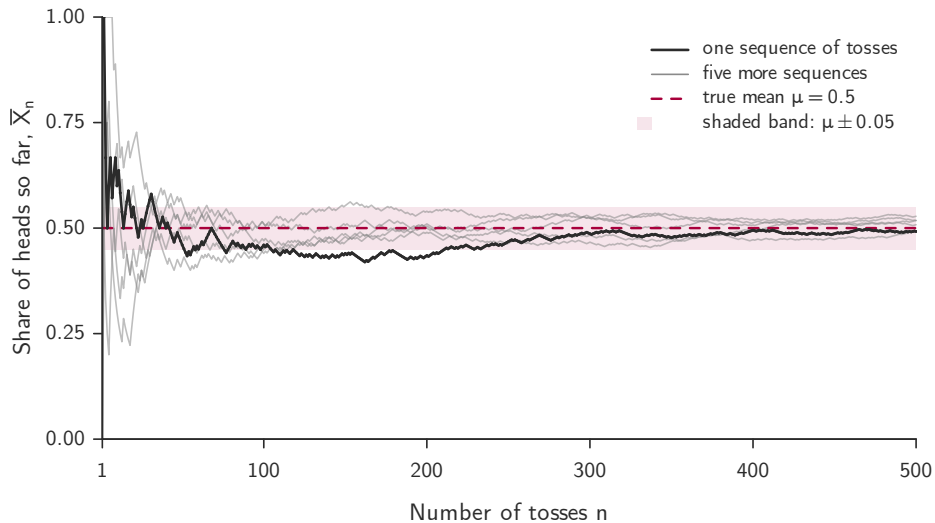
or equivalently,

$$\lim_{n \rightarrow \infty} \Pr(|\bar{X}_n - \mu| \geq \varepsilon) = 0$$

where $\bar{X}_n$ is the sample mean.

- LLN says that the sample average converges in probability to the (common) expected value μ as the sample size grows large.
- In other words, the sample mean is a consistent estimator for any population distribution with a finite mean.

Example: Mean of n independent tosses of a coin



What is Point Estimation?

Finite-sample Properties of Estimators

Asymptotic Properties of Estimators

- Consistency

- Asymptotic Normality

Asymptotic Distributions of Estimators

- In reality, the sample size never equals infinity.
- Therefore, an estimator has a **non-degenerate** (i.e. non-zero variance) sampling distribution in practical situations.
- It would be useful to know something about this distribution when the sample size is large but not infinite.
- How can we study the properties of large, but not infinitely large, samples?
- It turns out that we can use asymptotic theory to **approximate** the sampling distributions of estimators from finite large samples.
- Such approximated distributions are called the **asymptotic distributions** of estimators.

Definition (Convergence in Distribution)

Consider a sequence of random variables $W_1, W_2, \dots$, each with CDFs $F_1, F_2, \dots$. The sequence is said to **converge in distribution** to a limiting random variable W with CDF F if

$$\lim_{n \rightarrow \infty} F_n(x) = F(x),$$

for every point x at which F is continuous. We write this as

$$W_n \xrightarrow{d} W.$$

- As n grows, the distribution of W_n converges to the distribution of W .
- Convergence in probability is a special case of convergence in distribution in which the distribution converges to a **degenerate distribution** (i.e. a probability distribution which only takes a single value).

Limit Distribution of the $\sqrt{n}$ -scaled Sample Mean

- Consider the random variable $\bar{X}_n - \mu$, the fluctuation of the sample mean around the population mean.
- What does this converge to? By LLN, $\bar{X}_n - \mu \xrightarrow{P} \mu - \mu = 0$.
It converges to a degenerate distribution because

$$V[\bar{X}_n - \mu] = \sigma^2/n \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

- Now, consider $\sqrt{n}(\bar{X}_n - \mu)$, the fluctuation **scaled** by the factor of $\sqrt{n}$.
- Does this converge to a degenerate distribution? Its variance stays at σ^2 for every n :

$$V[\sqrt{n}(\bar{X}_n - \mu)] = n \cdot \sigma^2/n = \sigma^2.$$

(we carefully choose the scale $\sqrt{n}$ such that the variance neither shrinks to 0 nor blows up to ∞ .) This suggests, and the CLT on the next slide proves, that the limit is not degenerate.

- Then what non-degenerate distribution does this converge to? It turns out...

Theorem (Lindeberg–Lévy Central Limit Theorem)

Let $X_1, \dots, X_n$ be a sequence of i.i.d. random variables each with mean μ and variance $\sigma^2 < \infty$. Then, for *any* population distribution of X ,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

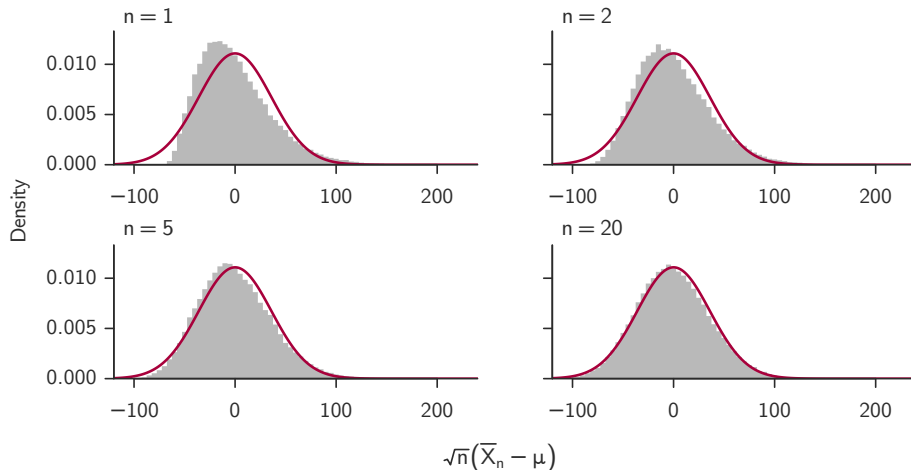
- As n grows, the distribution of the difference between the sample mean and its limit μ , when scaled by $\sqrt{n}$, converges to a normal distribution with mean 0 and variance σ^2 .
- CLT also implies that the *standardized* sample mean converges to a standard normal random variable:

$$Z_n \equiv \frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{V[\bar{X}_n]}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

- Note that CLT holds for a random sample from *any* population distribution with finite variance — what a convenient result!

Central Limit Theorem

bars: 100,000 simulated samples from the income population; curve: the limit $N(0, \sigma^2)$, $\sigma = 36.0$



Asymptotic Normality

- Now let's consider estimators from a large, but not infinitely large, sample.
- CLT implies that, for a sufficiently large sample, we have:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \underset{\text{approx.}}{\sim} \mathcal{N}(0, 1)$$

- This implies:

$$\bar{X}_n \underset{\text{approx.}}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

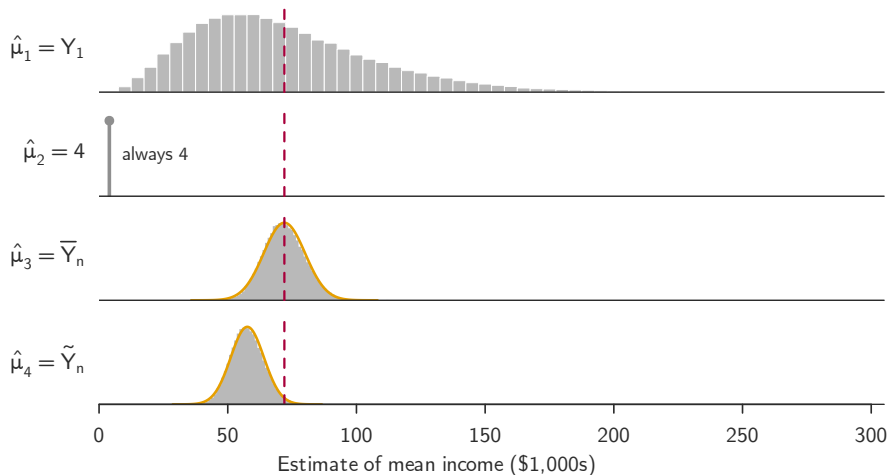
- That is, the (unstandardized) sample mean itself is also approximately normally distributed when sample size is large.
- An estimator is **asymptotically normal** if its asymptotic distribution is normal.

Verifying Asymptotic Normality by Simulation

100,000 samples of $n = 20$

dashed line: true mean

curve: normal approximation



Summary

Concept	Criteria	Intuition
Unbiasedness	$E[\hat{\mu}] = \mu$	Right on average
Efficiency	$V[\hat{\theta}_A] < V[\hat{\theta}_B]$	Lower variance (among unbiased estimators)
Consistency	$\hat{\mu}_n \xrightarrow{P} \mu$	Converge to estimand as $n \rightarrow \infty$
Asymptotic Normality	$\bar{X}_n \overset{\text{approx.}}{\sim} \mathcal{N}(\mu, \frac{\sigma^2}{n})$	Approximately normal in large n (sample mean shown)

Definition

The **analogy principle** tells us to estimate a characteristic in the population using the same characteristic in the sample.

- For example, following the analogy principle, we estimate the population mean μ_Y using the sample mean $\bar{Y}$:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

- To estimate σ_Y^2 , the analogy principle gives the **plug-in** estimator, which divides by n as the definition of σ^2 does:

$$\hat{\sigma}_Y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

(Lecture 2-2: dividing by $n - 1$ instead gives the unbiased **sample variance** S_n^2 .)

- As the sample size increases and the sample “approaches” the population we expect sample characteristic to converge to the population characteristics.