STA261 - Module 3

Hypothesis Testing

Rob Zimmerman

University of Toronto

July 16-18, 2024

Initial Hypotheses

- Consider our usual setup: we collect $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$ for some unknown $\theta \in \Theta$

- In Module 2, we learned how to produce the "best" point estimators of $\tau(\theta)$

- Now, we turn things around (sort of)

- Before observing $\mathbf{X} = \mathbf{x}$, we already have some conjecture/hypothesis about which specific value (or values) of $\theta \in \Theta$ generate $\mathbf{X}$

- Example 3.1:

Questions About Plausibility

- Suppose, for example, we initially suspect that $\theta = \theta_0$
- We find a good point estimator $\hat{\theta}(\mathbf{X})$ for θ, observe $\mathbf{X} = \mathbf{x}$, and produce the estimate $\hat{\theta}(\mathbf{x})$, which turns out to equal, say, $\theta_0 + 3$
- Is this evidence in favor of our initial suspicion, or against it?
- Is the difference of 3 "significant"?
- *Hypothesis testing* allows us to formulate this question rigorously (and answer it)

The Hypotheses in Hypothesis Testing

- **Null hypothesis significance testing (NHST)** (or **null hypothesis testing** or **statistical hypothesis testing**) is a framework for testing the plausibility of a statistical model based on observed data

- For better or worse, it has become a major component of statistical inference

- *Very* roughly speaking, NHST consists of three basic steps:

    1.

    2.

    3.

The “Hypothesis” in Hypothesis Testing

- Definition 3.1: A **hypothesis** is a statement about the statistical model that generates the data, which is either true or false.

- The negation of any hypothesis is another hypothesis, so they come in pairs

- Usually, we already have a parametric model $\{f_\theta : \theta \in \Theta\}$ in mind, and our hypotheses relate to the possible value (or values) of the parameter θ itself

- The two hypotheses in this setup can be written generically as $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta_0^c$, where $\Theta_0 \subset \Theta$ is some “default” set of parameters

- Example 3.2:

Kinds of Hypotheses

- We designate one hypothesis the **null hypothesis** (written H_0) and its negation the **alternative hypothesis** (written H_A or H_1)
- Mathematically speaking, any subjective meanings of the null and alternative hypotheses are irrelevant
- But in a scientific study, the null hypothesis typically represents the "status quo" or the "default" assumption
- The study is being conducted in the first place because we suspect the alternative hypothesis may be true instead

Simple and Composite Hypotheses

- Example 3.3:

- Example 3.4:

- Definition 3.2: Suppose a hypothesis H can be written in the form $H : \theta \in \Theta_0$ for some non-empty $\Theta_0 \subset \Theta$. If $|\Theta_0| = 1$, then H is a **simple hypothesis**. Otherwise, H is a **composite hypothesis**.

The Courtroom Analogy

- Consider a prosecution: the defendent is *innocent until proven guilty*
- But the whole point of the case is that the prosecutor suspects the defendent *is* guilty, and the purpose of the trial is to determine whether the evidence supports that guilt
- The jurors ask themselves: if the defendent really was innocent, how unlikely would this evidence be?
- If the evidence is overwhelmingly unlikely, the defendent is found guilty
- But if there's a *lack* of unlikely evidence, they find the defendent *not guilty*

A Motivating Example

- Example 3.5: Let $X_1, \ldots, X_{100} \overset{iid}{\sim} \mathcal{N}(\theta, 1)$, where $\theta \in \mathbb{R}$. Assess the plausibility that $\theta = 5$ if we observe $\bar{X} = -10$.

Hypothesis Tests and Rejection Regions

- Definition 3.3: A **hypothesis test** is a rule that specifies for which sample values the decision is made to reject H_0 in favour of H_A.

- Example 3.6:

- Definition 3.4: In a hypothesis test, the subset of the sample space for which H_0 will be rejected is called the **rejection region** (or **critical region**), and its complement is called the **acceptance region**.

- Given competing hypotheses H_0 and H_A, a hypothesis test is *characterized* by its rejection region $R \subseteq \mathcal{X}^n$

- In other words, $\mathbb{P}_\theta\left(\text{Reject } H_0\right) = \mathbb{P}_\theta\left(\mathbf{X} \in R\right)$

- Example 3.7:

Poll Time!

On Quercus: Module 3 - Poll 1

One-Tailed and Two-Tailed Tests

- If $\Theta \subseteq \mathbb{R}$ and H_0 is simple, then the rejection region is usually in both tails of the distribution:

- But if $H_0 : \theta \leq \theta_0$, then the rejection region is only in one tail:

- Definition 3.5: Suppose $\Theta \subseteq \mathbb{R}$. A **two-sided test** (or **two-tailed test**) has $H_0 : \theta = \theta_0$, for some $\theta_0 \in \Theta$. A **one-sided test** (or **one-tailed test**) has $H_0 : \theta \leq \theta_0$ or $H_0 : \theta \geq \theta_0$ for some $\theta_0 \in \Theta$.

Type I and Type II Errors

- Definition 3.6: A **type I error** is the rejection of H_0 when it is actually true. A **type II error** is the failure to reject H_0 when it is actually false.

- Example 3.8:

- Of course, we can never *know* if we are committing either of these errors

The Probability of Rejection

- Suppose the rejection region looks like $R = \{\mathbf{x} \in \mathcal{X}^n : \bar{x} \geq c\}$, for some $c \in \mathbb{R}$

- If we demand *very* strong evidence against H_0 before we would reject it, we might set c very high, which would make $\mathbb{P}_\theta(\mathbf{X} \in R) = \mathbb{P}_\theta(\bar{X} \geq c)$ very small under H_0

- In the standard framework, we choose the (low) probability *first*, and then calculate c based on that

- Example 3.9:

The Power Function

- Definition 3.7: The **power function** of a test with rejection region R is the function $\beta : \Theta \to [0,1]$ given by $\beta(\theta) = \mathbb{P}_\theta\left(\mathbf{X} \in R\right)$.

- Observe that

$$\beta(\theta) = \begin{cases} \mathbb{P}_\theta\left(\text{Type I error}\right), & \theta \in \Theta_0 \\ 1 - \mathbb{P}_\theta\left(\text{Type II error}\right), & \theta \in \Theta_0^c \end{cases}$$

- Definition 3.8: Let $\theta \in \Theta_0^c$. The **power** of a test at θ is defined as $\beta(\theta)$.

- Example 3.10:

The Power Function: Examples

- Example 3.11: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ with σ^2 known. Suppose a test of has a rejection region of the form $R = \{\mathbf{x} \in \mathcal{X}^n : \bar{x} > c\}$. Calculate the power function of this test.

Poll Time!

On Quercus: Module 3 - Poll 2

Size and the Probability of Rejection

- If we have a simple null hypothesis and $\mathbf{X}$ is continuous, we can often construct R so that $\mathbb{P}_{\theta_0}(\mathbf{X} \in R) = \alpha$, for some pre-chosen $\alpha \in (0, 1)$
- But for a more general null hypothesis $H_0 : \theta \in \Theta_0$, it's usually impossible to have $\mathbb{P}_\theta(\mathbf{X} \in R) = \alpha$ for all $\theta \in \Theta_0$
- Instead, we can try to ask for a "worst-case" probability
- Definition 3.9: The **size** of a test with rejection region R is a number $\alpha \in [0, 1]$ such that $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left(\mathbf{X} \in R\right) = \alpha$.
- Example 3.12:

Significance Levels

- A size-α test might be too much to ask for (especially when the underlying distribution is discrete)

- All we might be able to do is upper bound the worst-case probability

- Definition 3.10: The **level** (or **significance level**) of a test with rejection region R is a number $\alpha \in [0, 1]$ such that $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta (\mathbf{X} \in R) \leq \alpha$.

- Example 3.13:

Test Statistics

- A **test statistic** $T(\mathbf{X})$ is a statistic which is used to specify a hypothesis test
- The rejection region specifies which values of $T(\mathbf{X})$ have low probability under H_0
- If $R = \{\mathbf{x} \in \mathcal{X}^n : T(\mathbf{x}) \geq c\}$, then $\mathbb{P}_\theta(\mathbf{X} \in R) = \mathbb{P}_\theta(T(\mathbf{X}) \geq c)$, and evaluating that requires knowing the distribution of $T(\mathbf{X})$
- So a test statistic is only useful if we know its distribution under the null hypothesis
- Example 3.14:

p-Values

- Definition 3.11: Suppose that for every $\alpha \in (0, 1)$, we have a level-α test with rejection region R_α. For a given sample $\mathbf{X}$, the ***p*-value** is defined as

$$p(\mathbf{X}) = \inf\{\alpha \in (0, 1) : \mathbf{X} \in R_\alpha\}.$$

- The idea of a p-value may be the single most misinterpreted concept in statistics

p-Values Based On Test Statistics

- In non-specialist statistics courses, the p-value for a test with observed data $\mathbf{X} = \mathbf{x}$ is often defined as "the probability of obtaining data at least as extreme as the data observed, given that H_0 is true"

- At first glance, this bears no resemblance to the previous definition; however...

- Theorem 3.1: Suppose a test has rejection region of the form $R = \{\mathbf{x} \in \mathcal{X}^n : T(\mathbf{x}) \geq c\}$, for some test statistic $T : \mathcal{X}^n \to \mathbb{R}$. If we observe $\mathbf{X} = \mathbf{x}$, then our observed p-value is $p(\mathbf{x}) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left(T(\mathbf{X}) \geq T(\mathbf{x})\right)$.

- When H_0 is simple, that becomes $p(\mathbf{x}) = \mathbb{P}_{\theta_0}(T(\mathbf{X}) \geq T(\mathbf{x}))$

- Of course, the theorem also applies when the test specifies that low values of $T(\mathbf{x})$ are to be rejected

Poll Time!

On Quercus: Module 3 - Poll 3

Famous Examples: The Two-Sided Z-Test

- Example 3.15: Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ with $\mu \in \mathbb{R}$ and σ^2 known. Construct a size-α test of $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$ using the **Z-statistic**

$$Z(\mathbf{X}) = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}.$$

Famous Examples: The One-Sided Z-Test

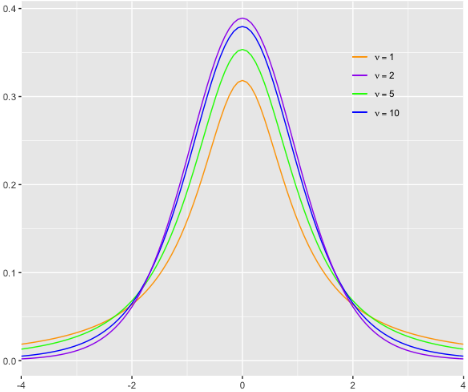
- Example 3.16: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ with $\mu \in \mathbb{R}$ and σ^2 known. Construct a size-α test of $H_0 : \mu \leq \mu_0$ versus $H_A : \mu > \mu_0$ using the Z-statistic.

The t-Distribution

- Definition 3.12: A real-valued random variable T is said to follow a **Student's t-distribution** with $\nu > 0$ degrees of freedom if its pdf is given by

$$f_T(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}.$$

We write this as $T \sim t_\nu$.



The t-Distribution: Important Properties

- Theorem 3.2: Let $Y, X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(0,1)$. Then

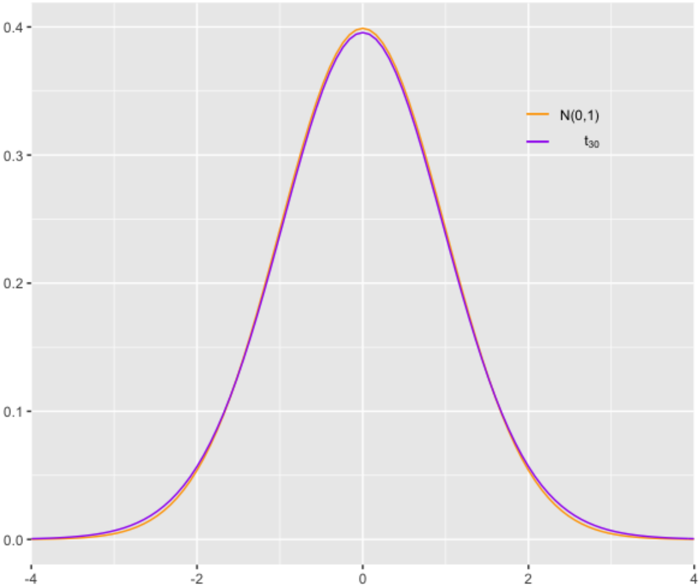
$$T = \frac{Y}{\sqrt{(X_1^2 + \cdots + X_n^2)/n}} \sim t_n.$$

-

- Theorem 3.3: Let $T_n \sim t_n$. Then $T_n \xrightarrow{d} Z$ as $n \to \infty$, where $Z \sim \mathcal{N}(0,1)$.

Proof.

A Great Approximation For Even Moderate n



The t-Distribution: More Important Properties

- The t-distribution is mainly used when we have $\mathcal{N}\left(\mu, \sigma^2\right)$ data and we're interested in μ, but σ^2 is unknown

- What happens if we swap σ^2 with S^2 in the Z-statistic?

- Theorem 3.4: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Then
$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}.$$

Proof.

Famous Examples: The Two-Sided t-Test

- Example 3.17: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Construct a size-α test of $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$ using the ***t*-statistic**

$$T(\mathbf{X}) = \frac{\bar{X} - \mu}{\sqrt{S^2/n}}.$$

Famous Examples: The One-Sided t-Test

- Example 3.18: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Construct a size-α test of $H_0 : \mu \geq \mu_0$ versus $H_A : \mu < \mu_0$ using the t-statistic.

Sample Size Calculations

- Usually, increasing our sample size increases the power of a test
- In real-world studies, obtaining a sample of independent data is typically quite expensive
- Whoever's paying for the study doesn't want experimenters collecting more data than necessary, since that costs money
- Moreoever, the larger the sample, the higher the chances of problems (errors in data entry, non-independence of some samples, etc.)
- So if we have demands for the power of our test at certain alternative parameters $\theta \in \Theta_0^c$, it's often useful to find the *minimum* sample size n that will give us that power

Sample Size Calculations

- Example 3.19: Suppose $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ where $\mu \in \mathbb{R}$ and σ^2 is known, and we want to test $H_0 : \mu \leq \mu_0$ versus $H_A : \mu > \mu_0$ using a test that rejects H_0 when $(\bar{X}_n - \mu_0)/\sqrt{\sigma^2/n} > c$, for some $c \in \mathbb{R}$. How can we choose c and n to obtain a size-0.1 test with a maximum Type II error probability of 0.2 if $\mu \geq \mu_0 + \sigma$?

The Problems With the p's

- Almost every scientific study that uses statistics will feature p-values somewhere

- The "strength" of a scientific conclusion often wrests upon those p-values

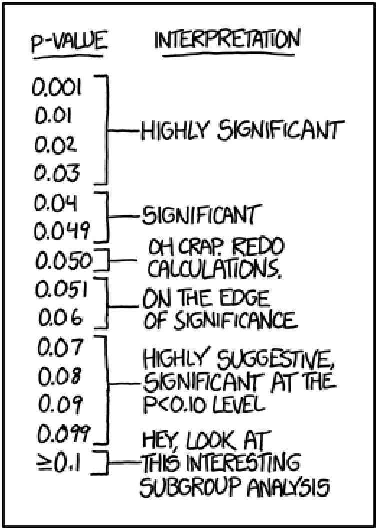
- Ronald Fisher suggested 5% as a reasonable significance level, and it's been widely adopted

-

- If every published study used significance levels of 5%, then on average, 1 out of every 20 studies make a type I error

- Think about how many scientific studies are published every day

The Problems With the p's



Source: https://xkcd.com/1478/

The Problems With the p's

- p-values lead to publication bias; the $p < 0.05$ threshold is so entrenched that a study result with $p = 0.06$ is considered a "negative" study
- Journals with limited space want to publish new, interesting, "positive" findings
- A study with $p > 0.05$ may contain important new information, but is far less likely to be published
- This pressure leads to ***p*-hacking**: "the misuse of data analysis to find patterns in data that can be presented as statistically significant, thus dramatically increasing and understating the risk of false positives."

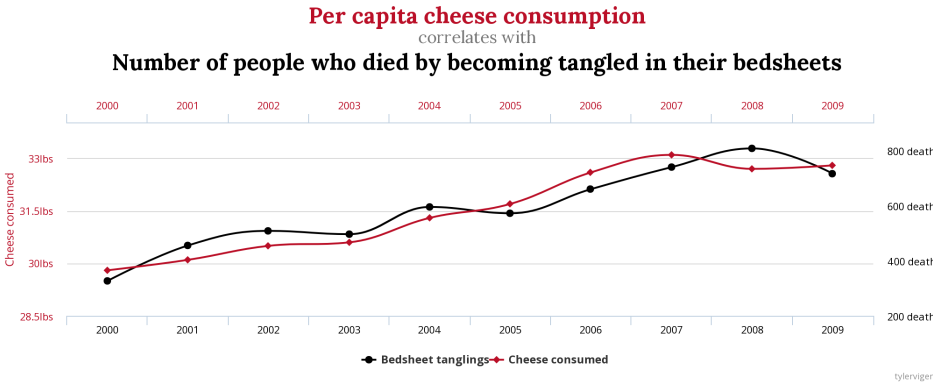
Examples of p-Hacking

- Changing α after seeing the data to declare the results statistically significant

- Increasing the size of the study population to produce a result that is statistically significant, but not *practically* significant

- Conducting multiple studies on the same data and "choosing" the one with significant results (this is called the **multiple comparisons problem**)

Should We Be Eating Less Cheese?



Source: https://www.tylervigen.com/

Poll Time!

On Quercus: Module 3 - Poll 4

Examples of p-Hacking

- Post-hoc analyses (i.e., testing hypotheses suggested by a given dataset)

- Outright fraud (such as "editing out" data points that sway the results away from the hoped-for conclusion, or simply lying about the p-value calculation in the hopes that no one will check)

- See also: the Replication Crisis

Bringing Back the Likelihood

- In Module 2, we saw that many common point estimators turned out to be MLEs

- It turns out that many common hypothesis tests are examples of an important kind of test based on the likelihood

- Definition 3.13: The **likelihood ratio test statistic** for testing $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta_0^c$ is defined as

$$\lambda(\mathbf{X}) = \frac{\sup_{\theta \in \Theta_0} L(\theta \mid \mathbf{X})}{\sup_{\theta \in \Theta} L(\theta \mid \mathbf{X})}.$$

  A **likelihood ratio test (LRT)** is any test that has a rejection region of the form $R = \{\mathbf{x} \in \mathcal{X}^n : \lambda(\mathbf{x}) \leq c\}$, for some $c \in [0, 1]$.

Poll Time!

On Quercus: Module 3 - Poll 5

LRTs: Examples

- Example 3.20: Show that the two-sided Z-test is an LRT.

LRTs: Examples

- Example 3.21: 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 $\theta \in \mathbb{R}$. Determine the LRT for testing $H_0 : \theta \leq \theta_0$ versus $H_A : \theta > \theta_0$.

Simple Tests Have Simple LRTs

- Theorem 3.5: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$. Suppose we want to test $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$ using an LRT. Then

$$\lambda(\mathbf{X}) = \frac{L(\theta_0 \mid \mathbf{X})}{L(\hat{\theta} \mid \mathbf{X})},$$

where $\hat{\theta}$ is the (unrestricted) MLE of θ based on $\mathbf{X}$.

- Example 3.22: Suppose $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}\,(0, \theta)$ where $\theta > 0$. Determine the LRT for testing $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$.

LRTs: Examples

- Example 3.23: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Bernoulli (θ) with $\theta \in (0, 1)$. Determine the LRT for testing $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$.

Making Life Easier With Sufficiency

- If $T(\mathbf{X})$ is some sufficient statistic with pdf/pmf $g_\theta(t)$, we might be interested in constructing an LRT based on its likelihood function $L^*(\theta \mid t) = g_\theta(t)$

- But would this change our conclusions?

- Theorem 3.6: Suppose $T(\mathbf{X})$ is sufficient for θ. If $\lambda(\mathbf{x})$ and $\lambda^*(T(\mathbf{x}))$ are the LRT statistics based on $\mathbf{X}$ and $T(\mathbf{X})$, respectively, then $\lambda^*(T(\mathbf{x})) = \lambda(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{X}^n$.

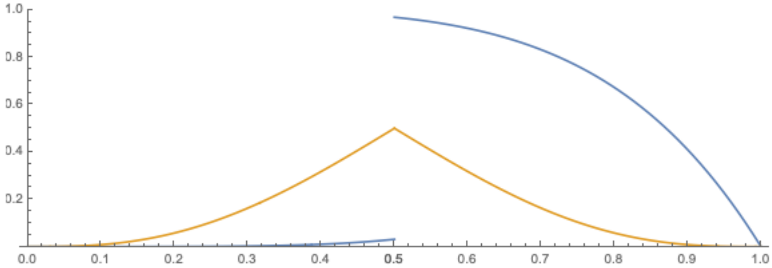
Proof.

Optimal Hypothesis Testing

- We have seen that there can be many tests of two competing hypotheses, with each test characterized by a rejection region
- What makes one test "better" than another?
- A natural idea is to try minimizing the probabilities of type I and type II errors
- Unfortunately, it's usually impossible to get both of these arbitrarily low

You Can't Get the Perfect Power Function

- Let $X \sim \text{Bin}(5, \theta)$, where $\theta \in (0, 1)$, and suppose we want to test $H_0 : \theta \leq \frac{1}{2}$ versus $H_A : \theta > \frac{1}{2}$; consider two different tests characterized by the following rejection regions: $R_1 = \{5\}$ and $R_2 = \{3, 4, 5\}$
-
-



A Compromise

- We have to settle on minimizing either type I error or type II error
- We will settle on the latter; that is, we fix a level α, and among all level-α tests, we try to find the one with the lowest probability of type II error
- This compromise isn't ideal for every real-life situation; sometimes, we care more about minimizing the probability of type I error
- Example 3.24:

Uniformly Most Powerful Tests

- Definition 3.14: A size-α (or level-α) test for testing $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta_0^c$ with power function $\beta(\cdot)$ is called a **uniformly most powerful (UMP) size-α (or level-α) test** if $\beta(\theta) \geq \beta'(\theta)$ for all $\theta \in \Theta_0^c$, where $\beta'(\cdot)$ is the power function of any other size-α (or level-α) test of the same hypotheses.

- UMP tests usually don't exist

- But when they do, how do we actually find them? How do we know that a test is UMP?

The Neyman-Pearson Lemma

- Theorem 3.7 (**Neyman-Pearson Lemma**): Consider testing $H_0 : \theta = \theta_0$ versus $H_A : \theta = \theta_1$. Consider a test whose rejection region R satisfies

$$\mathbf{x} \in R \text{ if } \frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} > c_0 \quad \text{and} \quad \mathbf{x} \in R^c \text{ if } \frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} < c_0$$

  for some $c_0 \geq 0$, and let $\alpha = \mathbb{P}_{\theta_0}(\mathbf{X} \in R)$. Then the test is a UMP level-α test. Moreover, *any* existing UMP level-α test has a rejection region that satisfies the above conditions.

- Why is the rejection region stated so strangely here? Why not just write $R = \left\{\mathbf{x} \in \mathcal{X}^n : \frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} > c_0\right\}$?

A Useful Corollary

- Theorem 3.8: Consider testing $H_0 : \theta = \theta_0$ versus $H_A : \theta = \theta_1$. Suppose $T(\mathbf{X}) \sim g_\theta$ is sufficient for θ. Then any test based on $T = T(\mathbf{X})$ with rejection region S is a UMP level-α test if it satisfies

$$t \in S \text{ if } \frac{g_{\theta_1}(t)}{g_{\theta_0}(t)} > k_0 \quad \text{and} \quad t \in S^c \text{ if } \frac{g_{\theta_1}(t)}{g_{\theta_0}(t)} < k_0$$

  for some $k_0 \geq 0$, where $\alpha = \mathbb{P}_{\theta_0}(T(\mathbf{X}) \in S)$.

The Neyman-Pearson Lemma: Examples

- Example 3.25: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ with $\mu \in \{\mu_0, \mu_1\}$ and σ^2 known. Find a UMP level-α test of $H_0 : \mu = \mu_0$ versus $H_A : \mu = \mu_1$, where $\mu_1 > \mu_A$.

Making Neyman-Pearson Useful

- There's one thing that keeps the Neyman-Pearson lemma from being useful in practice
- In real life, almost no one needs to test two simple hypotheses!
- On the other hand, one-sided tests are used in abundance
- Luckily, there's a way extend Neyman-Pearson that makes plenty of one-sided tests into UMP level-α tests
- We'll just look at a special case of this, which works when we have a sufficient statistic in an exponential family

The Karlin-Rubin Theorem

- Theorem 3.9 (**Karlin-Rubin**): Consider testing $H_0 : \theta \leq \theta_0$ versus $H_A : \theta > \theta_0$. Suppose $T = T(\mathbf{X}) \sim g_\theta$ is an $\mathbb{R}$-valued sufficient statistic for θ such that $g_{\theta_2}(t)/g_{\theta_1}(t)$ is monotone non-decreasing in t whenever $\theta_2 \geq \theta_1$. Then a test with rejection region $R = \{T > c_0\}$ is a UMP level-α test, where $\alpha = \mathbb{P}_{\theta_0}(T > c_0)$.

- By suitably restricting the entire parameter space, this also holds for a test of the form $H_0 : \theta = \theta_0$ versus $H_A : \theta > \theta_0$

- The analogous result holds when we want to test $H_0 : \theta \geq \theta_0$ versus $H_A : \theta < \theta_0$; then $g_{\theta_2}(t)/g_{\theta_1}(t)$ must be monotone non-increasing in t and the rejection region looks like $R = \{T < c_0\}$

The Neyman-Pearson Lemma: Examples

- Example 3.26: Show that the one-sided Z-test is a UMP level-α test.

The Neyman-Pearson Lemma: Examples

- Example 3.27: Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} \text{Poisson}(\lambda)$, where $\lambda > 0$. Explain how to produce a UMP level-α LRT for testing $H_0 : \lambda = \lambda_0$ versus $H_A : \lambda > \lambda_0$.

UMP Tests: Nonexistence

- Sadly, UMP tests usually don't always exist for a given pair of complementary hypotheses (especially for two-sided tests)

- Example 3.28: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ with $\mu \in \mathbb{R}$ and σ^2 known. Show there exists no UMP level-α test for $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$.