# STA261 - Module 4

## Intervals and Model Checking

Rob Zimmerman

University of Toronto

July 27-29, 2021

## Uncertainty in Point Estimates

- In Module 2, we learned how to produce the "best" point estimates of $\theta$ possible using statistics of our data

- The "best" unbiased estimator $\hat{\theta}(\mathbf{X})$ is the one that has the lowest possible variance among all unbiased estimators of $\theta$ $\hat{\theta}(\vec{X})$

- But even so, suppose we observe $\mathbf{X} = \mathbf{x}$ and calculate $\hat{\theta}(\mathbf{x})$; how do we know this is close to the true $\theta$? It may not be!

- We can't know for sure

- But we can use the data to get a range of *plausible* values of $\theta$

Eg: still assuming STA261 heights $\sim N(\mu, 1)$. I calculate $\bar{X}_n = 5'6.5''$

Probably more plausible that $\mu \in [4'6.5'', 6'6.5'']$ than $(2', 4']$

$= [\bar{X}_n - 1, \bar{X}_n + 1]$

# Random Sets

- Suppose for now that $\Theta \subseteq \mathbb{R}$
- If $\hat{\theta}(\mathbf{X})$ is a continuous random variable, then $\mathbb{P}_\theta\left(\theta = \hat{\theta}(\mathbf{X})\right) = 0$
  useless...
- But we can try to find a random set $C(\mathbf{X}) \subseteq \mathbb{R}$ based on $\mathbf{X}$ such that $\mathbb{P}_\theta\left(\theta \in C(\mathbf{X})\right) = 0.95$, for example
  a set which is a function of the random sample $\vec{X}$
  eg, $[\bar{X}_n - 1, \bar{X}_n + 1]$
- Example 4.1: Let $X \sim \mathcal{N}(\mu, 1)$ where $\mu \in \mathbb{R}$. Show that the region $C(X) = (X + z_{0.025}, X + z_{0.975})$ satisfies $\mathbb{P}_\mu(\mu \in C(X)) = 0.95$.

$$
\begin{aligned}
&\mathbb{P}_\mu(\mu \in C(X)) \\
&= \mathbb{P}_\mu(X + z_{0.025} < \mu < X + z_{0.975}) \\
&= \mathbb{P}_\mu(z_{0.025} < \mu - X < z_{0.975}) \\
&= \mathbb{P}(z_{0.025} < Z < z_{0.975}) \quad \text{where } Z \sim N(0,1) \text{ (check!)} \\
&= \Phi(z_{0.975}) - \Phi(z_{0.025}) \\
&= 0.95
\end{aligned}
$$

# Interval Estimators and Confidence Intervals

- Definition 4.1: An **interval estimate** of a parameter $\theta \in \Theta \subseteq \mathbb{R}$ is any pair of statistics $L, U : \mathcal{X}^n \to \mathbb{R}$ such that $L(\mathbf{x}) \leq U(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}^n$. The random interval $\underbrace{(L(\mathbf{X}), U(\mathbf{X}))}_{\text{interval with random endpoints!}}$ is called an **interval estimator**.

  Note: some authors include the endpoints!

- Example 4.2: $N(\mu,1)$: $(X_{(1)}, X_{(n)}+5)$   Bernoulli$(p)$: $(-\bar{X}_n+4, \bar{X}_n+5)$

- Definition 4.2: Suppose $\alpha \in [0,1]$. An interval estimator $(L(\mathbf{X}), U(\mathbf{X}))$ is a $(\mathbf{1}-\boldsymbol{\alpha})$**-confidence interval** for $\theta$ if $\mathbb{P}_\theta(L(\mathbf{X}) < \theta < U(\mathbf{X})) \geq 1-\alpha$ for all $\theta \in \Theta$. We refer to $1-\alpha$ as the **confidence level** of the interval.

- Example 4.3: $X \sim N(\mu,1) \Rightarrow$ We showed in Example 4.1 that

  $(X + z_{0.025}, X + z_{0.975})$

  is a $0.95$-confidence interval.

# One-Sided Intervals

- Definition 4.3: A **lower one-sided** confidence interval is a confidence interval of the form $(L(\mathbf{X}), \infty)$. An **upper one-sided** confidence interval is a confidence interval of the form $(-\infty, U(\mathbf{X}))$.

- Example 4.4: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1)$. Find a lower one-sided 0.5-confidence interval for $\mu$.

$0.5 = \mathbb{P}(Z < 0)$ where $Z \sim N(0,1)$

$= \mathbb{P}_\mu\left( \frac{\bar{X}_n - \mu}{\sqrt{1/n}} < 0 \right)$

$= \mathbb{P}_\mu( \bar{X}_n < \mu )$

$= \mathbb{P}_\mu( \mu \in (\bar{X}_n, \infty))$

So $(\bar{X}_n, \infty)$ is a lower one-sided 0.5-CI for $\mu$

But $(X_i, \infty)$ is ~~a~~ another one!

So $(1-\alpha)$-CIs are not unique.

# Confidence Intervals: Warmups

- The reason for the "$\geq 1-\alpha$" in the definition is that $\mathbb{P}_\theta\left(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})\right)$ may not be free of $\theta$, depending on the choices of $L(\mathbf{X})$ and $U(\mathbf{X})$

- The lower bound means we want $1-\alpha$ confidence even in the "worst case"; equivalently,

$$\inf_{\theta \in \Theta} \mathbb{P}_\theta\left(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})\right) \geq 1-\alpha$$

- Example 4.5: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Unif$(0,\theta)$, where $\theta \overset{>0}{\cancel{\in}} \mathbb{R}$. Find $a \in \mathbb{R}$ such that $(aX_{(n)}, 2aX_{(n)})$ is a 95% confidence interval for $\theta$.

u $F_\theta(x) = \left(\frac{x}{\theta}\right)$

$$\begin{aligned}
1-\alpha &= \mathbb{P}_\theta\left(\theta \in (aX_{(n)}, 2aX_{(n)})\right) \\
&= \mathbb{P}_\theta\left(aX_{(n)} < \theta < 2aX_{(n)}\right) \\
&= \mathbb{P}\left(\frac{\theta}{2a} < X_{(n)} < \frac{\theta}{a}\right) \\
&= F_{X_{(n)}}(\theta/a) - F_{X_{(n)}}(\theta/2a) \\
&= \left(\frac{\theta/a}{\theta}\right)^n - \left(\frac{\theta/2a}{\theta}\right)^n = \left(\frac{1}{a}\right)^n - \left(\frac{1}{2a}\right)^n
\end{aligned}$$

$\Rightarrow$ choose $a = \left(\frac{1-2^{-n}}{1-\alpha}\right)^{1/n}$, $\alpha = 0.05$

# Poll Time!

$\theta \in \mathbb{R} \Rightarrow \quad 3 < \theta < 5$ is either T or F

$\Rightarrow \mathbb{P}(3 < \theta < 5) \in \{0, 1\}$

If it's 0, then it's $< 0.95$

If it's 1, then it's $\geq 0.95$.

We don't know!

# Some Confidence Intervals Are Better Than Others

- A confidence interval is only useful when it tells us something we didn't know before collecting the data

- Example 4.6: Suppose $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, where $\theta \in (0,1)$. Find a 100% confidence interval for $\theta$.

  $(0,1)$. Not helpful!

  $(X_1 - 1, X_2 + 1)$ Not helpful!

  $(X_i - 200, \infty)$ Not helpful!

  A 100% confidence interval doesn't tell us anything!

- A good confidence interval shouldn't be any longer than necessary

- We interpret the length of the interval as a measure of how accurately the data allow us to know the true value of $\theta$

# Bringing Back Hypothesis Tests

- In Module 3, we learned about test statistics and rejection regions for hypothesis tests

- Pick some arbitrary $\theta_0 \in \Theta$, and suppose we want a level-$\alpha$ test of $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$ using a test statistic $T(\mathbf{X})$

- This means finding a rejection region $R_{\theta_0}$ such that

$$\mathbb{P}_{\theta_0}(T(\mathbf{X}) \in R_{\theta_0}) \leq \alpha$$

- This is equivalent to finding an *acceptance region* $A_{\theta_0} = R_{\theta_0}^c$ such that

$$\mathbb{P}_{\theta_0}(T(\mathbf{X}) \in A_{\theta_0}) \geq 1 - \alpha$$

# Confidence Intervals Via Test Statistics

- If the statement $T(\mathbf{X}) \in A_{\theta_0}$ can be manipulated into an equivalent statement of the form $L(\mathbf{X}) < \theta_0 < U(\mathbf{X})$, then

$$\mathbb{P}_{\theta_0}(L(\mathbf{X}) < \theta_0 < U(\mathbf{X})) \geq 1 - \alpha$$

- But $\theta_0 \in \Theta$ was arbitrary!

- So if we did this right, we must have

$$\mathbb{P}_{\theta}\left(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})\right) \geq 1 - \alpha \quad \text{for all } \theta \in \Theta$$

- This method of finding confidence intervals is called *inverting a hypothesis test*

- You can also go the other direction: start with a $(1-\alpha)$-CI $(L(\vec{X}), U(\vec{X}))$, and use it to design a level-$\alpha$ test (under the right conditions)

# Famous Examples: $Z$-Intervals

- Example 4.7: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ and $\sigma^2$ is known. Find a $(1-\alpha)$-confidence interval for $\mu$ by inverting the two-sided $Z$-test. Let $\mu_0 \in \mathbb{R}$. Need a level-$\alpha$ test of $H_0: \mu = \mu_0$ vs $H_A: \mu \neq \mu_0$.

From Example 3.15, $R_{\mu_0} = \left\{ \vec{x} \in \mathcal{X}^n : \left| \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2/n}} \right| > z_{1-\alpha/2} \right\}$

$$\Rightarrow A_{\mu_0} = \left\{ \vec{x} \in \mathcal{X}^n : \left| \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2/n}} \right| \leq z_{1-\alpha/2} \right\}$$

Then
$$\begin{aligned}
1-\alpha &= \mathbb{P}_\mu(\vec{X} \in A_\mu) \\
&= \mathbb{P}_\mu\left( -z_{1-\alpha/2} \leq \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leq z_{1-\alpha/2} \right) \\
&= \mathbb{P}_\mu\left( -z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}} - \bar{X}_n \leq -\mu \leq z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}} - \bar{X}_n \right) \\
&= \mathbb{P}_\mu\left( \bar{X}_n - z_{1-\alpha/2} \sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X}_n + z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}} \right)
\end{aligned}$$

$$\Rightarrow \text{Our } (1-\alpha)\text{-CI is } \left( \bar{X}_n - z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}},\ \bar{X}_n + z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}} \right)$$

"$Z$-interval"

# Famous Examples: One-Sided $Z$-Intervals

- Example 4.8: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ and $\sigma^2$ is known. Find a lower one-sided $(1-\alpha)$-confidence interval for $\mu$ by inverting an appropriate one-sided $Z$-test.

Example 3.16 : $R_{\mu_0} = \left\{ \vec{x} \in \mathcal{X}^n : \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} > z_{1-\alpha} \right\}$ gave us a size-$\alpha$ test

$$\Rightarrow A_{\mu_0} = \left\{ \vec{x} \in \mathcal{X}^n : \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} \leq z_{1-\alpha} \right\}$$

$$1 - \alpha = \mathbb{P}_\mu\left( \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \leq z_{1-\alpha} \right)$$

$$= \mathbb{P}_\mu\left( -\mu \leq z_{1-\alpha} \cdot \sqrt{\frac{\sigma^2}{n}} - \bar{X}_n \right)$$

$$= \mathbb{P}_\mu\left( \mu \geq \bar{X}_n - z_{1-\alpha} \cdot \sqrt{\frac{\sigma^2}{n}} \right)$$

$$\Rightarrow \text{Choose } \left( \bar{X}_n - z_{1-\alpha} \cdot \sqrt{\frac{\sigma^2}{n}}, \infty \right)$$

# Famous Examples: $t$-Intervals

- Example 4.9: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Find a $(1-\alpha)$-confidence interval for $\mu$ by inverting the two-sided $t$-test.

Example 3.17: $R_{\mu_0} = \left\{ \vec{x} \in \mathcal{X}^n : \left| \frac{\bar{x} - \mu_0}{\sqrt{s^2/n}} \right| > t_{n-1,1-\alpha/2} \right\}$

So $1-\alpha = \mathbb{P}_\mu\left( -t_{n-1,1-\alpha/2} \le \frac{\bar{X}_n - \mu}{\sqrt{S^2/n}} \le t_{n-1,1-\alpha/2} \right)$

$= \mathbb{P}_\mu\left( \bar{X}_n - t_{n-1,1-\alpha/2} \cdot \sqrt{S^2/n} \le \mu \le \bar{X}_n + t_{n-1,1-\alpha/2} \cdot \sqrt{S^2/n} \right)$

So choose $\left( \bar{X}_n - t_{n-1,1-\alpha/2} \cdot \sqrt{\frac{S^2}{n}} , \bar{X}_n + t_{n-1,1-\alpha/2} \cdot \sqrt{\frac{S^2}{n}} \right)$

"$t$ - Interval"

# Famous Examples: One-Sided $t$-Intervals

- Example 4.10: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Find an upper one-sided $(1-\alpha)$-confidence interval for $\mu$ by inverting an appropriate one-sided $t$-test.

Consider $H_0: \mu \geq \mu_0$ vs $H_A: \mu < \mu_0$. By Example 3.18,

$R_{\mu_0} = \{ \vec{x} \in \mathcal{X}^n : \frac{\bar{x} - \mu_0}{\sqrt{s^2/n}} < t_{n-1,\alpha} \}$ give us a size-$\alpha$ test.

$$\Rightarrow 1-\alpha = \mathbb{P}_\mu\left( \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \geq t_{n-1,\alpha} \right)$$

$$= \mathbb{P}_\mu\left( \mu \leq \bar{X}_n - t_{n-1,\alpha} \cdot \sqrt{\frac{S^2}{n}} \right)$$

So choose $\left( -\infty, \bar{X}_n - t_{n-1,\alpha} \sqrt{S^2/n} \right)$

# An LRT-Based Interval

$F_\theta(x) = 1 - e^{-(x-\theta)}$

- Example 4.11: Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f_\theta(x) = e^{-(x-\theta)} \cdot \mathbb{1}_{x \geq \theta}$, where $\theta \in \mathbb{R}$. Find a $(1-\alpha)$-confidence interval for $\theta$ by inverting an LRT.

From Example 3.21, the LRT of $H_0: \theta \leq \theta_0$ vs $H_A: \theta > \theta_0$ had rejection region

$$R_{\theta_0} = \{\vec{x} \in \mathcal{X}^n : x_{(1)} \geq \theta_0 - \frac{\log(c)}{n} \text{ OR } x_{(1)} < \theta_0\}.$$

$$\Rightarrow A_{\theta_0} = \{\vec{x} \in \mathcal{X}^n : x_{(1)} + \frac{\log(c)}{n} < \theta_0 \text{ AND } x_{(1)} \geq \theta_0\} = \{\vec{x} \in \mathcal{X}^n : x_{(1)} + \frac{\log(c)}{n} < \theta_0 < x_{(1)}\}$$

If we can choose $c$ to make a size-$\alpha$ test, then $\left(X_{(1)} + \frac{\log(c)}{n}, X_{(1)}\right)$ will be a $(1-\alpha)$-CI for $\theta$.

How? $$1-\alpha = P_\theta\Big(X_{(1)} \leq \theta - \frac{\log(c)}{n} \wedge \underbrace{X_{(1)} \geq \theta_0}_{\text{always true!}}\Big)$$
$$= 1 - \left(1 - F_\theta\left(\theta - \frac{\log(c)}{n}\right)\right)^n$$
$$= 1 - \left(1 - 1 + \exp\left(-\left[\theta - \frac{\log(c)}{n} - \theta\right]\right)\right)^n$$
$$= 1 - c$$

$\Rightarrow$ choose $c = \alpha$. So $\left(X_{(1)} + \frac{\log(\alpha)}{n}, X_{(1)}\right)$ is a $(1-\alpha)$-CI for $\theta$.

# Functions of the Data *and* the Parameter

- In constructing our confidence intervals, we've often encountered statements that look like
$$\mathbb{P}_\theta\left(a < Q(\mathbf{X}, \theta) < b\right) \geq 1 - \alpha,$$
where $Q : \mathcal{X}^n \times \Theta \to \mathbb{R}$ is a function of the data $\mathbf{X}$ *and* the parameter $\theta$, and $a, b$ are constants

- We were able to choose those constants $a$ and $b$ because we knew exactly what the distribution of $Q(\mathbf{X}, \theta)$ was

- We could then "invert" the statement $a < Q(\mathbf{X}, \theta) < b$ to produce a confidence interval for $\theta$

- Example 4.12: $N(\mu, \sigma^2)$, $\sigma^2$ known: $\mathbb{P}_\mu\left(-z_{1-\alpha/2} < \overbrace{\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}}}^{= Q(\vec{X}, \mu) \sim N(0,1) \text{ under } \mu} < z_{1-\alpha/2}\right) = 1 - \alpha$

- Example 4.13: $\text{Unif}(0, \theta)$: $\mathbb{P}_\theta\left(\frac{1}{2a} \leq \underbrace{\frac{X_{(n)}}{\theta}}_{= Q(\vec{X}, \theta)\text{, distribution free of } \theta} \leq \frac{1}{a}\right) = 1 - \alpha$ where $a$ chosen as before

# Pivotal Quantities

- The key in these examples was that the *distribution* of $Q(\mathbf{X}, \theta)$ is free of $\theta$
- Definition 4.4: A random variable $Q(\mathbf{X}, \theta)$ is a **pivotal quantity** (or **pivot**) for $\theta$ if its distribution is free of $\theta$.
- So if $\mathbf{X} \sim f_{\theta_1}$ and $\mathbf{Y} \sim f_{\theta_2}$, then $Q(\mathbf{X}, \theta_1) \overset{d}{=} Q(\mathbf{Y}, \theta_2)$
- Every ancillary statistic is a pivotal quantity
- Example 4.14: $N(\mu, \sigma^2)$, $\sigma^2$ known: $\mathbb{P}_\mu\left(-z_{1-\alpha/2} < \overbrace{\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}}}^{= Q(\vec{X}, \mu) \sim N(0,1) \text{ under } \mu} < z_{1-\alpha/2}\right) = 1-\alpha$
- Example 4.15: $Exp(\lambda)$: $Q(\vec{X}, \lambda) = \frac{X_1}{\lambda} \sim Exp(1)$ is pivotal.

# Poll Time!

# Confidence Intervals from Pivotal Quantities

- Example 4.16: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\lambda)$, $\lambda > 0$. Show that $Q(\mathbf{X}, \lambda) = 2\lambda \sum_{i=1}^{n} X_i$ is a pivotal quantity, and use it to find a $1-\alpha$ confidence interval for $\lambda$.

Use mgfs! $M_{\Sigma X_i}(t) = \left(\frac{\lambda}{\lambda - t}\right)^n, \ t < \lambda$

$\Rightarrow M_{2\lambda \Sigma X_i}(t) = \left(\frac{\lambda}{\lambda - 2\lambda t}\right)^n = (1-2t)^{-n} \Rightarrow Q(\vec{X}, \lambda) \sim \chi^2_{(2n)}$ free of $\lambda$! So pivotal.

Set $1-\alpha = P(a < 2\lambda \Sigma X_i < b)$ for some $a, b \in \mathbb{R}$

where $a$ and $b$ satisfy $1-\alpha = F_{\chi^2_{(2n)}}(b) - F_{\chi^2_{(2n)}}(a)$.

Can pick, for example, $a = 0 \Rightarrow b = \chi^2_{(2n),1-\alpha}$.

Then $1-\alpha = P\left(0 < \lambda < \frac{\chi^2_{(2n),1-\alpha}}{2\Sigma X_i}\right)$

$\Rightarrow$ choose $\left(0, \frac{\chi^2_{(2n),1-\alpha}}{2\Sigma X_i}\right)$

# Finding Pivotal Quantities

- There's no all-purpose strategy to finding pivotal quantities, but there's a neat trick that sometimes lets us pull one out of the pdf of a statistic $T(\mathbf{X})$

- Theorem 4.1: Suppose that $T(\mathbf{X}) \sim f_\theta$ is univariate and continuous, such that the pdf can be expressed as

$$f_\theta(t) = g(Q(t,\theta)) \cdot \left| \frac{\partial}{\partial t} Q(t,\theta) \right|$$

  for some function $g(\cdot)$ which is free of $\theta$ and some function $Q(t,\theta)$ which is continuously differentiable and one-to-one as a function of $t$ (i.e., with $\theta$ fixed). Then $Q(T(\mathbf{X}),\theta)$ is a pivot.

*Proof.* Fix $\theta \in \Theta$ and let $h_\theta(q)$ be the pdf of $Q(T(\vec{X}),\theta)$. We'll just write $Q(T(\vec{X}))$. Let $Q^{-1}(q)$ be the functional inverse of $Q(t)$. Then

$$\begin{aligned} h_\theta(q) &= f_\theta(Q^{-1}(q)) \cdot \left| \frac{\partial}{\partial q} Q^{-1}(q) \right| \\ &= f_\theta(Q^{-1}(q)) \cdot \left| \frac{\partial}{\partial t} Q(t) \Big|_{t=Q^{-1}(q)} \right|^{-1} \quad \text{by the inverse function theorem} \\ &= g(Q(Q^{-1}(q))) \cdot \left| \frac{\partial}{\partial t} Q(t) \Big|_{t=Q^{-1}(q)} \right| \cdot \left| \frac{\partial}{\partial t} Q(t) \Big|_{t=Q^{-1}(q)} \right|^{-1} = g(q) \quad \text{which is free of } \theta. \ \square \end{aligned}$$

# Finding Pivotal Quantities: Examples

- Example 4.17: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$ where $\theta > 0$. Find a pivotal quantity based on $T(\mathbf{X}) = X_{(n)}$, and use it to construct a $1-\alpha$ confidence interval for $\theta$.

The pdf of $T(\vec{X}) = X_{(n)}$ is $n \cdot f_\theta(t) \cdot F_\theta(t)^{n-1}$

$$= n \cdot \frac{1}{\theta} \cdot \left(\frac{t}{\theta}\right)^{n-1}$$

$$= \frac{nt^{n-1}}{\theta^n}$$

$$= \underbrace{1}_{g(\cdot)} \cdot \left|\frac{\partial}{\partial t} \underbrace{\frac{t^n}{\theta^n}}_{Q(t,\theta)}\right|$$

By Theorem 4.1,

$Q(X_{(n)}, \theta) = \dfrac{X_{(n)}^n}{\theta^n}$ is a pivotal quantity.

What's its distribution? $P\left(\dfrac{X_{(n)}^n}{\theta^n} \le x\right)$

$$= P(X_{(n)} \le \theta x^{1/n})$$

$$= F_\theta(\theta x^{1/n})^n$$

$$= \left(\frac{\theta x^{1/n}}{\theta}\right)^n$$

$$= x$$

So $Q(X_{(n)}, \theta) \sim \text{Unif}(0,1)$.

Choose $a, b$ s.t. $P\left(a < \dfrac{X_{(n)}^n}{\theta^n} < b\right) = 1-\alpha$

For example, $a = \alpha/2$, $b = 1-\alpha/2$ works.

Then $1-\alpha = P\left(\alpha/2 < \dfrac{X_{(n)}^n}{\theta^n} < 1-\alpha/2\right)$

$$= P\left(\frac{X_{(n)}^n}{1-\alpha/2} \le \theta^n \le \frac{X_{(n)}^n}{\alpha/2}\right)$$

$\Rightarrow$ Choose $\left(\dfrac{X_{(n)}}{(1-\alpha/2)^{1/n}}, \dfrac{X_{(n)}}{(\alpha/2)^{1/n}}\right)$.

# Finding Pivotal Quantities: Examples

- Example 4.18: Let $X \sim f_\theta(x) = \frac{2(\theta - x)}{\theta^2} \cdot \mathbb{1}_{0 \le x \le \theta}$, where $\theta > 0$. Find a pivotal quantity based on $X$, and use it to construct a $1-\alpha$ confidence interval for $\theta$.

Observe if $Q(x,\theta) = \frac{\theta - x}{\theta}$, then $f_\theta(x) = g(Q(x,\theta)) \cdot \left|\frac{\partial}{\partial x} Q(x,\theta)\right|$ where $g(x) = 2x$.

By Theorem 4.1, $Q(X,\theta) = \frac{\theta - X}{\theta}$ is a pivotal quantity, with distribution...?

$$\mathbb{P}\left(\frac{\theta - X}{\theta} \le x\right)$$

$$= \mathbb{P}(X \ge (1-x)\cdot\theta)$$

$$= \int_{(1-x)\cdot\theta}^{\theta} \frac{2(\theta - t)}{\theta^2}\, dt$$

$$= x^2, \quad x \in (0,1).$$

Plenty of choices for $a,b$ in $1-\alpha = \mathbb{P}\left(a < \frac{\theta - X}{\theta} < b\right) = b^2 - a^2$

If we choose $a = 0$, then $b = \sqrt{1-\alpha}$. Therefore,

$$1-\alpha = \mathbb{P}\left(0 < \frac{\theta - X}{\theta} < \sqrt{1-\alpha}\right)$$

$$= \mathbb{P}\left(X < \theta < \frac{X}{1-\sqrt{1-\alpha}}\right)$$

$$0 < 1 - \frac{X}{\theta} < \sqrt{1-\alpha}$$

$$-1 < -\frac{X}{\theta} < \sqrt{1-\alpha} - 1$$

$$-1 > -\frac{\theta}{X} > \frac{1}{\sqrt{1-\alpha} - 1}$$

$$-X > -\theta > \frac{X}{\sqrt{1-\alpha} - 1}$$

$$X < \theta < \frac{X}{1-\sqrt{1-\alpha}}$$

$\Rightarrow$ Choose $\left(X, \frac{X}{1-\sqrt{1-\alpha}}\right)$.

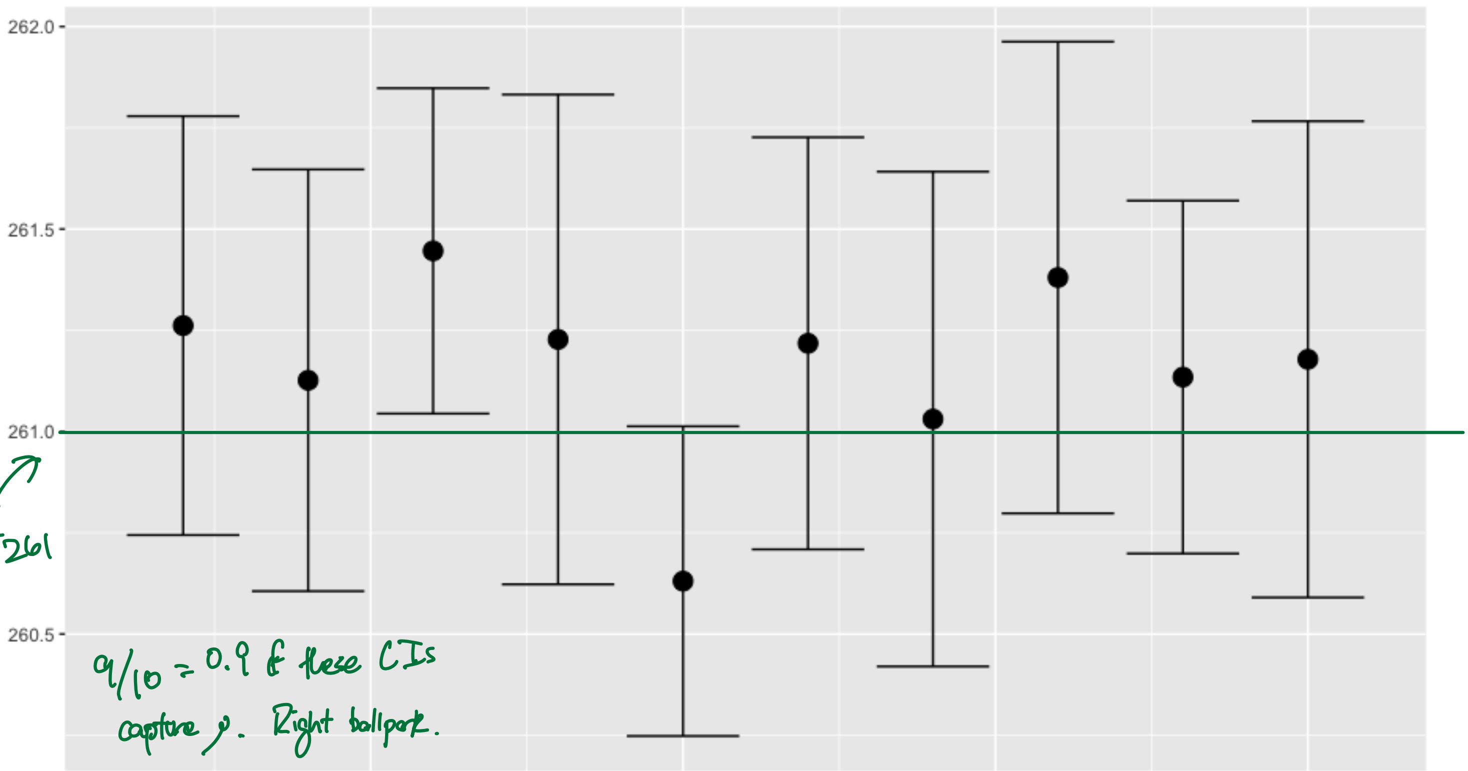
# Confidence Intervals: Interpretations

- Confidence intervals are almost as widely misinterpreted as $p$-values
- Suppose that in a published scientific study, you see a stated 95% confidence interval such as $(0.932, 1.452)$
- How do you interpret this correctly?
- Should we be surprised if we try and reproduce the study and calculate a 95% confidence interval of $(0.824, 1.734)$?
- What about $(-0.232, 1.440)$?

# Poll Time!

# Confidence Intervals: Interpretations

- Here are ten observed 95% $Z$-intervals for $\mu$ calculated from ten random samples of size $n = 15$ from a $\mathcal{N}(\mu, 1)$ distribution:




True $\mu = 261$

$9/10 = 0.9$ of these CIs capture $\mu$. Right ballpark.

# Questioning Our Assumptions...

- All of the theory we've done up to this point has depended on the assumption of an underlying statistical model
- When we say "Suppose $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$...", we're assuming the data follows one of the distributions in the parametric family $\{f_\theta : \theta \in \Theta\}$ and only the parameter $\theta$ is unknown
- If we get the statistical model wrong, then any inferences we make about $\theta$ are likely to be completely invalid
- So it's extremely important to be able to check that statistical model assumption

# Nothing Is Certain

- Of course, we can't *know* for sure that a model is correct
- Unless we generate the data, but then what's the point?
- But we can perform checks that give us confidence in our assumptions
- This is called *model checking*
- We will study two kinds of model checks: visual diagnostics and goodness-of-fit tests

# Histograms: Preliminaries

- Suppose we have iid data $X_1, X_2, \ldots, X_n$, which we hypothesize are distributed according to a cdf $F_\theta$
- Let's group the range of the data into bins $[h_1, h_2], (h_2, h_3], \ldots, (h_{m-1}, h_m]$
- By the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \in (h_j, h_{j+1}]} \xrightarrow{p} \mathbb{P}_\theta(X_i \in (h_j, h_{j+1}])$$

$$= \mathbb{P}_\theta(h_j < X_i \leq h_{j+1})$$

$$= F_\theta(h_{j+1}) - F_\theta(h_j)$$

- So if $n$ is large and we're correct about $F_X$, then

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \in (h_j, h_{j+1}]} \approx F_\theta(h_{j+1}) - F_\theta(h_j)$$

# The Histogram Density Function

- If, in addition, we believe $F_{\cancel{X}\theta}$ is continuous with pdf $f_{\cancel{X}\theta}$, then there exists $h^* \in (h_j, h_{j+1})$ such that

$$\frac{1}{n(h_{j+1} - h_j)} \sum_{i=1}^{n} \mathbb{1}_{X_i \in (h_j, h_{j+1}]} \approx \frac{F_\theta(h_{j+1}) - F_\theta(h_j)}{h_{j+1} - h_j} = f_\theta(h^*)$$

by the mean value theorem!

- Definition 4.5: Given data $X_1, \ldots, X_n$ and a partition $h_1 < h_2 < \cdots < h_m$, the **density histogram function** is defined as

$$\hat{f}_n(t) = \begin{cases} \frac{1}{n(h_{j+1} - h_j)} \sum_{i=1}^{n} \mathbb{1}_{X_i \in (h_j, h_{j+1}]}, & t \in (h_j, h_{j+1}] \\ 0, & \text{otherwise} \end{cases}$$

# Histograms

- If we believe that our observed data $x_1, \ldots, x_n$ are realizations of $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$, then the observed $\hat{f}_n(t)$ should look like a "discretized" version of $f_\theta(t)$

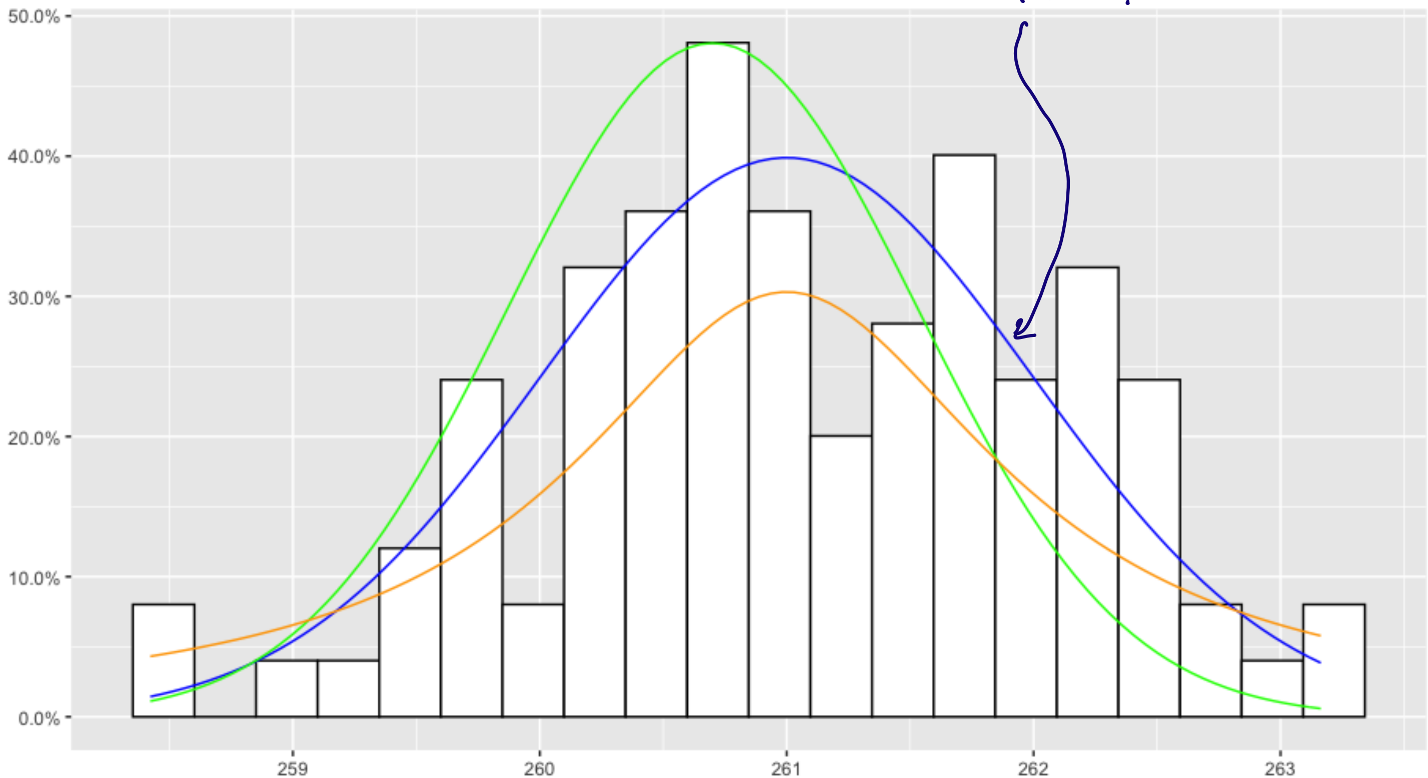
- ...and the resemblance should improve as $n$ gets larger and each bin size $h_{j+1} - h_j$ gets smaller

- Definition 4.6: A plot of a density histogram function $\hat{f}_n(t)$ with vertical lines drawn at each $h_j$ is called a **histogram**. A histogram where each bin width $h_{j+1} - h_j = 1$ is called a **relative frequency plot**.

# Histograms: An Example

- Here's a histogram ($n = 100$) overlaid with three hypothesized pdfs; which is more likely to have generated the data?

This one! The blue one.



# Poll Time!

# Empirical CDFs

- We might prefer to deal with the cdf $F_{\cancel{X}\theta}$ instead

- If we fix any $t \in \mathbb{R}$, then the law of large numbers says that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \le t} \xrightarrow{p} \mathbb{P}_\theta\left(X_i \le t\right)$$

- So if $n$ is large and we're correct about $F_{\cancel{X}\theta}$, then

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \le t} \approx F_\theta(t)$$

- Definition 4.7: Given observations $\cancel{x}X_1, \dots, \cancel{x}X_n$, the **empirical distribution function (ecdf)** is defined as

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \le t}$$

# Empirical CDFs Are Nice

- If we believe that our observed data $x_1, \ldots, x_n$ are realizations of $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_{\cancel{X}\theta}$, then $\hat{F}_n(t)$ should look like $F_\theta(t)$

- In fact, a famous result called the **Glivenko-Cantelli theorem** says that if $F_X$ really is the true cdf, then $\hat{F}_n(t) \longrightarrow F_\theta(t)$ as $n \to \infty$ in a *much* stronger sense than convergence in probability

"Uniformly almost surely" $\mathbb{P}_\theta\left( \sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F_\theta(t)| > \varepsilon \right) \xrightarrow{n \to \infty} 0$

- Theorem 4.2: For any fixed $t \in \mathbb{R}$, the ecdf $\hat{F}_n(t)$ is an unbiased estimator of $F_\theta(t)$, and it has a lower variance than $\mathbb{1}_{X_i \le t}$.

*Proof.*

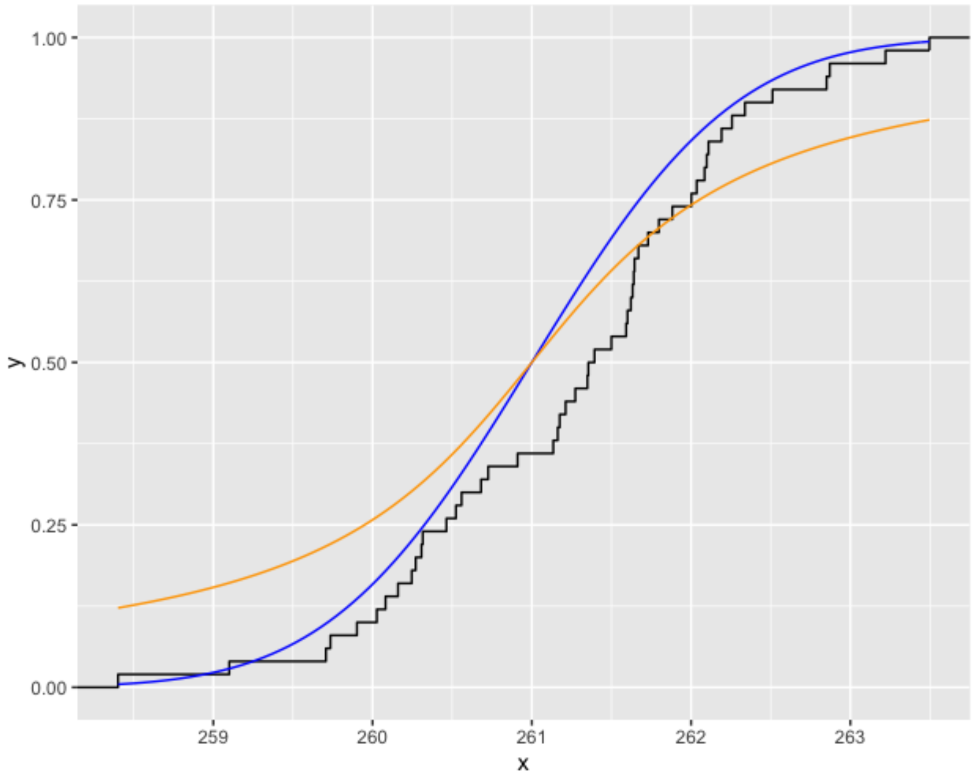
$$\mathbb{1}_{X_i \le t} \sim \text{Bernoulli}(\mathbb{P}_\theta(\mathbb{1}_{X_i \le t} = 1))$$
$$= \text{Bernoulli}(\mathbb{P}_\theta(X_i \le t))$$
$$= \text{Bernoulli}(F_\theta(t))$$

$$\mathbb{E}_\theta[\hat{F}_n(t)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\mathbb{1}_{X_i \le t}] = F_\theta(t)$$

$$\text{Var}_\theta(\hat{F}_n(t)) = \frac{1}{n}\text{Var}_\theta(\mathbb{1}_{X_i \le t}) = \frac{1}{n} F_\theta(t) \cdot (1 - F_\theta(t)) \le F_\theta(t) \cdot (1 - F_\theta(t)) = \text{Var}_\theta(\mathbb{1}_{X_i \le t}). \ \square$$

# Empirical CDFs: An Example

- Here's an ecdf ($n = 50$) overlaid with two hypothesized cdfs; which is more likely to have generated the data?



# Poll Time!

$X_1, \ldots, X_n \overset{iid}{\sim} N(0,1)$.

What's $\mathbb{E}[\hat{F}_n(0)]$ ?

$$\mathbb{E}[\hat{F}_n(0)] = F(0)$$
$$= \Phi(0)$$
$$= 1/2$$

# Bringing Back Ancillarity and Sufficiency

- We know from Module 1 that if $\mathbf{X} \sim f_\theta$, the distribution of an ancillary statistic $S(\mathbf{X})$ is free of $\theta$
- But if we've gotten the model $\{f_\theta : \theta \in \Theta\}$ wrong, $S(\mathbf{X})$ could very well depend on $\theta$! (or some unknown parameter in the "true" model)
- So some ancillary statistics provide a model check: if our realization $S(\mathbf{x})$ is "surprising", we have evidence against the model being true
- Similarly, if $T(\mathbf{X})$ is sufficient for $\theta$, then $\mathbf{X} \mid T(\mathbf{X}) = t$ shouldn't depend on $\theta$
- This leads to the idea of **residual analysis**
- Loosely speaking, residuals are based on the information in the data that is left over after we have fit the model

(actually no formal definition of "residual"!)

# Residual Plots

- Example 4.19: Let $X_1, \ldots, X_n$ be a random sample from a suspected $\mathcal{N}(\mu, \sigma^2)$ distribution, where $\mu \in \mathbb{R}$ and $\sigma^2$ is known. If we're correct, then $R(\mathbf{X}) = (X_1 - \bar{X}, \ldots, X_n - \bar{X})$ is ancillary for $\mu$, because

$$X_i - \bar{X} \sim \mathcal{N}\left(0, \frac{n-1}{n}\sigma^2\right), \quad i = 1, \ldots, n$$

Assignment 1 Q17

and therefore **standardized residuals**

$$R_i^*(\mathbf{X}) := \frac{X_i - \bar{X}}{\sqrt{\frac{n-1}{n}\sigma^2}} \sim \mathcal{N}(0, 1).$$

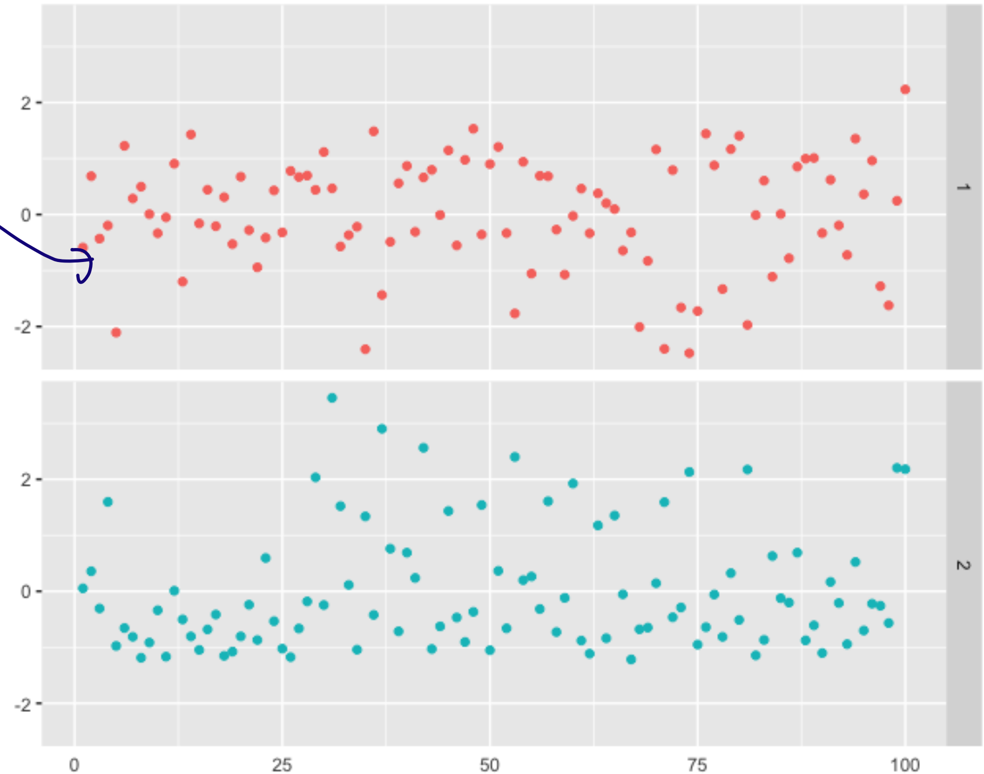
So if we're right about $\mathcal{N}(\mu, \sigma^2)$, then a scatterplot of the residuals shouldn't exhibit any discernable pattern, and should mostly stay within $(-3, 3)$

If $\sigma^2$ is unknown, can swap with $S^2$, whence $R_i^*(\vec{X}) \sim t_{(n-1)}$

# Residual Plots

- Example 4.20: Here are two standardized residual plots constructed from two samples ($n = 100$) with equal variances $\sigma^2$; which looks more like it came from a $\mathcal{N}(\mu, \sigma^2)$ distribution?

It was this one!



# Probability Plots

- Probability plots extend this idea

- We need a fundamental result of probability theory first

- Theorem 4.3 (**Probability integral transform**): Let $X$ be a continuous random variable with cdf $F_\theta(x)$, and let $U = F_\theta(X)$. Then $U \sim$ Unif $(0, 1)$.

  (STA257)

- The order statistics of $U_1, \ldots, U_n \overset{iid}{\sim}$ Unif $(0, 1)$ follow a Beta distribution: $U_{(j)} \sim$ Beta $(j, n-j+1)$, and so $\mathbb{E}\left[U_{(j)}\right] = \frac{j}{n+1}$ Assignment 0.

- This suggests a recipe: If we hypothesize $X_1, \ldots, X_n \sim F_\theta$, then plot $F_\theta(X_{(j)})$ versus $\frac{j}{n+1}$, $j = 1, \ldots, n$.

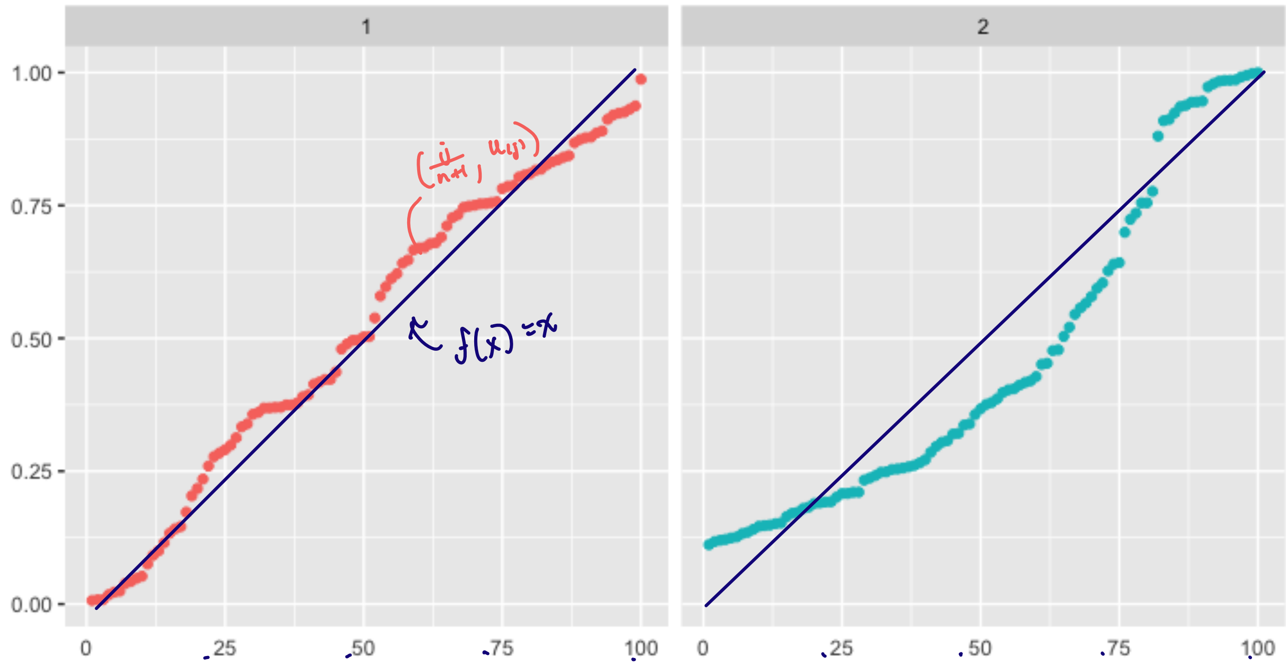
  $= [F_\theta(X)]_{(j)}$ since cdfs are increasing

  $= U_{(j)}$ if $F_\theta$ is correct

  If it doesn't look like a straight line, we should question $F_\theta$.

# Probability Plots

- Example 4.21: Here are two probability plots constructed from the standardized residuals as before, using $F_\theta(x) = \Phi(x)$. Which looks more like it came from a $\mathcal{N}\left(\mu, \sigma^2\right)$ distribution?



# Q-Q Plots

- We could also go in the other direction by looking at the quantiles
- Definition 4.8: Let $X$ be a random variable with cdf $F_\theta$. The **inverse cdf** (or the **quantile function**) is defined by $F_\theta^{-1}(t) = \inf\{x : F_\theta(x) \geq t\}$.

  "generalized inverse"
- When $X$ is continuous, the inverse cdf is simply the functional inverse of $F_\theta$
- There are plenty of software algorithms that can estimate the quantiles from a sample $x_1, \ldots, x_n$
- If we hypothesize $X_1, \ldots, X_n \sim F_\theta$ and we can compute $F_\theta^{-1}$, then we have another recipe for model checking:

  Plot the observed quantiles versus the theoretical ones!

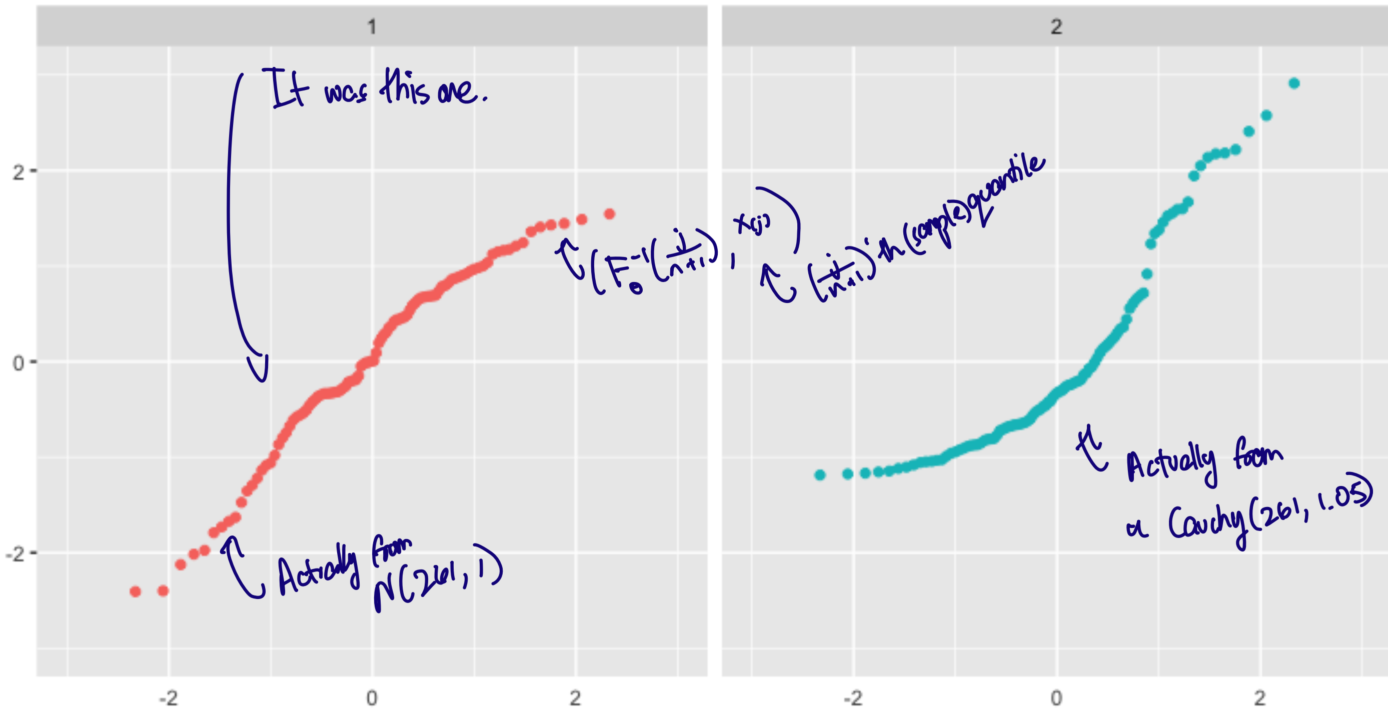
  If it doesn't look like $y=x$, we should question $F_\theta$.

# Q-Q Plots

By far, the most popular use is for $F_\theta = \Phi(\cdot)$.

Want to see if the $N(0,1)$ distribution does a good job of capturing the EXTREME observations

- Example 4.22: Here are two Q-Q plots constructed from the standardized residuals as before, using $F_\theta^{-1}(x) = \Phi^{-1}(x)$. Which looks more like it came from a $\mathcal{N}\left(\mu, \sigma^2\right)$ distribution?



# Q-Q Plots

- Q-Q plots are most frequently used as a test for Normality
- But technically there's no reason why we can't use them to compare *any* two distributions, observed or hypothesized
- ...provided we can actually compute (or estimate) their quantiles, of course
- Q-Q plots are particularly useful when we want to see how the "outliers" in our data compare to the extreme values predicted by the tails of a hypothesized distribution

# Goodness of Fit Tests

- Instead of using visual diagnostics, we can use hypothesis tests as model checks
- This time, the null hypothesis $H_0$ is that the model $\{f_\theta : \theta \in \Theta\}$ for our data is "correct" Eg: $H_0$: the data is Normally distributed $H_0$: the observations are independent
  $H_0$: two distributions generating two different samples are independent
- As usual, we have a test statistic $T(\mathbf{X})$ that follows some known distribution under $H_0$
- An observed value $T(\mathbf{x})$ which is very unlikely under $H_0$ (as quantified by a $p$-value) provides evidence that the model is wrong
- Such hypothesis tests are called **goodness of fit tests**

# Towards a Foundational Test

- Suppose we observe iid random variables $W_1, W_2, \ldots, W_n$ taking values in sample space $\mathcal{X} = \{1, 2, \ldots, k\}$, which we think of as *labels* or *categories*
- We want to test whether the $W_i$'s are distributed according to some hypothesized probability measure $\mathbb{P}_0$ on $\mathcal{X}$
- Let $X_i = \sum_{j=1}^{n} \mathbb{1}_{W_j = i}$ and let $p_i = \mathbb{P}_0(\{i\})$ so that under $H_0$,

$$(X_1, X_2, \ldots, X_k) \sim \text{Multinomial}(n, p_1, \ldots, p_k)$$

$X_i = \sum_{j=1}^{n} Y_j$, $Y_j \sim \text{Bernoulli}(p_i)$

- Now define

$$R_i = \frac{X_i - \mathbb{E}[X_i]}{\sqrt{\text{Var}(X_i)}} \stackrel{H_0}{=} \frac{X_i - np_i}{\sqrt{np_i(1-p_i)}}$$

- Since $R_i \xrightarrow{d} \mathcal{N}(0,1)$ under $H_0$, it's tempting to think $\sum_{i=1}^{k} R_i^2 \xrightarrow{d} \chi^2_{(k)}$, but that's not true They're not independent! For the multinomial, we need $\sum_{i=1}^{k} X_i = n$

# Pearson's Chi-Squared Test

- Instead, we have the following result

- Theorem 4.4: If $(X_1, X_2, \ldots, X_k) \sim$ Multinomial$(n, p_1, \ldots, p_k)$, then

$$\sum_{i=1}^{k}(1-p_i)R_i^2 = \sum_{i=1}^{k}\frac{(X_i - np_i)^2}{np_i} \xrightarrow{d} \underbrace{\chi^2_{(k-1)}}_{\text{Asymptotic distribution under } H_0}.$$

- The statistic $\chi^2(\mathbf{X}) = \sum_{i=1}^{k}\frac{(X_i - np_i)^2}{np_i}$ is called a **chi-square statistic**, and it's almost always written as

$$\chi^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i}$$

$O_i$ = "observed" $i$

$E_i$ = "expected" $i$

- The chi-squared test is an *approximate test*, because the test statistic only has the $\chi^2_{(k-1)}$ distribution in the limit (more on this in Module 5)

# A Famous Example: Fisher and Mendel's Pea Data

- Mendelian laws of inheritance establish relative frequencies of dominant and recessive phenotypes across new generations

- Gregor Mendel was known for his pioneering experiments with pea plants in the mid-1800s

- If you cross smooth, yellow male peas with wrinkled, green female peas, Mendelian inheritance predicts these relative frequencies of traits in the progeny:

| | Yellow | Green |
|---|---|---|
| Smooth | $\frac{9}{16}$ | $\frac{3}{16}$ |
| Wrinkled | $\frac{3}{16}$ | $\frac{1}{16}$ |

# A Famous Example: Fisher and Mendel's Pea Data

- Mendel crossed 556 such pairs of peas together and recorded the following counts:

OBSERVED

| | Yellow | Green |
|---|---|---|
| Smooth | 315 | 108 |
| Wrinkled | 102 | 31 |

(EXPECTED)

| | Yellow | Green |
|---|---|---|
| Smooth | 312.75 | 104.25 |
| Wrinkled | 104.25 | 34.75 |

- Example 4.23: Do these results support the predicted frequencies?

$$\chi^2 = \frac{(315-312.75)^2}{312.75} + \frac{(108-104.25)^2}{104.25} + \frac{(102-104.25)^2}{104.25} + \frac{(31-34.75)^2}{34.75} \simeq 0.6043$$

Our p-value is $p(\vec{x}) = \mathbb{P}(\chi^2_{(3)} \geq 0.6043)$

$= 1 - \mathbb{P}(\chi^2_{(3)} < 0.6043)$

$\simeq 0.895$

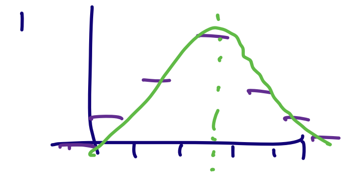
Check out the "Mendelian paradox"

So we fail to reject $H_0$.

# Extending the Chi-Squared Test

- What if our hypothesized distribution is not categorical, but quantitative?

- We can still use a chi-squared test – but how?



- The trick is to partition the sample space $\mathcal{X}$ into $k$ disjoint subsets $\mathcal{X}_1, \ldots, \mathcal{X}_k$, and let $X_i = \sum_{j=1}^{n} \mathbb{1}_{W_j \in \mathcal{X}_i}$ and $p_i = \mathbb{P}_0(\mathcal{X}_i)$

Eg: $\mathcal{X} = \mathbb{R}$. Maybe $\mathcal{X}_1 = (-\infty, -3]$, $\mathcal{X}_2 = (-3, 3]$, $\mathcal{X}_3 = (3, 27]$, $\mathcal{X}_4 = (27, \infty)$

- The finer our partition, the better we can distinguish between distributions

- But of course, we need to increase our sample size accordingly so that each category gets sufficiently "filled" with data

Guideline: each $\mathcal{X}_i$ should have at least 5 observations before doing this

If you have 0 observations in $\mathcal{X}_i$, then you can't hypothesize anything other than $p_i = 0$

# A Famous Example: Testing for Uniformity

- There are many reasons why we might want to test whether some data $U_1, \ldots, U_n$ arises from a Unif$(0,1)$ distribution

* Probability plot : use probability integral transform to make $U_1, \ldots, U_n \overset{iid}{\sim} \text{Unif}(0,1)$ under $H_0: F_\theta$. Basically quantitative version of visual probability plot check

* Random number generation: when simulating random data from some distribution, we need to start with $U_1, \ldots, U_n \overset{iid}{\sim} \text{Unif}(0,1)$ and then transform them. Can't truly generate random numbers, but we can make a sequence "look" random enough.

- We can use a chi-squared test for this by binning $[0,1]$ into $k$ equal-sized sub-intervals of length $1/k$, and letting $X_i = \sum_{j=1}^{n} \mathbb{1}_{U_j \in (\frac{i-1}{k}, \frac{i}{k}]}$ and $p_i = 1/k$

$= \mathbb{P}_0(U_j \in (\frac{i-1}{k}, \frac{i}{k}])$.

# A Famous Example: Testing for Uniformity

- Example 4.24: How can we test whether an iid sequence $U_1, U_2, \ldots U_n$ arises from a Unif$(0,1)$ distribution using $10$ categories?

Partition $[0,1]$ into $[0,\frac{1}{10}], (\frac{1}{10},\frac{2}{10}], \ldots, (\frac{9}{10},1]$

Let $X_i = \sum_{j=1}^{n} \mathbb{1}_{U_j \in [\frac{i-1}{10}, \frac{i}{10}]}$, $i = 1, \ldots, 10$

OR let $V_i = \lceil 10 U_i \rceil$, so $V_1, \ldots, V_n \overset{iid}{\sim} \text{Unif}(\{1,2,\ldots,10\})$ under $H_0$

and let $X_i = \sum_{j=1}^{n} \mathbb{1}_{V_j = i}$

Then carry out a $\chi^2$ goodness of fit test with $\chi^2 = \sum_{i=1}^{10} \frac{(X_i - n/10)^2}{n/10} \overset{H_0}{\underset{\text{approx}}{\sim}} \chi^2_{(9)}$.

Note: this is known to be a low-powered test.

There are way better randomness tests out there! The "diehard tests" are standard

# Other Goodness of Fit Tests

- Pearson's chi-squared isn't the only goodness of fit test out there; there are countless others

- Many apply to one particular parametric family specifically

  Eg: Normality testing: Shapiro-Wilk, Anderson-Darling, Jarque-Bera, ---

- Others are completely generic and test for equality between *any* two distributions

  Eg: Kolmogorov-Smirnov, Cramer-von Mises by far the most popular

- These latter tests allow us to compare an ecdf $\hat{F}_n$ to a hypothesized cdf $F_\theta$

  Very helpful! Basically a quantitative version of the visual ecdf-vs-$F_\theta$ check

# Other Goodness of Fit Tests

- In most cases, the distribution of the test statistic under $H_0$ is only known in the limit as $n \to \infty$
- Even then, cutoffs often can't be calculated exactly and require simulations to approximate
- When there's more than one test out there for the same thing, it's always a good idea to read up on the benefits/drawbacks of each one before deciding which to use
- One might have a lower probability of Type I error, another might higher power for lower sample sizes, another might be more robust to outliers, and so on

Very active area of research!