STA261 - Module 4

Intervals and Model Checking

Rob Zimmerman

University of Toronto

July 26-28, 2022

Uncertainty in Point Estimates

- In Module 2, we learned how to produce the "best" point estimates of θ 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 θ

- 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 θ?

- We can't know for sure

- But we can use the data to get a range of *plausible* values of θ

Random Sets

- Suppose for now that $\Theta \subseteq \mathbb{R}$
- If $\hat{\theta}(\mathbf{X})$ is a continuous random variable, then $\mathbb{P}_\theta\left(\theta = \hat{\theta}(\mathbf{X})\right) = 0$
- 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
- 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 $\mathbb{P}_\mu(\mu \in C(X)) = 0.95$.

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 $(L(\mathbf{X}), U(\mathbf{X}))$ is called an **interval estimator**.

- Example 4.2:

- 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 θ if $\mathbb{P}_\theta\left(L(\mathbf{X}) < \theta < U(\mathbf{X})\right) \geq 1 - \alpha$ for all $\theta \in \Theta$. We refer to $1 - \alpha$ as the **confidence level** of the interval.

- Example 4.3:

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 μ.

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

Poll Time!

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 $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim}$ Bernoulli (θ), where $\theta \in (0, 1)$. Find a 100% confidence interval for θ.

- 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 θ

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-α 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_{θ_0} such that

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

- This is equivalent to finding an *acceptance region* $A_{\theta_0} = R^c_{\theta_0}$ 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*

-

Famous Examples: Z-Intervals

- Example 4.7: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ where $\mu \in \mathbb{R}$ and σ^2 is known. Find a $(1-\alpha)$-confidence interval for μ by inverting the two-sided Z-test.

Famous Examples: One-Sided Z-Intervals

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

Famous Examples: t-Intervals

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

Famous Examples: One-Sided t-Intervals

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

An LRT-Based Interval

- 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 θ.

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 θ, 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 θ

- Example 4.12:

- Example 4.13:

Pivotal Quantities

- The key in these examples was that the *distribution* of $Q(\mathbf{X}, \theta)$ is free of θ
- Definition 4.4: A random variable $Q(\mathbf{X}, \theta)$ is a **pivotal quantity** (or **pivot**) for θ if its distribution is free of θ.
- 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:
- Example 4.15:

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 λ.

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

Proof.

Finding Pivotal Quantities: Examples

- Example 4.17: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ 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 θ.

Finding Pivotal Quantities: Examples

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

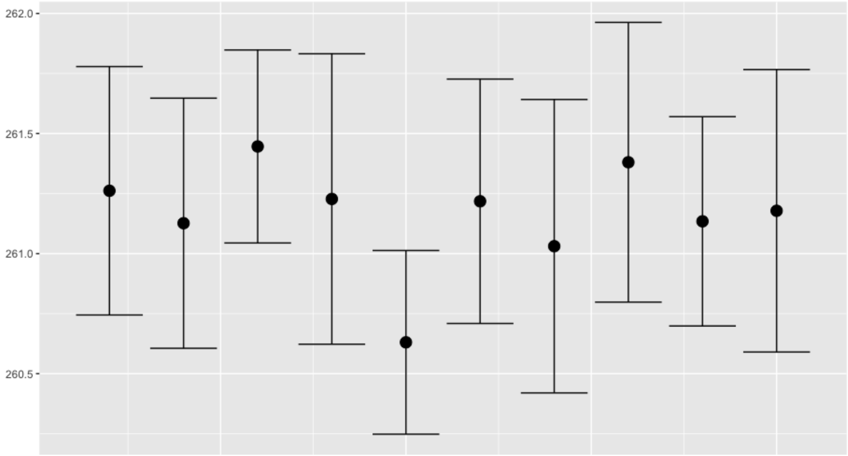
Confidence Intervals: Interpretations

- 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 μ calculated from ten random samples of size $n = 15$ from a $\mathcal{N}(\mu, 1)$ distribution:



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 θ is unknown

- If we get the statistical model wrong, then any inferences we make about θ 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
-
- 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 $X_1, X_2, \dots, X_n$, which we hypothesize are distributed according to a cdf F_θ

- Let's group the range of the data into bins $[h_1, h_2], (h_2, h_3], \dots, (h_{m-1}, h_m]$

- By the law of large numbers,

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

- 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_θ is continuous with pdf f_θ, then there exists $h^* \in (h_j, h_{j+1})$ such that

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

- 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

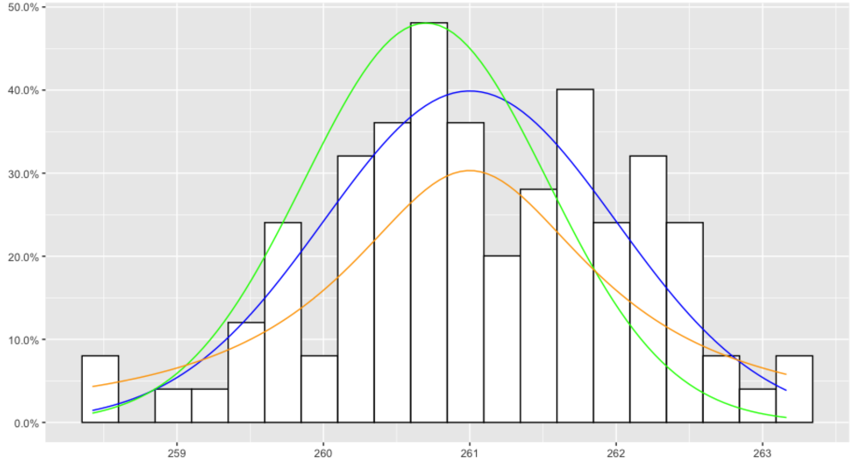
- If we believe that our observed data $x_1, \ldots, x_n$ are realizations of $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$, then the observed $\hat{f}_n(t)$ should look like a "discretized" version of $f_\theta(t)$

- ...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?



Poll Time!

Empirical CDFs

- We might prefer to deal with the cdf F_θ instead

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

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

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

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

- Definition 4.7: Given observations $X_1, \ldots, X_n$, the **empirical distribution function (ecdf)** is defined as

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

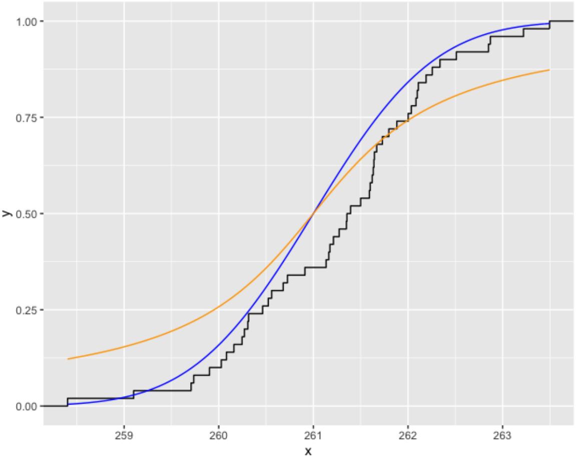
Empirical CDFs Are Nice

- If we believe that our observed data $x_1, \ldots, x_n$ are realizations of $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_\theta$, 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
- 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 $\mathbb{1}_{X_i \leq t}$.

Proof.

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!

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 θ
- But if we've gotten the model $\{f_\theta : \theta \in \Theta\}$ wrong, $S(\mathbf{X})$ could very well depend on θ!
- 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 θ, then $\mathbf{X} \mid T(\mathbf{X}) = t$ shouldn't depend on θ
- 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

Residual Plots

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

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

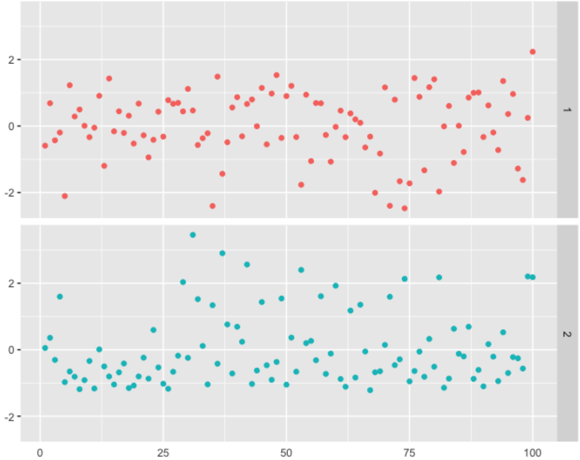
and therefore **standardized residuals**

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

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

Residual Plots

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



Probability Plots

- Probability plots extend this idea

- We need a fundamental result of probability theory first

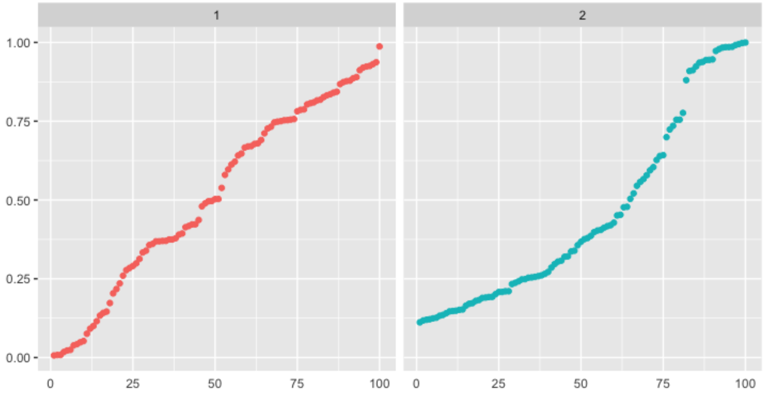
- Theorem 4.3 (**Probability integral transform**): Let X be a continuous random variable with cdf $F_\theta(x)$, and let $U = F_\theta(X)$. Then $U \sim \text{Unif}\,(0, 1)$.

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

- This suggests a recipe:

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?

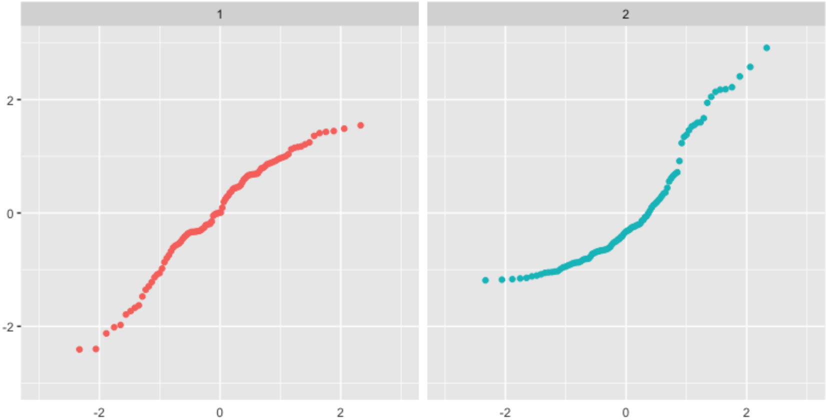


Q-Q Plots

- 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_θ. The **inverse cdf** (or the **quantile function**) is defined by $F_\theta^{-1}(t) = \inf\{x : F_\theta(x) \geq t\}$.
- When X is continuous, the inverse cdf is simply the functional inverse of F_θ
- 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_θ^{-1}, then we have another recipe for model checking:

Q-Q Plots

- 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"
- 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,

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

- Now define

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

- Since $R_i \xrightarrow{d} \mathcal{N}\left(0, 1\right)$ 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

Pearson's Chi-Squared Test

- Instead, we have the following result
- Theorem 4.4: If $(X_1, X_2, \dots, X_k) \sim \text{Multinomial}(n, p_1, \dots, p_k)$, then

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

- 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

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

- 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:

	Yellow	Green
Smooth	315	108
Wrinkled	102	31

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

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 $\mathcal{X}_1, \dots, \mathcal{X}_k$, and let $X_i = \sum_{j=1}^{n} \mathbb{1}_{W_j \in \mathcal{X}_i}$ and $p_i = \mathbb{P}_0(\mathcal{X}_i)$
- 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

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

- 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 $p_i =$

A Famous Example: Testing for Uniformity

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

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
- Others are completely generic and test for equality between *any* two distributions
- These latter tests allow us to compare an ecdf $\hat{F}_n$ to a hypothesized cdf F_θ

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