# 9.5: Some Comments on Selecting a Test Procedure

Taylor

University of Virginia

# Motivation

Now we'll talk a little about optimality of tests.

To state our first theorem we need this definition: A **simple hypothesis** is one in which, when true, completely specifies the distribution of the samples $X_i$s.

E.g. $H_a : \lambda = 2$ is simple. $H_a : \lambda > 2$ is not simple.

# Nota Bene

In this section we prefer to talk about *power* instead of type 2 error. The relationship is as follows:

$$\text{Power}(\theta) = 1 - \beta(\theta)$$

or

$$\text{Power} = 1 - \beta.$$

# The Neyman-Pearson Theorem

Suppose we are testing a simple null hypothesis versus a simple alternative hypothesis. $H_0 : \theta = \theta_0$ versus $H_a : \theta = \theta_a$. Let $k$ be a positive number. Let the rejection region for our test be

$$R_k^* = \left\{ (x_1, \ldots, x_n) : \frac{f(x_1, \ldots, x_n; \theta_a)}{f(x_1, \ldots, x_n; \theta_0)} \geq k \right\}.$$

After we select this there will be some type 1 error $\alpha^*$ and type 2 error $\beta^*$.

**Theorem**

*NPT: For any other test procedure with type 1 error probability $\alpha$ satisfying $\alpha \leq \alpha^*$, the probability of type two error $\beta$ will satisfy $\beta > \beta^*$.*

# The Neyman-Pearson Theorem

In other words: this is the most powerful/smallest type 2 error test we can get if our hypotheses are simple and we constrain our $\alpha$.

We'll extend this result to cases where our hypotheses can more closely resemble the ones we were using in the previous chapters.

# Example 9.20 on page 470

Consider randomly selecting $n = 5$ new vehicles of a certain type and determining the number of major defects on each one. Letting $X_i$ denote the number of such defects for the $i$th selected vehicle ($i = 1, \ldots, 5$), suppose that the $X_i$s form a random sample from a Poisson distribution with parameter $\lambda$. Let's find the best test for testing $H_0 : \lambda = 1$ versus $H_a : \lambda = 2$.

# Example 9.20 on page 470

$$\frac{f(x_1,\ldots,x_5;\theta_a)}{f(x_1,\ldots,x_5;\theta_0)} = \left[\frac{e^{-2\times 5}2^{\sum_i x_i}}{\prod_i x_i!}\right] \div \left[\frac{e^{-1\times 5}1^{\sum_i x_i}}{\prod_i x_i!}\right]$$
$$= \left[\frac{e^{-2\times 5}2^{\sum_i x_i}}{e^{-1\times 5}1^{\sum_i x_i}}\right]$$
$$= e^{-5}2^{\sum_i x_i}$$

So

$$R^* = \{(x_1,\ldots,x_5) : e^{-5}2^{\sum_i x_i} \geq k\}$$

# Example 9.20 on page 470

We haven't picked $k$ yet. We just have to pick it in accordance with our desired type 1 error.

$$
\begin{aligned}
& e^{-5} 2^{\sum_i x_i} \geq k \\
\iff & 2^{\sum_i x_i} \geq k' \\
\iff & \sum_i x_i \geq k''
\end{aligned}
$$

Instead of picking $k$, let's pick $k''$. It will be easier to choose because we can calculate probabilities with $\bar{X}$.

# Example 9.20 on page 470

So having a test with the rejection region

$$R^* = \left\{ (x_1, \ldots, x_5) : e^{-5} 2^{\sum_i x_i} \geq k \right\}$$

is the same as having one with this rejection region:

$$R^* = \left\{ (x_1, \ldots, x_5) : \sum_i x_i \geq c \right\}.$$

They both tell us to "reject when $\sum_i X_i$ is bigger than some constant." We pick $c$ to control the type 1 error $\alpha = .05$. So we pick a c such that

$$P\left( \sum_i X_i \geq c | \theta = \theta_0 \right) \leq \alpha = .05$$

We know $\sum_i X_i \sim$ Poisson(5) under the null, so $c = 10$. This gives us $\alpha = .032$.

# Example 9.20 on page 470

So our test is: "reject $H_0$ when $\sum_i x_i \geq 10$." This gives us an $\alpha = .032$. Let's find $\beta$

$$
\begin{aligned}
\beta &= \beta^*(\theta_a) \\
&= P\left(\sum_i X_i < 10 \text{ when } H_a \text{ is true}\right) \\
&= P\left(\sum_i X_i \leq 9 \,\middle|\, \sum_i X_i \sim \text{Poisson}(10)\right)
\end{aligned}
$$

# Example 9.21 on page 471

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1)$. Consider testing $H_0 : \mu = \mu_0$ versus $H_a : \mu = \mu_a$ where $\mu_a > \mu_0$. This is kind of like a right-tailed test except our hypotheses are simple.

Our likelihood ratio is

$$\frac{(\frac{1}{2\pi})^{n/2} \exp\left[-(1/2)\sum_i (x_i - \mu_a)^2\right]}{(\frac{1}{2\pi})^{n/2} \exp\left[-(1/2)\sum_i (x_i - \mu_0)^2\right]} = \left[e^{-n(\mu_a^2 - \mu_0^2)/2}\right] \cdot \left[e^{(\mu_a - \mu_0)\sum_i x_i}\right]$$

So we reject when $\sum_i X_i$ is big. Be able to show this; we spent a while on this in class.

# Proof of NPT

WLOG assume we're dealing with a discrete joint pmf.
A few preliminaries:

1. $R^* = \{\mathbf{x} : f(\mathbf{x}; \theta_a) \geq kf(\mathbf{x}; \theta_0)\}$
2. $R$ is the rejection region of some other test
3. $\alpha^* = P(R^* | \theta = \theta_0)$
4. $\alpha = P(R | \theta = \theta_0)$
5. $1 - \beta^* = P(R^* | \theta = \theta_a)$
6. $1 - \beta = P(R | \theta = \theta_a)$

Keep in mind that

$$\left\{ \frac{f(\mathbf{x}; \theta_a)}{f(\mathbf{x}; \theta_0)} \geq k \right\} \Leftrightarrow \{f(\mathbf{x}; \theta_a) \geq kf(\mathbf{x}; \theta_0)\} \Leftrightarrow \{f(\mathbf{x}; \theta_a) - kf(\mathbf{x}; \theta_0) \geq 0\}$$

# Proof of NPT

WTS $\beta > \beta^*$. I think this proof is easier to follow backwards:

$$\begin{aligned}
\beta - \beta^* &\geq \beta - \beta^* - k(\alpha^* - \alpha) \\
&= (1 - \beta^*) - (1 - \beta) - k\alpha^* + k\alpha \\
&= P(R^*|\theta = \theta_a) - P(R|\theta = \theta_a) - kP(R^*|\theta = \theta_0) + kP(R|\theta = \theta_0) \\
&= [P(R^*|\theta = \theta_a) - kP(R^*|\theta = \theta_0)] - [P(R|\theta = \theta_a) - kP(R|\theta = \theta_0)]
\end{aligned}$$

# Proof of NPT

Let's break down the two differences a bit more...

$$
\begin{aligned}
P(R^*|\theta = \theta_a) - kP(R^*|\theta = \theta_0) = P(R^* \cap R|\theta = \theta_a) + P(R^* \cap R'|\theta = \theta_a) \\
- kP(R^* \cap R|\theta = \theta_0) - kP(R^* \cap R'|\theta = \theta_0)
\end{aligned}
$$

$$
\begin{aligned}
P(R|\theta = \theta_a) - kP(R|\theta = \theta_0) = P(R \cap R^*|\theta = \theta_a) + P(R \cap R^{*'}|\theta = \theta_a) \\
- kP(R \cap R^*|\theta = \theta_0) - kP(R \cap R^{*'}|\theta = \theta_0)
\end{aligned}
$$

so the difference between these two guys is

$$
P(R^* \cap R'|\theta = \theta_a) - P(R \cap R^{*'}|\theta = \theta_a) - kP(R^* \cap R'|\theta = \theta_0) + kP(R \cap R^{*'}|\theta = \theta_0
$$

# Proof of NPT

going back to our proof...

$$
\begin{aligned}
\beta - \beta^* &\geq \beta - \beta^* - k(\alpha^* - \alpha) \\
&= (1 - \beta^*) - (1 - \beta) - k\alpha^* + k\alpha \\
&= P(R^*|\theta = \theta_a) - P(R|\theta = \theta_a) - kP(R^*|\theta = \theta_0) + kP(R|\theta = \theta_0) \\
&= [P(R^*|\theta = \theta_a) - kP(R^*|\theta = \theta_0)] - [P(R|\theta = \theta_a) - kP(R|\theta = \theta_0)] \\
&= P(R^* \cap R'|\theta = \theta_a) - kP(R^* \cap R'|\theta = \theta_0) \\
&- \left[P(R \cap R^{*'}|\theta = \theta_a) - kP(R \cap R^{*'}|\theta = \theta_0)\right]
\end{aligned}
$$

# Proof of NPT

For concision, denote $h(\mathbf{x}) = f(\mathbf{x};\theta_a) - kf(\mathbf{x};\theta_0)$

On $R^*$, $h \geq 0$. On $R^{*'}$, $h < 0$. So

$$P(R^* \cap R' | \theta = \theta_a) - kP(R^* \cap R' | \theta = \theta_0) = \sum_{R^* \cap R'} h(\mathbf{x}) \geq 0$$

$$P(R \cap R^{*'} | \theta = \theta_a) - kP(R \cap R^{*'} | \theta = \theta_0) = \sum_{R \cap R^{*'}} h(\mathbf{x}) < 0$$

# Proof of NPT

going back to our proof one more time

$$
\begin{aligned}
\beta - \beta^* &\geq \beta - \beta^* - k(\alpha^* - \alpha) \\
&= (1 - \beta^*) - (1 - \beta) - k\alpha^* + k\alpha \\
&= P(R^*|\theta = \theta_a) - P(R|\theta = \theta_a) - kP(R^*|\theta = \theta_0) + kP(R|\theta = \theta_0) \\
&= [P(R^*|\theta = \theta_a) - kP(R^*|\theta = \theta_0)] - [P(R|\theta = \theta_a) - kP(R|\theta = \theta_0)] \\
&= P(R^* \cap R'|\theta = \theta_a) - kP(R^* \cap R'|\theta = \theta_0) \\
&- \left[P(R \cap R^{*'}|\theta = \theta_a) - kP(R \cap R^{*'}|\theta = \theta_0)\right] \\
&> 0
\end{aligned}
$$

# NPT

Isn't it more realistic to consider tests of the form, say, $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$?

Yes, but we can still use NPT here.

First, what do we mean by the "best" test?

# Uniformly Most Powerful Test: the definition

The "U" in uniformly needs some explaining. Instead of $\beta > \beta^*$, we want

$$\beta(\theta) > \beta^*(\theta)$$

for any $\theta$ that work with $H_a$, or in other words, for all $\theta \in \Theta_a$.

Note: $\Theta_a$ is the set of all parameters that work with $H_a$. For example, if our test was $H_0 : \mu = 2$ versus $H_a : \mu > 2$, we would have $\Theta_a = (2, \infty)$

# Uniformly Most Powerful Test: the definition

A **uniformly most powerful level $\alpha$ test**, is one for which the power is maximized for any $\theta$ that satisfies $H_a$.

In other words:

**Theorem**

*A test with $\alpha^*$ and $\beta^*(\theta)$ is UMP if: for any other test with $\alpha$ and $\beta(\theta)$, $\alpha \leq \alpha^*$ implies $1 - \beta(\theta) < 1 - \beta^*(\theta)$ for all $\theta \in \Theta_a$.*

# UMP Tests

UMP tests do not always exist. One of the goals of this class is to practice identifying when we can use NPT (a theorem testing two simple hypotheses against each other) to identify UMP tests in certain situations with more complicated hypotheses. This why NPT is very important; it's a simple theorem that has a lot of use in identifying something very important (UMP tests) that don't always exist.

# Using NPT to find UMP Tests

We just used the NPT to find tests for hypotheses like these: $H_0 : \theta = \theta_0$ versus $H_a : \theta = \theta_a$. WLOG let's assume $\theta_a > \theta_0$. We write out that likelihood ratio, that gives us a test statistic, then we pick a $k$ based on $\alpha$ and $\theta_0$.

What about hypotheses like these: $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$? $H_a$ is **composite** here. NPT doesn't apply *directly* here.

What we could do is pick a particular $\theta_a \in \Theta_a$. For this parameter we could apply NPT by writing out that likelihood ratio, then this gives us a test statistic, and then we pick a cutoff $k$ based on $\alpha$. Does this help us?

The answer is sometimes: if we get the same test (test stat. and cutoff) for any $\theta_a$, then we're done. Think about that for a second: the definition of a UMP test is that it's most powerful for *all* $\theta_a \in \Theta_a$. Even though NPT's statement only mentions simple versus simple tests, it can still be used sometimes to find a UMP simple versus composite test.

# Using NPT to find UMP Tests

Here's an example where this works. Let's go back to the example where $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$. $H_0 : \lambda = \lambda_0$ versus $H_a : \lambda < \lambda_0$. Pick any $\lambda_{alt} < \lambda_0$. Using the same math as on slide 6,
$\frac{f(\mathbf{x};\lambda_{alt})}{f(\mathbf{x};\lambda_0)} = e^{-n(\lambda_{alt}-\lambda_0)} \left(\frac{\lambda_{alt}}{\lambda_0}\right)^{\sum X_i}$.

So for any $\lambda_{alt}$ we pick that's possible, we get the same test of the form "reject when $\sum X_i \leq c$." It's the same $c$ every time because $c$ is chosen using $\lambda_0$ and the inequality doesn't flip different ways for different $\lambda_{\text{alt}}$s.

# NPT

NPT gives us UMP tests for one sided alternative hypotheses when we're testing the mean of normal data. However it doesn't give us a UMP test for a two-sided alternative. It also sometimes fails us for when we're testing hypotheses about $\mu$ AND $\sigma^2$.

# Likelihood Ratio Tests

An alternative procedure: LRTs. They look similar, but they are not the same.

Instead of

$$\frac{f(x_1,\ldots,x_n;\theta_a)}{f(x_1,\ldots,x_n;\theta_0)} > k \iff \frac{f(x_1,\ldots,x_n;\theta_0)}{f(x_1,\ldots,x_n;\theta_a)} < 1/k$$

we'll have

$$\frac{\sup_{\theta\in\Theta_0} f(x_1,\ldots,x_n;\theta)}{\sup_{\theta\in\Theta} f(x_1,\ldots,x_n;\theta)} \leq k$$

where $\Theta = \Theta_0 \cup \Theta_a$.

## LRTs

Reject $H_0$ if

$$\frac{\sup_{\theta\in\Theta_0} f(x_1,\ldots,x_n;\theta)}{\sup_{\theta\in\Theta} f(x_1,\ldots,x_n;\theta)} \leq k$$

where $\Theta = \Theta_0 \cup \Theta_a$, and $k$ is chosen to be such that the probability of type 1 error is $\alpha$.

# Likelihood Ratio Tests

How do we find

$$\frac{\sup_{\theta \in \Theta_0} f(x_1, \ldots, x_n; \theta)}{\sup_{\theta \in \Theta} f(x_1, \ldots, x_n; \theta)}?$$

we find the denominator by finding the maximum likelihood estimate $\hat{\theta}$ and plugging it into the likelihood function. This is the invariance principle of MLEs!

We find the numerator by finding another maximum likelihood estimate, but now the parameter space is constrained. We plug this $\hat{\theta}_0$ into our likelihood to get the numerator.

# Example 9.24 on page 475

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$. Let's test $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$. We could also write this as $H_0 : \mu = \mu_0$ and $\sigma^2 > 0$ versus $H_a : \mu \neq \mu_0$ and $\sigma^2 > 0$.

# Example 9.24 on page 475

The denominator is always easier to find. We showed earlier that our MLE estimates for normal data were $\hat{\mu} = \bar{X}$ and $\hat{\sigma^2} = \frac{\sum_i (x_i - \bar{x})^2}{n}$. Our likelihood is $(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{\sum_i (x_i - \mu)^2}{2\sigma^2}\right]$. So we just plug $\hat{\mu}$ and $\hat{\sigma^2}$ in for $\mu$ and $\sigma^2$, respectively.

With the numerator, we don't have to estimate $\mu$ since it is given to us as $\mu_0$. Now the constrained likelihood is $(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{\sum_i (x_i - \mu_0)^2}{2\sigma^2}\right]$. If we maximize that with respect to $\sigma^2$, we get $\hat{\sigma^2}_0 = \frac{\sum_i (x_i - \mu_0)^2}{n}$.

# Example 9.24 on page 475

Finally, we plug in our likelihoods. Our rule is reject $H_0$ when

$$\left(\frac{\sum_i (x_i - \bar{x})^2}{\sum_i (x_i - \mu_0)^2}\right)^{n/2} \leq k.$$

With a bit of algebra, we can show that this is the same as rejecting when

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} \geq c \text{ or } \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \leq -c$$

which is that 2-tailed t test we learned about already. (Verifying this is also a quiz question).

# Likelihood Ratio Test: a convenient result

Another convenient result:

$$-2\log\left(\frac{\sup_{\theta\in\Theta_0} f(x_1,\ldots,x_n;\theta)}{\sup_{\theta\in\Theta} f(x_1,\ldots,x_n;\theta)}\right) \overset{\text{approx.}}{\sim} \chi^2_\nu$$

where $\nu$ is the number of parameters involved in the null hypothesis (often 1 in this class).

This means that, if our data set is large, then we don't need to do any algebra to simplify the expression enough to find a null distribution.