STAT 339: Statistical Theory

Frequentist Parameter Estimation

Anthony Scotina

Reminder!

Personal and General Reflections on 50 years of Teaching Statistics

- Event for *Undergraduate Teaching Award*, Boston Chapter of the ASA

🗓️Tuesday, January 25

⏱️6-7pm ET

📍 Register here

💰 **FREE** to attend

🙏Please go, if you can!

Examples

1. **Clinical Trial**: What is the risk of major adverse cardiovascular events (MACE) for T2D patients while taking certain treatment regimens?
   - Estimating p, the *unknown* proportion of MACE for a large group of T2D patients taking a specific treatment
2. **Piéchart Emporium**: What is the average wait time at the checkout counter for PE customers?
   - Estimating μ, the *unknown* average wait time for PE customers
3. **Cell Phone Batteries**: How can we best quantify battery life in a certain type of smart phone?
   - Estimating μ, the *unknown* average battery life

Considerations: What is the *best* estimator? How do we determine what makes an estimator *best*?

Estimators and Estimates

In general, we will refer to θ as the **target parameter** of interest.

- Can be equal to μ, p, σ^2, etc., but we'll use θ as a "catch-all".

To *estimate* one (or more) parameters, we need **data**!

- For example, suppose the *average* wait time of a *random sample* of 20 PE customers was **five minutes**.
    - This is a **point estimate** - it is an estimate of θ in the form of a *single value*.

A **point estimator** (or *statistic*), $\hat{\theta}$, is the rule/formula used to calculate the value of an *estimate* based on *sample data*.

Examples:

- $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$
- $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2}$

Estimators and Estimates

Unbiased Estimation

Bias of Point Estimators

Estimators are *not* perfect! Some are **good**, some are **bad**, and others are 💩💩💩

Let $\hat{\theta}$ be a point estimator for the parameter θ. Then $\hat{\theta}$ is an **unbiased estimator** if $E(\hat{\theta}) = \theta$. If $E(\hat{\theta}) \neq \theta$, then $\hat{\theta}$ is *biased*.

- The **bias** of a point estimator $\hat{\theta}$ is given by $Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$.

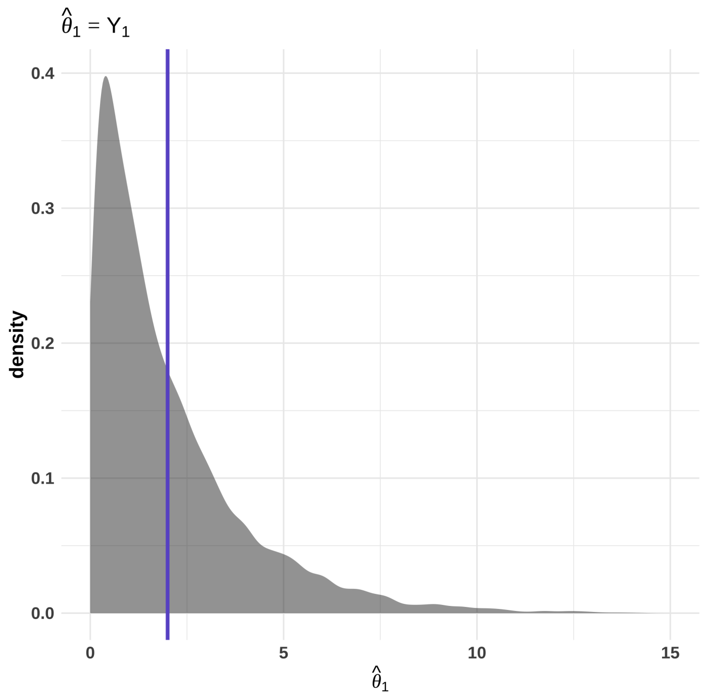
Ideally, the expected value of our estimator $\hat{\theta}$ will equal the parameter (θ) that we're trying to estimate.

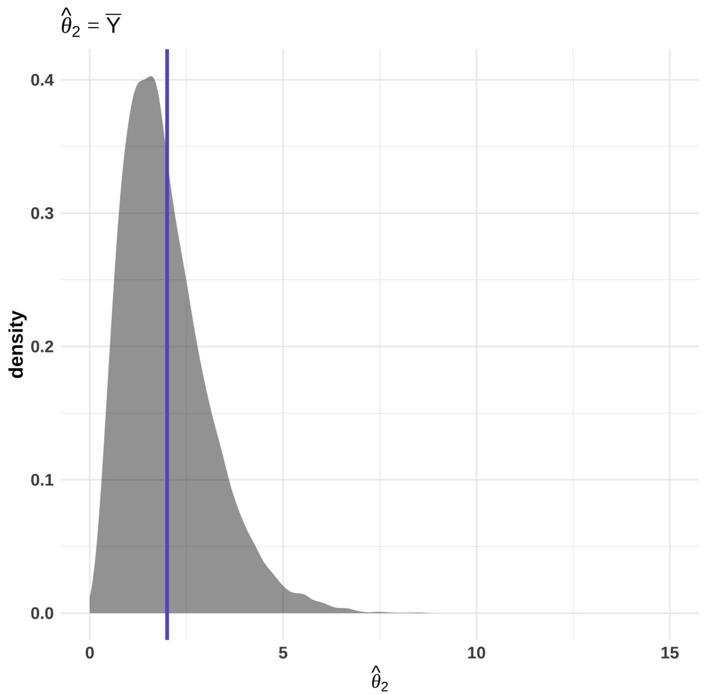
- But we also want $\hat{\theta}$ to have a **small variance** - this means a higher fraction of $\hat{\theta}$ values (in *repeated sampling*) will be *close* to θ.

Two Unbiased Estimators

$$Y_1, Y_2, Y_3 \sim Exponential(2)$$

- Suppose $\theta = E(Y_i) = 2$. let's try to *estimate* θ using different $\hat{\theta}$.





Mean Square Error (MSE)

The **mean square error (MSE)** of a point estimator is the *average of the square of the distance between the estimator and target parameter*:

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

It can be shown that

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2.$$

- In other words, $MSE(\hat{\theta})$ is a function of *both* the **variance** and **bias** of $\hat{\theta}$.

Note: For *unbiased* estimators, $MSE(\hat{\theta}) = Var(\hat{\theta})$.

Biased Estimators

If an estimator $\hat{\theta}$ is **biased** we can usually *correct* it to make it *unbiased*.

Example

Suppose that $\hat{\theta}$ is an estimator for a parameter θ and $E(\hat{\theta}) = a\theta + b$ for some nonzero constants a and b.

1. In terms of a, b, and θ, what is $Bias(\hat{\theta})$?
2. Find a function of $\hat{\theta}$, say, $\hat{\theta}^*$, that is an unbiased estimator for θ.
3. Express $MSE(\hat{\theta}^*)$ as a function of $Var(\hat{\theta})$.

Order Statistics as Estimators

Let $Y_1, Y_2, \ldots, Y_n \sim Uniform(0, \theta)$, where the *target parameter* is θ.

- Because θ is the upper bound of the *support* for the Y_i, let's try to use

  $$Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$$

  as an estimator for θ.

- Is $\hat{\theta} = Y_{(n)}$ *unbiased* for θ?

From STAT 338: The PDF for $Y_{(n)}$ is

$$g_{(n)}(y) = n[F(y)]^{n-1} f(y),$$

where $f(y)$ is the PDF for Y, and $F(y) = P(Y \leq y)$.

Common Unbiased Point Estimators

Sample Mean

Suppose $Y_1, \ldots, Y_n$ are a *random sample* from some population with mean μ and variance σ^2.

- Our *target parameter* is $\theta = \mu$. Let's show that $\hat{\theta} = \bar{Y}$ is **unbiased**.

Sample Variance

It turns out that

$$S^{2*} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

is **biased** for σ^2.

- How can we find an **unbiased** estimator for σ^2? 🤔

Bias/Variance Trade-off

While **unbiased** estimators sound *desirable*, they are not always the *best* estimators.

In general, we'd like for $Bias(\hat{\theta})$ to be close to zero. But we also want $Var(\hat{\theta})$ to be close to zero!

- Higher variance means that estimates might be *very different* across **repeated samples**.
- Ideally, $MSE(\hat{\theta})$ will be *as small as possible*.

Bias/Variance Trade-off

$Y_1, Y_2, \ldots, Y_n \sim iid\, Uniform(0, \theta)$.

Consider three estimators for θ:

1. $\hat{\theta}_1 = 2\bar{Y}$
2. $\hat{\theta}_2 = Y_{(n)}$.
3. $\hat{\theta}_3 = 2Y_1$.

Let's find the bias and variance for each.

Methods of Estimation

The Method of Moments

Finding Estimators

Up to this point, we've mostly used *intuition* to find estimators $\hat{\theta}$ of θ.

- The **sample mean**, $\bar{Y}$, *seems* like it would be a good estimator for the **population mean**, μ.

- The **sample variance**, s^2, *seems* like it would be a good estimator for the **population variance**, σ^2.

But what if we wanted to find estimators for the α and β parameters, using a sample of observations from the $Gamma(\alpha, \beta)$ distribution?

- $E(Y) = \alpha\beta$, but we want to find estimators for *each* of α and β!

Two estimation techniques

1. *Method of Moments*

2. *Method of Maximum Likelihood*

Method of Moments

Recall: The kth moment of a random variable Y is

$$\mu_k^{'} = E(Y^k)$$

- Therefore, $\mu_1^{'} = E(Y)$, $\mu_2^{'} = E(Y^2)$, etc.

We define the kth **sample moment** as the average,

$$m_k^{'} = \frac{1}{n}\sum_{i=1}^{n} Y_i^k.$$

Method of Moments (MOM): Set $\mu_k^{'} = m_k^{'}$, for $k = 1, 2, \ldots, t$ ($t=$ number of parameters to be estimated) and *solve* for the parameter(s) of interest.

Uniform MOM Estimator

Let $Y_1, Y_2, \ldots, Y_n \sim iid\, Uniform(0, \theta)$.

- $\mu_1^{'} = E(Y) = \theta/2$
- $m_1^{'} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

MOM Estimators for Gamma parameters

Let $Y_1, Y_2, \ldots, Y_n \sim iid\ Gamma(\alpha, \beta)$, where α and β are unknown.

- Find the MOM estimators for α and β.
- $\mu_1^{'} = E(Y) = \alpha\beta$.
  - Set this equal to $\frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}$.
- $\mu_2^{'} = E(Y^2) = Var(Y) + [E(Y)]^2 = \alpha\beta^2 + \alpha^2\beta^2$.
  - Set this equal to $\frac{1}{n}\sum_{i=1}^{n} Y_i^2$.

We need to solve the system of equations for α and β.

- $\tilde{\alpha} = \frac{n\bar{Y}^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$
- $\tilde{\beta} = \frac{\bar{Y}}{\tilde{\alpha}} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n\bar{Y}}$

MOM Estimators for Normal parameters

Suppose we have a random sample $Y_1, Y_2, \ldots, Y_n \sim iid\ Normal(\mu, \sigma^2)$.

- Find the MOM estimators for μ and σ^2.
- $\tilde{\mu} = \bar{X} \implies$ **unbiased** for μ!
- $\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \implies$ **biased** for σ^2

Pros and Cons of MOM

Benefits

- Simple to use (just equate sample and population moments)
- Can be used to estimate multiple parameter families

Limitations

- Generate *biased* estimators in many cases
- Need the moments to exist! (Sorry, Cauchy distribution…)
- MLEs are typically *closer* to the target quantity…

Methods of Estimation

The Method of Maximum Likelihood

Likelihood Function

Setting: $Y_1, Y_2, \ldots, Y_n$ are *iid* from a distribution with parameter θ (which might be a single value or a *vector* of multiple parameters).

- The **likelihood function**, $f(\mathbf{y} \mid \theta)$, gives the *likelihood* of observing our sample

  $$(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$$

  when the parameter is θ.
  - For simplicity, we define $\mathbf{y} = (y_1, y_2, \ldots, y_n)$.

Example (*Continuous* random sample)

$$\begin{aligned} f(\mathbf{y} \mid \theta) &= f(y_1, y_2, \ldots, y_n \mid \theta) \\ &= f(y_1 \mid \theta) \times f(y_2 \mid \theta) \times \cdots \times f(y_n \mid \theta) \end{aligned}$$

Note: The likelihood function is sometimes written as $L(\mathbf{y} \mid \theta)$ or $L(\theta)$.

Maximum Likelihood Estimation

Intuition: Choose $\hat{\theta}$ as the estimate of θ that **maximizes** the likelihood function!

- In this context, $\hat{\theta}$ is called the **maximum likelihood estimator (MLE)**.

Example: Moose's favorite toys

This box came with 45 balls. Sadly, Moose lost most of them under furniture, and there are **four left**.

- Some are *red*, and some are *yellow*, but we don't know *exactly* how many of each.
- Moose really only cares about the **red** balls, so let's try to *estimate* how many are red!

Moose's Favorite Toys

We have four balls - some are **red**, and some are yellow. Let's try to estimate *how many red balls there are among the four remaining*.

- I allow Moose to choose three of these balls *at random*. Suppose all three are red; yay!
- If our sample yields *three red balls*, what would be a good estimate of the total number of red balls remaining, n_r?

The parameter, n_r can be *either* **3** or **4**. We know that Moose choose $Y = 3$ red balls, so n_r *cannot* equal 0, 1, or 2.

- Let's find the *likelihood* of obtaining our sample, in two separate worlds: one with $n_r = 3$, and one with $n_r = 4$

$$P(Y = 3 \mid n_r = 3) = \frac{\binom{3}{3}\binom{1}{0}}{\binom{4}{3}} = 0.25$$

Moose's Favorite Toys

We have four balls - some are **red**, and some are yellow. Let's try to estimate *how many red balls there are among the four remaining.*

- I allow Moose to choose three of these balls *at random*. Suppose all three are red; yay!
- If our sample yields *three red balls*, what would be a good estimate of the total number of red balls remaining, n_r?

The parameter, n_r can be *either* **3** or **4**. We know that Moose choose $Y = 3$ red balls, so n_r *cannot* equal 0, 1, or 2.

- Let's find the *likelihood* of obtaining our sample, in two separate worlds: one with $n_r = 3$, and one with $n_r = 4$

$$P(Y = 3 \mid n_r = 4) = \frac{\binom{4}{3}}{\binom{4}{3}} = 1$$

Because $n_r = 4$ *maximizes* the likelihood of the *observed sample*, our **MLE** of n_r is $\hat{n}_r = 4$.

Lifetimes of Electrical Components

Suppose the lifetimes of electrical components (in years), Y, are modeled from an exponential distribution. That is, $Y_1, Y_2, \ldots, Y_n \sim Exponential(\theta)$.

- We observe a sample of $n = 5$ component lifetimes: $\mathbf{y} = (3, 1.5, 2, 1.7, 2.1)$. Let's find the MLE $\hat{\theta}_{MLE}$ for θ that *maximizes* the likelihood of this sample.

1. *Write likelihood*:

$$
\begin{aligned}
L(\theta) = f(\mathbf{y} \mid \theta) &= f(y_1 \mid \theta) \times \cdots \times f(y_5 \mid \theta) \\
&= \left(\frac{1}{\theta}\right) e^{-y_1/\theta} \times \cdots \times \left(\frac{1}{\theta}\right) e^{-y_5/\theta} \\
&= \frac{1}{\theta^5} \exp\left(\frac{-\sum_{i=1}^{5} y_i}{\theta}\right) \\
&= \frac{1}{\theta^5} \exp\left(\frac{-10.3}{\theta}\right)
\end{aligned}
$$

Lifetimes of Electrical Components

Suppose the lifetimes of electrical components (in years), Y, are modeled from an exponential distribution. That is, $Y_1, Y_2, \ldots, Y_n \sim Exponential(\theta)$.

- We observe a sample of $n = 5$ component lifetimes: $\mathbf{y} = (3, 1.5, 2, 1.7, 2.1)$. Let's find the MLE $\hat{\theta}_{MLE}$ for θ that *maximizes* the likelihood of this sample.

2. *Take derivative of* **log-likelihood** with respect to θ:

- $\log L(\theta) = -5 \log \theta - (10.3/\theta)$
- $\frac{d \log L(\theta)}{d\theta} = (-5/\theta) + (10.3/\theta^2)$

Lifetimes of Electrical Components

Suppose the lifetimes of electrical components (in years), Y, are modeled from an exponential distribution. That is, $Y_1, Y_2, \ldots, Y_n \sim Exponential(\theta)$.

- We observe a sample of $n = 5$ component lifetimes: $\mathbf{y} = (3, 1.5, 2, 1.7, 2.1)$. Let's find the MLE $\hat{\theta}_{MLE}$ for θ that *maximizes* the likelihood of this sample.

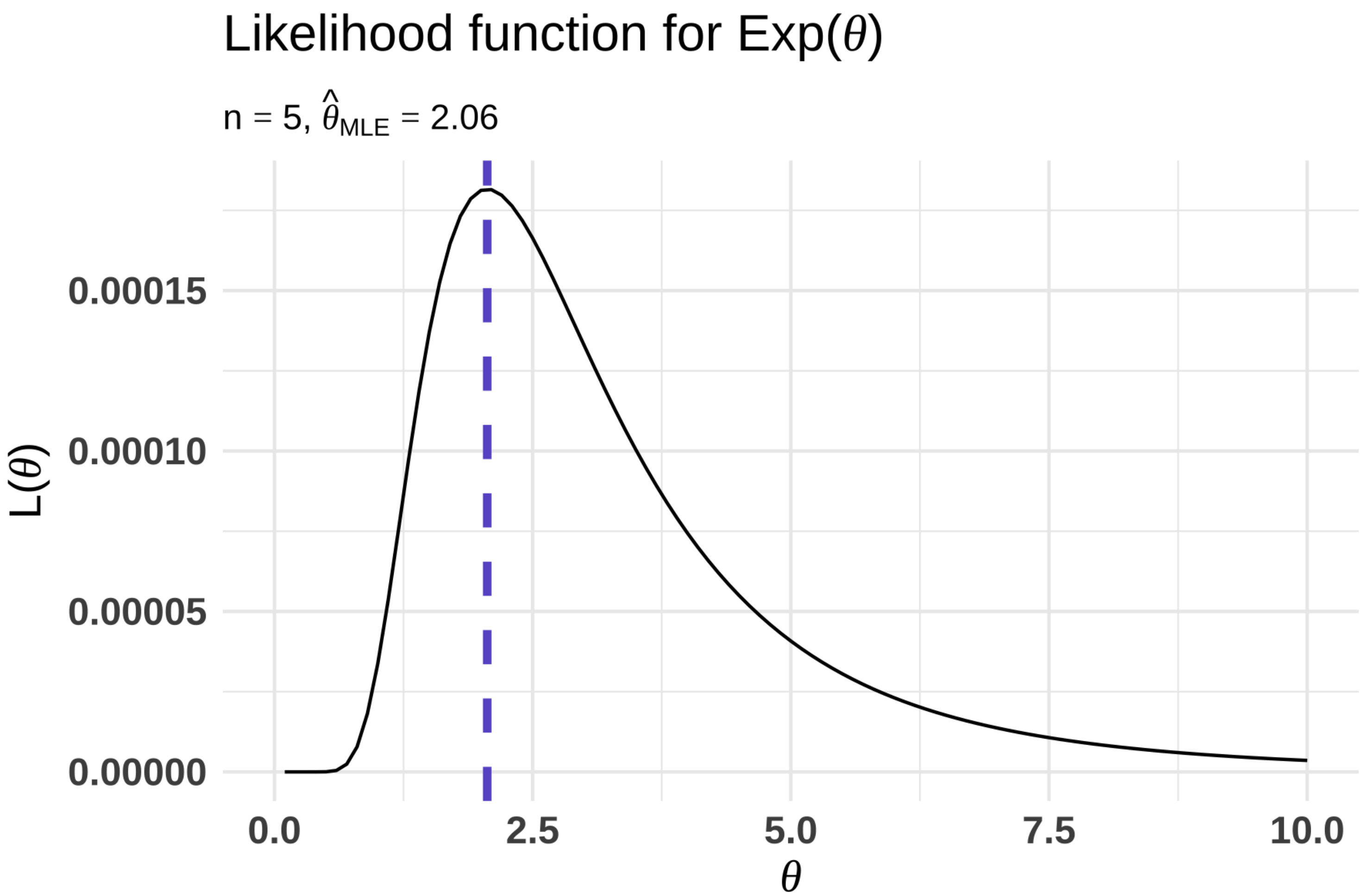
3. *Solve for θ:*

- $(-5/\theta) + (10.3/\theta^2) = 0 \implies \theta = 10.3/5 = 2.06$

Therefore, $\hat{\theta}_{MLE} = 2.06$. Because the *data* $\mathbf{y}$ are *observed* here, 2.06 is a maximum likelihood **estimate** of θ.

4. (Bonus) Take second derivative of log-likelihood, make sure it is negative at $\theta = 2.06$.

Exponential Likelihood



Normal Distribution MLEs

Suppose that $Y_1, Y_2, \ldots, Y_n$ form a *random sample* from a $Normal(\mu, \sigma^2)$ distribution.

- Find the MLEs of μ and σ^2.

Note: $\theta = (\mu, \sigma^2)$, so we need to take two different derivatives of $\log L(\theta)$.

Solution

- $\hat{\mu}_{MLE} = \bar{Y}$
- $\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$ (a **biased** estimator!)

Uniform MLE

Suppose that $Y_1, Y_2, \ldots, Y_n$ form a *random sample* from a $Uniform(0, \theta)$ distribution.

- Find the MLE of θ.

1. *Write likelihood*:

$$
\begin{aligned}
L(\theta) &= f(y_1 \mid \theta) \times \cdots \times f(y_n \mid \theta) \\
&= \frac{1}{\theta^n}, \quad \text{if } 0 \le y_i \le \theta
\end{aligned}
$$

- The first derivative of $L(\theta)$ does *not* equal zero for *any* $\theta > 0$.
- However, $1/\theta^n$ **increases** as θ *decreases*, so we want to select θ to be as small as possible in order to maximize the likelihood.
    - *One constraint*: All of the y_i values are between 0 and θ.
    - The *smallest value* of θ that satisfies this constraint is $Y_{(n)} = \max(Y_1, \ldots, Y_n)$

Therefore, $\hat{\theta}_{MLE} = Y_{(n)}$.

Pros and Cons of MLE

Benefits

- MLEs are *invariant*! This means that, if $\hat{\theta}$ is an MLE for θ, then $g(\hat{\theta})$ is an MLE for $g(\theta)$.
- MLEs are *consistent*.

Limitations

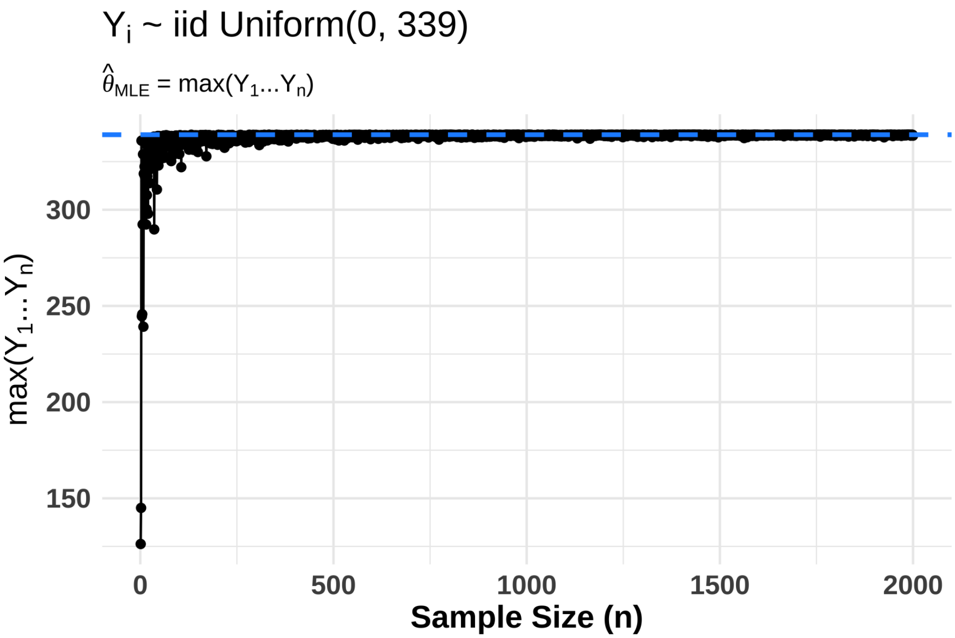
- MLEs do not always exist.
- The MLE is **NOT** the most likely parameter, *given the data* ($E(\theta\mid Y)$). It estimates the parameter θ that maximizes the distribution of $Y \mid \theta$
    - In other words, the MLE gives the parameter estimate most likely to have produced the observed data.

Consistency of the MLE

The estimator $\hat{\theta}_n$ is said to be **consistent** for θ if, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) = 1.$$

- Basically, this means that if n is **large enough**, there is a probability of 1 that $\hat{\theta}_n$ will be *very close to* θ.



Data Reduction: Sufficiency

Sufficient Statistics

Most of the estimators we've chosen have *seemed* like they would be good estimators.

- The sample mean, $\bar{Y}$, is *probably* a solid estimator for the population mean μ.

Once we calculate $\bar{Y}$, the actual sample values $Y_1, \ldots, Y_n$ are *no longer important*; the information in the sample is *summarized* by $\bar{Y}$.

- But does this summary retain all of the information about μ contained in the original n observations?

A statistic that summarizes *all* information in a sample about a target parameter is said to be **sufficient**.

- We'll use sufficient statistics to help determine *best* (unbiased) estimators.

Factorization Criterion

Theorem

Let U be a statistic based on the random sample $Y_1, \ldots, Y_n$. Then U is a **sufficient statistic** for the estimation of a parameter θ if and only if the likelihood $L(\theta) = f(y_1, \ldots, y_n \mid \theta)$ can be factored into two nonnegative functions:

$$L(\theta) = g(u, \theta) \times h(y_1, \ldots, y_n)$$

where:

- $g(u, \theta)$ is a function *only* of u and θ, and
- $h(y_1, \ldots, y_n)$ is *not* a function of θ.

Process for Finding a Sufficient Statistic

1. Write out the **likelihood**, $L(\theta) = f(y_1 \mid \theta) \times \cdots \times f(y_n \mid \theta)$.
2. Given some statistic U, check if $L(\theta)$ can be broken down into $g(u, \theta)$ and $h(y_1, \ldots, y_n)$.
    - **Note**: There are often *more than one* sufficient statistic for any parameter.

Sufficient Statistic Examples

1. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample such that $Y_i \sim Exponential(\theta)$ with PDF

$$f(y_i \mid \theta) = \frac{1}{\theta} e^{-y_i/\theta}, \quad y_i > 0.$$

   Show that $U = \bar{Y}$ is a sufficient statistic for θ.

2. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample such that $Y_i \sim Beta(\theta, 1)$ with PDF

$$f(y_i \mid \theta) = \theta y^{\theta - 1}, \quad 0 < y < 1.$$

   Show that $U = \prod_{i=1}^{n} Y_i = Y_1 \times \cdots \times Y_n$ is a sufficient statistic for θ.

Rao-Blackwell Theorem

Let $\hat{\theta}$ be an unbiased estimator for θ If $\hat{\theta}$ has a smaller variance than *all other unbiased estimators* for θ, then $\hat{\theta}$ is the **best unbiased estimator (BUE)** (the "boo").

Rao-Blackwell Theorem

Let $h(U)$ be some function of a statistic, U. If:

- U is a sufficient statistic for θ
- $E[h(U)] = \theta$

then it follows that $\hat{\theta} = h(U)$ is the *best unbiased estimator* for θ.

Sampling Distributions of Estimators

Recap

What have we done so far?⌛⌛⌛

We've used **point estimators** (or *statistics*), $\hat{\theta}$, to *estimate* unknown target **parameters**, θ.

- These estimators are functions of:
  - observable random variables in a sample
  - known constants (usually the sample size, *n*)
- While *unknown*, θ is assumed to be **fixed** at some value.

Because statistics are *functions* of random variables…

All statistics are random variables!

Because **all statistics are random variables**, all statistics have *probability distributions* that illustrate (among other things) how much they *vary from sample to sample.*

- These "special" probability distributions are called **sampling distributions**.

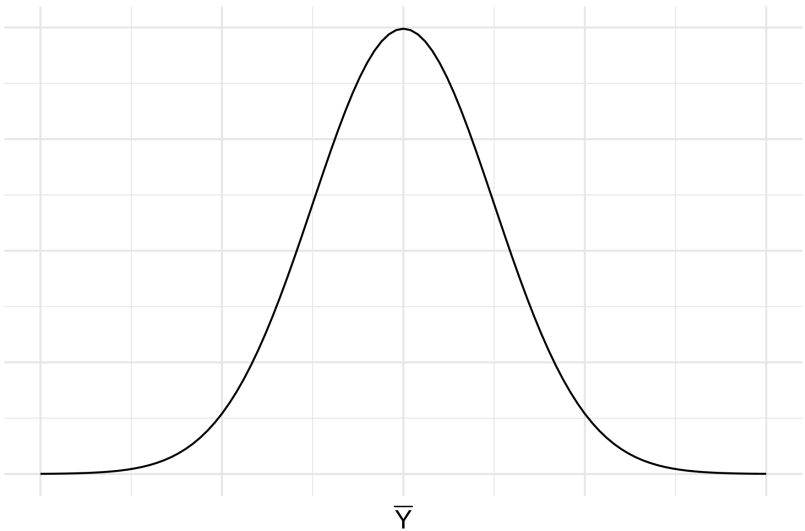
Why sampling distributions?

Before the sample has been taken, we can use the *sampling distribution* of $\hat{\theta}$ to calculate the probability that $\hat{\theta}$ will be close to θ.

Example

Let $Y_1, Y_2, \ldots, Y_n \sim iid\ Normal(\mu, \sigma^2)$. Then:

$$\bar{Y} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right).$$



Chi-Squared Distribution

Again, suppose that $Y_1, Y_2, \ldots, Y_n \sim iid\ Normal(\mu, \sigma^2)$.

- Though *now* we want to work with the *sample variance*, S^2.

1. **Unbiased Estimator**:

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

2. **MLE (and MOM Estimator)**:

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

It turns out that, except for a *scale factor*, the sample variance follows a χ^2 (*chi-squared*) distribution with $n-1$ **degrees of freedom**.

Chi-Squared Distribution

Theorem

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a $Normal(\mu, \sigma^2)$ distribution. Then

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

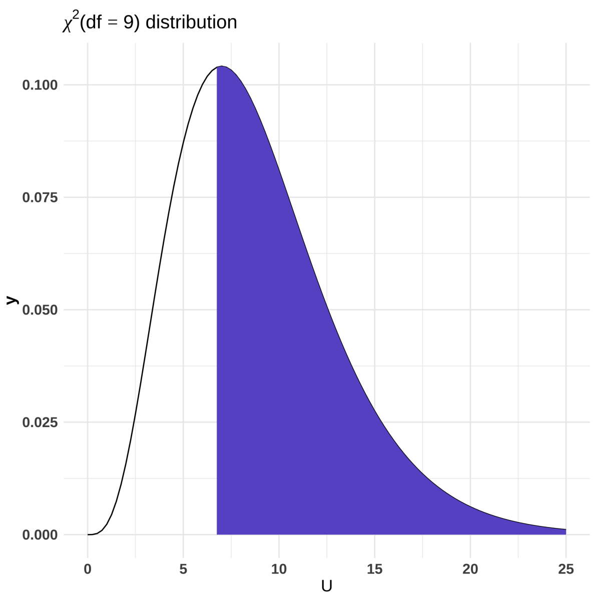
Example

Suppose $Y_1, Y_2, \ldots, Y_{10} \sim iid\ Normal(\mu, \sigma^2 = 4)$.

- μ is *unknown*, and σ^2 is *known*.

Because $n = 10$ and $\sigma^2 = 4$, the *sampling distribution* $U = 9S^2/4 \sim \chi^2(df = 9)$.

- Let's use this to find $P(S^2 > 3)$.



$$P(S^2 > 3) = P\left(\frac{9S^2}{4} > \frac{9 \times 3}{4}\right) = P(U > 6.75)$$

```
1 - pchisq(6.75, df = 9)
```

```
## [1] 0.6631296
```

Student's t Distribution

When the population standard deviation, σ, is *unknown*, it can be estimated by $S = \sqrt{S^2}$, and the quantity

$$T = \frac{\bar{Y} - \mu}{s/\sqrt{n}}$$

is used in certain procedures for inference about μ.

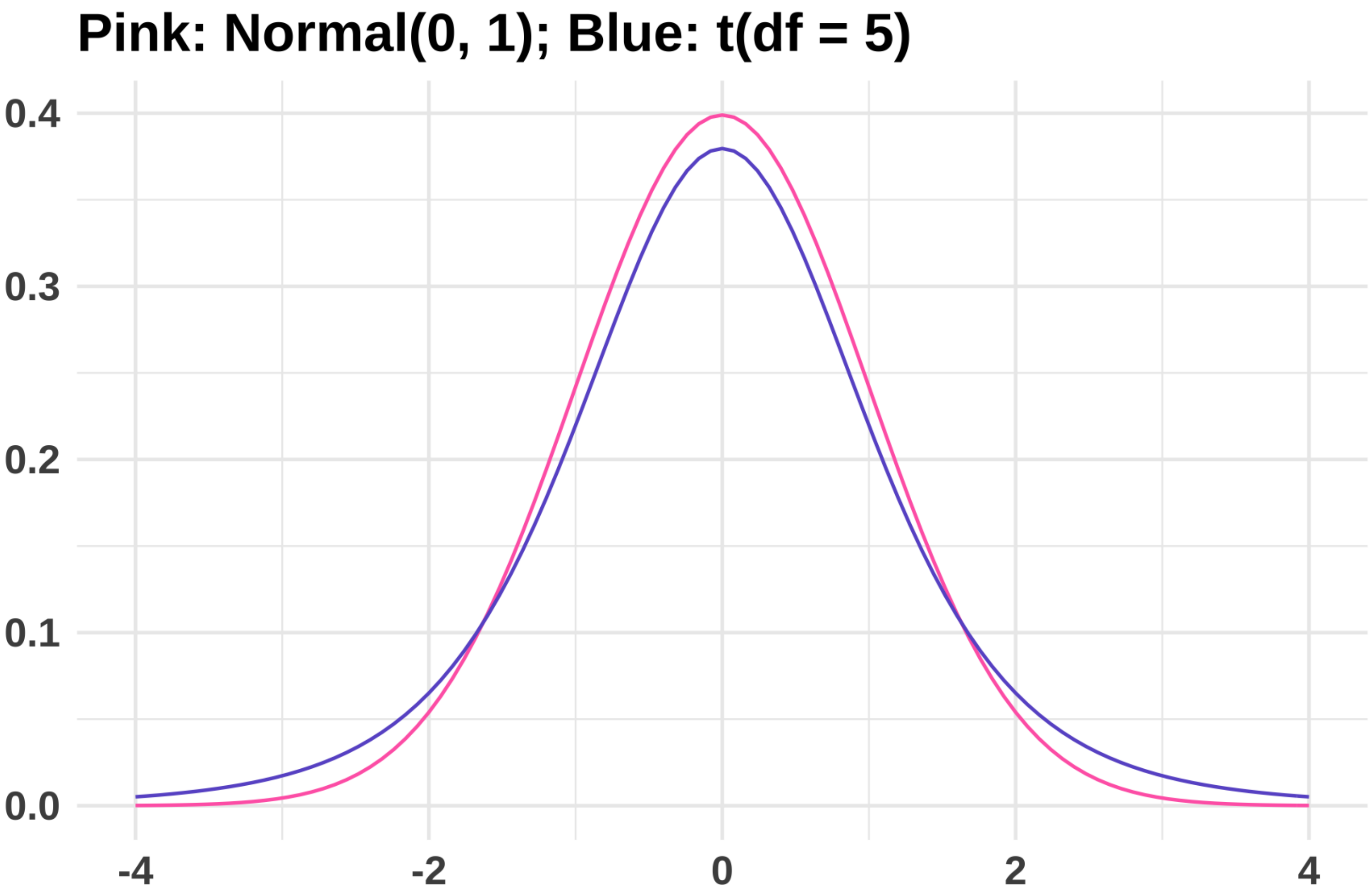
- This quantity, T, has a **t distribution** with $n - 1$ degrees of freedom!

Definition: Let Z be a **standard Normal** random variable, and let W be a χ^2-distributed random variable with ν degrees of freedom. Then, if Z and W are *independent*,

$$T = \frac{Z}{\sqrt{W/\nu}}$$

is said to have a **t distribution** with ν degrees of freedom.

Student's t vs. Normal



F Distribution

Suppose now that we are comparing the variances from **two normal samples**:

- $X_1, X_2, \ldots, X_n \sim N(\mu_X, \sigma_X^2)$
- $Y_1, Y_2, \ldots, Y_n \sim N(\mu_Y, \sigma_Y^2)$

Question

Are the sample data consistent with the *assumption* that $\sigma_X^2 = \sigma_Y^2$?

- We know that S_X^2 and S_Y^2 are unbiased estimators of σ_X^2 and σ_Y^2, respectively.
  - Let's look at the *ratio*, S_X^2 / S_Y^2.
- It turns out that, if we divide each S^2 by its respective σ^2, then the ratio

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(df_1 = n_X - 1, df_2 = n_Y - 1)$$

F Distribution

General Definition

Let W_1 and W_2 be *independent* χ^2-distributed random variables with ν_1 and ν_2 df, respectively. Then

$$F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F(\nu_1, \nu_2).$$

Example: Suppose...

- $X_1, \ldots, X_{21} \sim N(\mu_X, \sigma_X^2)$, $S_X^2 = 994.7$
- $Y_1, \ldots, Y_{15} \sim N(\mu_Y, \sigma_Y^2)$, $S_Y^2 = 250.3$

Is it reasonable to assume that $\sigma_X^2 = \sigma_Y^2$?

IF $\sigma_X^2 = \sigma_Y^2$, then $\frac{S_X^2}{S_Y^2} \sim F(20, 14)$:

$$P(\frac{S_X^2}{S_Y^2} > \frac{994.7}{250.3}) = P(F > 3.97) = 0.006$$

Recap

We have developed **sampling distributions** of *statistics* calculated by using observations in random samples from **Normal** populations.

If $Y_1, \ldots, Y_n \sim iid\ N(\mu, \sigma^2)$, then...

1. $\sqrt{n}(\bar{Y} - \mu)/\sigma \sim N(0, 1)$
2. $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$
3. $\sqrt{n}(\bar{Y} - \mu)/S \sim t(n-1)$
4. $F = (S_1^2/\sigma_1^2)/(S_2^2/\sigma_2^2) \sim F(n_1 - 1, n_2 - 1)$, provided that the samples are *independent.*

These sampling distributions will help us quite a bit later on with **confidence intervals** and **hypothesis tests**!

Frequentist Estimation

This first unit of STAT 339 has been devoted to estimation from a **frequentist perspective**.

- Frequentists view probability as a representation of a *long-run frequency* over a *large* (sometimes infinite) number of repetitions of an experiment.
- The true value of a parameter, θ, is **fixed** and **unknown**.

In the next unit, we will focus on estimation from a **Bayesian perspective**.

- Bayesians view probability as a representation of a *relative plausibility* of an event.
- Parameters, θ, are themselves treated as *random variables*, assigned some **prior distribution**.
    - Gives weight to prior knowledge.

While we will study various procedures through both *frequentist* and *Bayesian* lenses, these are **not** competing!

- Both perspectives aim to learn from data, both use data to fit models, evaluate hypotheses, etc.