Statistical Inference

Lecture 07b

ANU - RSFAS

Last Updated: Wed Apr 18 13:29:40 2018

Hypothesis Testing

- So far we have focused on statistical point estimation.
- In many situations, however, the simple estimation of a population characteristic is not the final desired outcome of a statistical analysis.
- We may want to use our estimates to decide whether some previously proposed theory or statement regarding the population of interest is actually true (or at least is plausible given the information provided by the observations at hand).
- This is, of course, the standard framework of statistical hypothesis testing which is familiar from any introductory unit in basic statistics.

Hypothesis Testing

Steve Stern's Notes

- Consider the following situation:
  - Suppose that we have purchased a light-bulb based on its advertised claim that the mean lifetime of such bulbs is at least 1000 hours.
  - If we then observe the lifetime of the actual bulb we purchased, we have some data with which to assess the advertising claim.
  - This simple scenario is precisely the framework of statistical hypothesis testing.
  - Suppose we believe that the lifetime of the population of bulbs in question is exponentially distributed with mean parameter θ

$$p(x;\theta) = \frac{1}{\theta} exp\left(-\frac{x}{\theta}\right) \quad \text{for } \theta \in \Theta$$

$E(x) = \theta$

Hypothesis Testing

- We can formulate a hypothesis test as: $H_0 : \theta \geq 1000$
- **Definitions:** Suppose that $X_1, \ldots, X_n$ represent a simple random sample from a parametric family with density function $f(x; \theta)$ for some parameter $\theta \in \Theta$.
  - A statistical hypothesis is simply a subset of the parameter space, Θ.
  - Any statistical hypothesis of interest, often termed the null hypothesis, is associated with a competing alternative hypothesis.
  - A null hypothesis and its alternative form a partition of the parameter space Θ consisting of the sets Θ_0 and

$$\Theta_1 = \Theta_0^c \cap \Theta$$

Hypothesis Testing

Definition: A hypothesis testing procedure or hypothesis test is a rule that specifies:

1. For which sample values the decision is made to accept H_0.
2. For which sample values H_0 is rejected and H_1 is accepted as true.

The subset of the space for which H_0 will be rejected is called the rejection region or critical region (C). The complement of the rejection region is called the acceptance region.

Hypothesis Testing

- For our light bulb example:
  - We can define a test which rejects H_0 if X is less than 1,000 hours.
  - $C = \{X < 1000\}$
- More generally, we can define a statistical test in terms of a rejection region (C) which is just a set for some statistic $T(X_1, \dots, X_n)$:

$$C = \{X \in \mathcal{X} : T(X) < k\}$$

Type I and Type II Errors

- Common sense would indicate that the test described in the example of the previous section; namely, rejecting the null hypothesis that the mean lifetime of the bulbs is at least 1000 hours based on a single observation being less than 1000 hours, is not a very good test.
- We will make errors.
- Consider the following possibilities:
  - Type I Error: Reject H_0 given that it is true. Thus the observations fall in the rejection region C when in fact that null hypothesis, H_0, is true.
  - Type II Error: Do not Reject H_0 when it is false: Thus the observed data values fall outside the rejection region when in fact the null hypothesis is false.

Type I and Type II Errors

	Decision	
Truth	Accept H_0	Accept H_1
H_0	Correct Decision	Type I Error
H_1	Type II Error	Correct Decision

Type I and Type II Errors

- Probability of a Type I error (α): $H_0: \theta \geq 1000$

$$
\begin{aligned}
P(C) &= P_{1000}(X < 1000 | H_0 \text{ is true }) \quad \theta = 1000 \\
&= \int_0^{1000} \frac{1}{1000} exp\left(-\frac{x}{1000}\right) dx \\
&= 1 - exp(-1000/1000) = 0.632 = \alpha.
\end{aligned}
$$

- What if $\theta = 1500$:

$$
\begin{aligned}
P(C) &= P_{1500}(X < 1000 | H_0 \text{ is true }) \\
&= \int_0^{1000} \frac{1}{1500} exp\left(-\frac{x}{1500}\right) dx \\
&= 1 - exp(-1000/1500) = 0.077 = \alpha
\end{aligned}
$$

Type I and Type II Errors

- Let's determine the probability of a Type II error.
- Note that we specified $H_0 : \theta \geq 1000$. This means:

$$H_1 : \theta < 1000$$

- Picking a specific value in this region ($\theta = 500$), we have:

$$\begin{aligned} P(C^c) &= P_{500}(X > 1000 | H_0 \text{ is false }) \\ &= \int_{1000}^{\infty} \frac{1}{500} exp\left(-\frac{x}{500}\right) dx \\ &= exp(-1000/500) = 0.135 \end{aligned}$$

$= P(\text{Type II error})$

- There is a strong relationship between Type I and Type II errors. Note that for a given value of θ, only one type of error can occur (since for any given θ, H_0 either is or is not true).

Type I and Type II Errors

Definition 4.2: The probability of a Type I error, α in a test of hpotheses is called the **size** or **signifigance level** of the test. The complement of the probability of a Type II error

$$\eta(\theta) = 1 - \beta, \quad \leftarrow \text{power of the test}$$

is the **power** of the test.

- **Power = 1 - P(Type II Error)**
  $= 1 - P(\boldsymbol{X} \in C^c | H_1 \text{ is true}) = P(\boldsymbol{X} \in C | H_1 \text{ is true})$
- Given that H_1 is true, what is the probability I reject H_0

Type I and Type II Errors

- Consider the light bulb example:

$$\eta(\theta) = P(X \in C) = P(X < 1000) = 1 - exp(-1000/\theta)$$

- The size of the test determined by $C = \{X < 1000\}$.
- Again recall: $H_0 : \theta \geq 1000$.
- The power function is a decreasing function of θ in this case. So to maximize it we set $\theta = 1000$.

$$\begin{aligned} max_{\theta \in \Theta_0} \eta(\theta) &= max_{\theta \geq 1000} 1 - exp(-1000/\theta) \\ &= 1 - exp(-1000/1000) = 0.632 = \alpha \end{aligned}$$

Type I and Type II Errors

- It is standard to focus on test which have sizes 0.05 or 0.01.
- If we focus on tests with rejection regions of the form $C = \{X < k_\alpha\}$, we can choose k_α such that:

maximize the power when $\theta \in \Theta_0$

$$\begin{aligned} max_{\theta \in \Theta_0}\ \eta(\theta) &= max_{\theta \geq 1000} 1 - exp(-k_\alpha/\theta) \\ &= 1 - exp(-k_\alpha/1000) = \alpha \end{aligned}$$

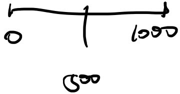
Based on this $k_\alpha = -1000 ln(1-\alpha)$ so at $\alpha = 0.05$ we have:

$$C = \{X < 51.29\}$$

Type I and Type II Errors

- What actually is the power based on this rejection region if truly $\theta = 500 \in \Theta_1$?

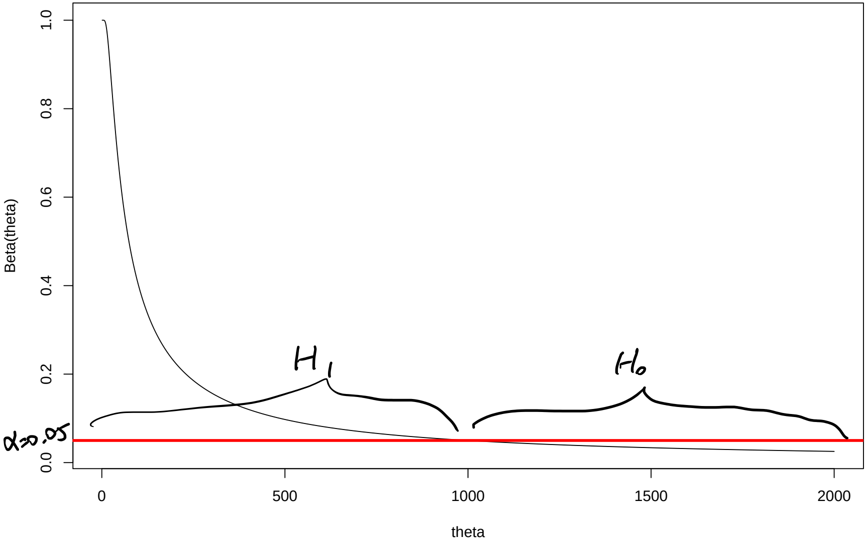
$H_0 : \theta \geq 1000$, $H_1 : \theta < 1000$



$$\begin{aligned}
\eta(\overset{\theta}{\overset{\shortparallel}{500}}) &= P_{500}(X \in C) \\
&= P_{500}(X < 51.29) \\
&= \int_0^{51.29} \frac{1}{500} exp(-x/500) dx \\
&= 1 - exp(-51.29/500) = 0.0975.
\end{aligned}$$

- So this test has less than a 10% chance of detecting even this drastic departure from the null hypothesis based on $\alpha = 0.05$!!

The Power Function is a Function! $\eta(\theta)$



- Red $= 0.05$

Type I and Type II Errors

- Unfortunately, if our power is not as large as we like, we cannot simply change a rejection region of the form $C = \{X < k\}$ to increase the power without simultaneously affecting the size of our test (α).
- Our task, then, is to find tests (or equivalently rejection regions) of a given size which have the best possible power when $\theta \in \Theta_1$.
- We can increase our sample size (not always possible) or by finding a "good" test statistic.

Essential Nature of a Hypothesis Test (Experimental Design, Hoff 2009)

- Given H_0, H_1 and data $\boldsymbol{x} = \{x_1, \ldots, x_n\}$:

1. From the data, compute a relevant test statistic $T(\boldsymbol{x})$: The test statistic $T(\boldsymbol{x})$ should be chosen so that it can differentiate between H_0 and H_1 in ways that are scientifically relevant. Typically, $T(\boldsymbol{x})$ is chosen so that

$$T(\boldsymbol{x}) \text{ is probably } \begin{cases} \text{small under } H_0 \\ \text{large under } H_1 \end{cases}$$

2. Obtain a null distribution: A probability distribution over the possible outcomes of $T(\boldsymbol{X})$ under H_0. Here, $X = \{X_1, \dots, X_n\}$ are potential experimental results that could have happened under H_0.

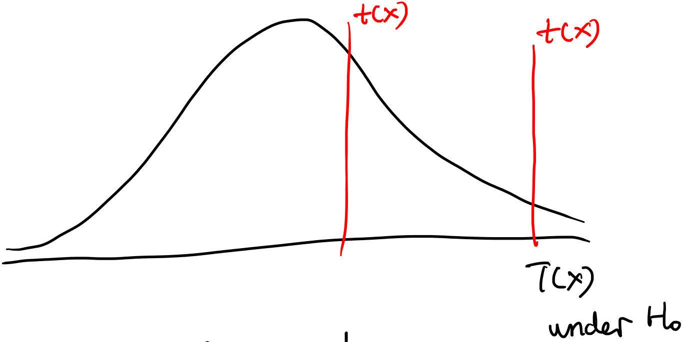
3. Compute the p-value: The probability under H_0 of observing a test statistic $T(\boldsymbol{X})$ as or more extreme than the observed statistic $t(\boldsymbol{x})$.

$$\text{p-value} = P(T(\boldsymbol{X}) \geq t(\boldsymbol{x}) | H_0)$$

observed test statistic

If the p-value is small $\Rightarrow$ evidence against H_0
If the p-value is large $\Rightarrow$ not evidence against H_0

- See pg 77 for a discussion of p-values (and the ASA discussion posted on Wattle).



$P(T(x) \mid H_0)$ = p-value

vs

$P(H_0 \mid T(x))$ = ?

p-values

Example: Suppose that $X_1, \ldots, X_n$ are a random sample from a normal distribution with mean μ and unit variance. Consider testing:

when Ho is true

$$\begin{aligned} H_0 &: \quad \mu \leq \mu_0 \\ H_1 &: \quad \mu > \mu_0 \end{aligned}$$

- We can show that that $C = \left\{ \frac{\bar{X} - \mu_0}{1/\sqrt{n}} \geq k \right\}$. So

$$Z \downarrow \qquad T(\boldsymbol{X}) = \frac{\bar{X} - \mu_0}{1/\sqrt{n}}$$

observed

$$\text{p-value} = P\left(\frac{\bar{X} - \mu_0}{1/\sqrt{n}} \geq \frac{\bar{x} - \mu_0}{1/\sqrt{n}} \right) = P\left(Z \geq \frac{\bar{x} - \mu_0}{1/\sqrt{n}} \right)$$

- The probability, under H_0, of getting the observed test statistic or something more extreme (based on the rejection region).

Neyman-Pearson Set-up

- Consider simple hypotheses - those which consist of only a single parameter value.
- We will examine the case of a statistical test for which both the null and alternative hypotheses are simple.
- Suppose that $X_1, \ldots, X_n$ are a sample from a population characterized by a probability model with density function $f(x;\theta)$ for $\theta \in \Theta$ where $\Theta = \{\theta_0, \theta_1\}$.
- We shall focus on:

$$\begin{aligned} H_0: &\quad \theta = \theta_0 \\ H_1: &\quad \theta = \theta_1 \end{aligned}$$

Neyman-Pearson Lemma

- Consider the likelihood-ratio:

$$\frac{L(\theta_1; x)}{L(\theta_0; x)}$$

$$\lambda(\boldsymbol{x}) = \frac{L(\theta_0; \boldsymbol{x})}{L(\theta_1; \boldsymbol{x})}$$

suggests θ_0 maximizes the likelihood more than θ_1

- The test we shall define has a critical region of the form

$$C = \{\lambda(\boldsymbol{x}) \leq k\}$$

- The ratio of the likelihood for any given sample at each of the two possible parameter values is precisely a relative measure of how plausible the two hypotheses are.
- In other words, when $\lambda(\boldsymbol{x})$ is very small, this is strong evidence that the observations arose from the alternative hypothesis rather than the null hypothesis.

Neyman-Pearson Set-up

- It should seem intuitively reasonable that the likelihood ratio is a good method of distinguishing between samples which support the null hypothesis versus samples which support the alternative hypothesis.
- From what we have done, we know for a given α we could compare the power $\eta(\theta)$. $\alpha \to k \to \eta(\theta)$
- We would like to find a uniformily most powerful test . . .

$$\eta(\theta) \geq \eta(\theta^*)$$

- It turns out that N-P tests lead to UMP tests.

Example Suppose that $X_1, \ldots, X_n$ are a random sample from a normal distribution with mean μ and unit variance. Further, suppose that we know $\mu \in \{0, 1\}$. We wish to test:

$$
\begin{aligned}
H_0 : &\qquad \mu = 0 \\
H_1 : &\qquad \mu = 1
\end{aligned}
$$

$$\lambda(\vec{x}) = \frac{L(\theta_0)}{L(\theta_1)}$$

$\theta_0 = 0$

$\theta_1 = 1$

$$\begin{aligned}
\lambda(\boldsymbol{x}) &= \frac{exp\left(-\frac{1}{2}\sum_{i=1}^{n} X_i^2\right)}{exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i - 1)^2\right)} \\
&= exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left[X_i^2 - (X_i - 1)^2\right]\right) \\
&= exp\left(\frac{n}{2} - \sum_{i=1}^{n} X_i\right)
\end{aligned}$$

- So we get the rejection region: $C = \{\lambda(\vec{x}) \leq k\}$

$$
\begin{aligned}
C &= \left\{ exp\left(\frac{n}{2} - \sum_{i=1}^{n} X_i\right) \leq k \right\} \\
&= \left\{ \frac{n}{2} - \sum_{i=1}^{n} X_i \leq log(k) \right\} \\
&= \left\{ -\sum_{i=1}^{n} X_i \leq log(k) - \frac{n}{2} \right\} \\
&= \left\{ \sum_{i=1}^{n} X_i \geq -log(k) + \frac{n}{2} \right\} \\
&= \left\{ \bar{X} \geq -log(k)/n + \frac{1}{2} \right\} \\
&= \left\{ \bar{X} \geq k^* \right\}
\end{aligned}
$$

$\theta_0 = 0$

$$
\begin{aligned}
P_{H_0}(C) &= P_{H_0}\left(\bar{X} \geq k^*\right) = \alpha \\
&= P_{H_0}\left(\frac{\bar{X} - 0}{1/\sqrt{n}} \geq k^{**}\right) = \alpha \\
&= P_{H_0}\left(Z \geq k^{**}\right) = \alpha \quad = 0.05
\end{aligned}
$$

- If $\alpha = 0.05$ then c^{**} is 1.644854.

$\alpha = 0.05$

```
qnorm(0.95)
```

$P\left(\frac{\bar{X}-0}{1/\sqrt{n}} \geq 1.645\right)$

```
## [1] 1.644854
```

then reject H_0

- This is a UMP test!

$P(X=1) = \theta$

Example Suppose that $X_1, \ldots, X_{10}$ are a random sample from a Bernoulli distribution with parameter θ. Further, suppose that we wish to test:

$$\begin{aligned} H_0 : \quad & \theta = 0.5 \\ H_1 : \quad & \theta = 0.2 \end{aligned}$$

- Let's get the likelihood:

$$L(\theta; \boldsymbol{x}) = \theta^{\sum x_i}(1-\theta)^{10-\sum x_i} = \theta^{10\bar{x}}(1-\theta)^{10-10\bar{x}}$$

$$\lambda(\boldsymbol{x}) = \frac{0.5^{10\bar{x}}(1-0.5)^{10-10\bar{x}}}{0.2^{10\bar{x}}(1-0.2)^{10-10\bar{x}}}$$
$$= \left(\frac{5}{8}\right)^{10} 4^{10\bar{x}}$$

- So we get the rejection region:

$$C = \left\{\left(\frac{5}{8}\right)^{10} 4^{10\bar{x}} \leq k\right\}$$
$$= \left\{10\bar{x} \leq log_4\left[\left(\frac{8}{5}\right)^{10} k\right]\right\}$$
$$= \{10\bar{x} \leq k^*\}$$

$\{n\bar{x} \leq k^*\} = \{\sum x_i \leq k^*\}$

- Let's get a UMP test for $\alpha = 0.01$.

$$
\begin{aligned}
P_{H_0}(C) &= P(10\bar{X} \le k^*) = 0.01 \\
&= P\left(\sum_{i=1}^{n} X_i \le k^*\right) = 0.01
\end{aligned}
$$

- Recall that under H_0: $\sum_{i=1}^{n} X_i \sim \mathrm{binomial}(n = 10, p = 0.5)$.
- Due to the discreteness, we can't find a k^* such that we achieve $\alpha = 0.01$.

```r
qbinom(0.01, 10, 0.5)
```

n θ_0

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

- The closest we can find is $k^* = 1$.

```r
pbinom(1, 10, 0.5)
```

```
## [1] 0.01074219
```

- So we have a UMP test of size $\alpha = 0.01074$, which is close to $\alpha = 0.01$.

$$P_{H_0}(C) = P(10\bar{X} \leq 1) = 0.01074$$

Neyman-Pearson Lemma

Section 4.2:

- Suppose that H_0 and H_1 are simple hypotheses and that the test that rejects H_0 whenever the likelihood ratio is less than k has significance level α.
- **Lemma 4.2:** Then any other test for which the significance level is less than or equal to α has power less than or equal to that of the likelihood ratio test.

Proof: The proof is on pg. 73, but is not that enlightening so I won't present it, but is interesting in terms of playing with sets.

Neyman-Pearson Lemma - Section 4.4

- On the surface, it seems the N-P Lemma is too simple to be of any real use. Can we push the result a bit? Let's consider the example.

Example: Suppose that $X_1, \ldots, X_n$ are a random sample from a normal distribution with mean μ and unit variance. Consider testing:

$$\begin{aligned} H_0: \quad & \mu = \mu_0 = 0 \\ H_1: \quad & \mu = \mu_1 \end{aligned}$$

Where $\mu_1 > 0 = \mu_0$.

Neyman-Pearson Lemma

$$\lambda(\boldsymbol{x}) = \frac{exp\left(-\frac{1}{2}\sum_{i=1}^{n} X_i^2\right)}{exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu_1)^2\right)}$$

$$= exp\left(\frac{n\mu_1^2}{2} - n\mu_1\bar{X}\right)$$

Neyman-Pearson Lemma

$\mu_1 > 0 ; n > 0$

$\therefore -n\mu_1 < 0$

- So we get the rejection region:

$$
\begin{aligned}
C &= \left\{ exp\left(\frac{n\mu_1^2}{2} - n\mu_1\bar{X}\right) \leq k \right\} \\
&= \left\{ \left(\frac{n\mu_1^2}{2} - n\mu_1\bar{X}\right) \leq log(k) \right\} \\
&= \left\{ \bar{X} > \frac{\mu_1}{2} - \frac{1}{n\mu_1} log(k) \right\} \\
&= \left\{ \bar{X} > k^* \right\} \\
&= \left\{ \frac{\bar{X} - 0}{1/\sqrt{n}} \geq k^{**} \right\} = \{ Z \geq k^{**} \}
\end{aligned}
$$

under H_0

- We assumed that $\mu_1 > 0$, so we get the sign switch.
- If $\alpha = 0.05$ then $k^{**} = 1.64$.

Neyman-Pearson Lemma

- The UMP test has the same rejection region as our previous example: $H_0 : \mu = 0$ vs $H_1 : \mu = 1$.
- This test is actually UMP for $H_0 : \mu = 0$ vs $H_1 : \mu > 0$.
- It can also be shown that the test is UMP for $H_0 : \mu \leq 0$ vs $H_1 : \mu > 0$.

Neyman-Pearson Lemma

- What if we wanted to test: $H_0 : \mu = 0$ vs $H_1 : \mu < 0$?
- We get a UMP test with rejection region:
- So we get the rejection region:

$$
\begin{aligned}
C &= \left\{ exp\left(\frac{n\mu_1^2}{2} - n\mu_1\bar{X}\right) \leq k \right\} \\
&= \left\{ \left(\frac{n\mu_1^2}{2} - n\mu_1\bar{X}\right) \leq log(k) \right\} \\
&= \left\{ \bar{X} \leq \frac{\mu_1}{2} - \frac{1}{n\mu_1} log(k) \right\} \\
&= \left\{ \bar{X} \leq k^* \right\} \\
&= \left\{ \frac{\bar{X} - 0}{1/\sqrt{n}} \leq k^{**} \right\} = \{Z \leq k^{**}\}
\end{aligned}
$$

Neyman-Pearson Lemma

* How does that compare to a Maximum Likleihood Ratio Test (Generalized Likelihood Ratio Test) [an extension we will discuss shortly]? For:

$$H_0: \quad \mu = \mu_0$$
$$H_1: \quad \mu \neq \mu_0$$

- Let's have $\mu_0 = 0$. We will show the rejection region is:

$$\left\{|Z| > \sqrt{n}\sqrt{[-2log(c)]/n}\right\} = \{|Z| > k^*\}$$

- So we will reject H_0 if:

$$\left\{\left|\frac{(\bar{x} - 0)}{1/\sqrt{n}}\right| > 1.96\right\} \quad \checkmark$$

Neyman-Pearson Lemma

- Let's plot the power for the three tests for $n = 10, \mu_0 = 0, \alpha = 0.05$:

1. $H_0 : \mu = 0$ vs $H_1 : \mu > 0$

under H_0

$\Rightarrow \alpha = 0.05$

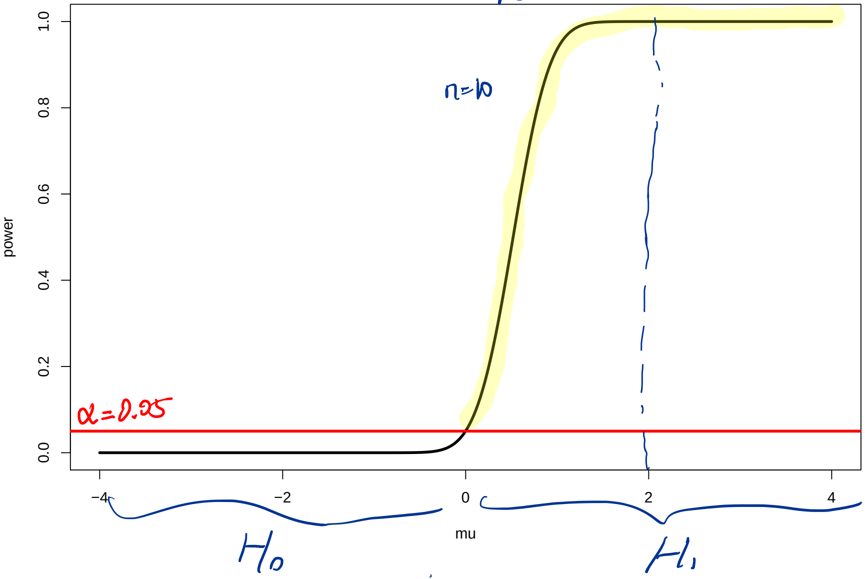
under H_1

$$\begin{aligned}
\eta(\mu) &= P\left(\frac{\bar{X} - \mu_0}{1/\sqrt{n}} \geq 1.64\right) \\
&= P\left(\frac{\bar{X} - \mu}{1/\sqrt{n}} - \frac{\mu_0 - \mu}{1/\sqrt{n}} \geq 1.64\right) \\
&= P\left(\frac{\bar{X} - \mu}{1/\sqrt{n}} \geq 1.64 + \frac{\mu_0 - \mu}{1/\sqrt{n}}\right) \\
&= P\left(Z \geq 1.64 - \frac{\mu}{1/\sqrt{n}}\right) = 1 - P\left(Z < 1.64 - \sqrt{n}\mu\right)
\end{aligned}$$

$\mu_0 = 0$

```r
mu <- seq(-4,4, by=0.01)
n <- 10
power <- 1 - pnorm(1.64 - sqrt(n)*mu)
plot(mu, power, lwd=3, type="l")
abline(h=0.05, lwd=3, col="red")
```

Power Curve



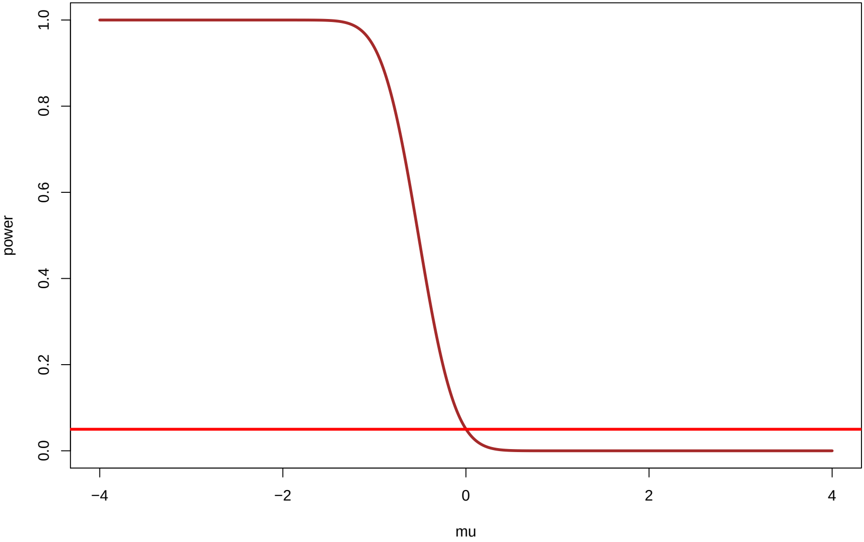
Neyman-Pearson Lemma

- Let's plot the power for the three tests for $n = 10, \mu_0 = 0, \alpha = 0.05$:

2. $H_0 : \mu = 0$ vs $H_1 : \mu < 0$

$$
\begin{aligned}
\eta(\mu) &= P\left(\frac{\bar{X} - \mu_0}{1/\sqrt{n}} \leq -1.64\right) \\
&= P\left(\frac{\bar{X} - \mu}{1/\sqrt{n}} - \frac{\mu_0 - \mu}{1/\sqrt{n}} \leq -1.64\right) \\
&= P\left(\frac{\bar{X} - \mu}{1/\sqrt{n}} \leq -1.64 + \frac{\mu_0 - \mu}{1/\sqrt{n}}\right) \\
&= P\left(Z \leq -1.64 - \sqrt{n}\mu\right)
\end{aligned}
$$

```r
mu <- seq(-4,4, by=0.01)
n <- 10
power <- pnorm(-1.64 - sqrt(n)*mu)
plot(mu, power, lwd=3, type="l", col="brown")
abline(h=0.05, lwd=3, col="red")
```



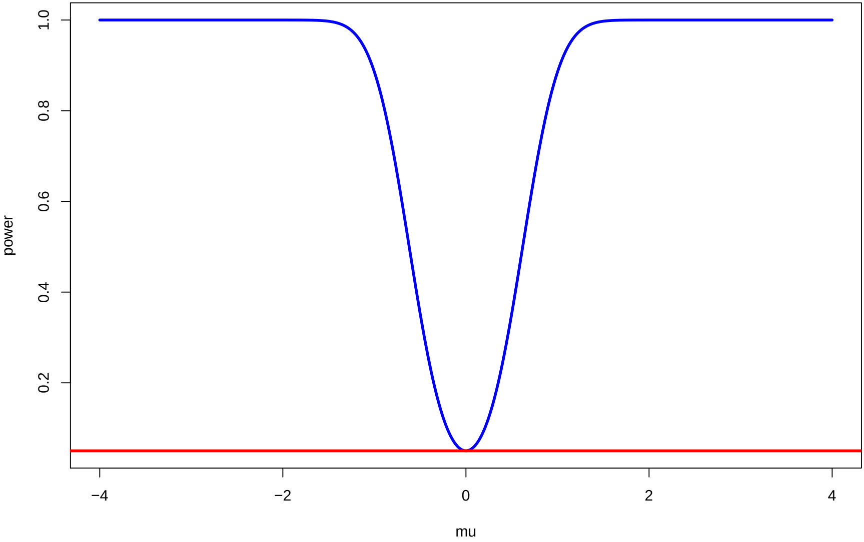
power
1.0
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4
mu

Neyman-Pearson Lemma

3. $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$

$$
\begin{aligned}
\eta(\mu) &= P\left(\frac{\bar{X}-\mu_0}{1/\sqrt{n}} \geq 1.96\right) + P\left(\frac{\bar{X}-\mu_0}{1/\sqrt{n}} \leq -1.96\right) \\
&= P\left(\frac{\bar{X}-\mu}{1/\sqrt{n}} \geq 1.96 + \frac{\mu_0-\mu}{1/\sqrt{n}}\right) + P\left(\frac{\bar{X}-\mu}{1/\sqrt{n}} \leq -1.96 + \frac{\mu_0-\mu}{1/\sqrt{n}}\right) \\
&= P\left(Z \geq 1.96 + \frac{\mu_0-\mu}{1/\sqrt{n}}\right) + P\left(Z \leq -1.96 + \frac{\mu_0-\mu}{1/\sqrt{n}}\right) \\
&= 1 - P\left(Z < 1.96 + \frac{\mu_0-\mu}{1/\sqrt{n}}\right) + P\left(Z \leq -1.96 + \frac{\mu_0-\mu}{1/\sqrt{n}}\right) \\
&= 1 - P\left(Z < 1.96 - \sqrt{n}\mu\right) + P\left(Z \leq -1.96 - \sqrt{n}\mu\right)
\end{aligned}
$$

```r
mu <- seq(-4,4, by=0.01)
n <- 10
power <- 1 - pnorm(1.96 - sqrt(n)*mu) +
  pnorm(-1.96 - sqrt(n)*mu)
plot(mu, power, lwd=3, type="l", col="blue")
abline(h=0.05, lwd=3, col="red")
```


power
1.0
0.8
0.6
0.4
0.2
−4
−2
0
2
4
mu

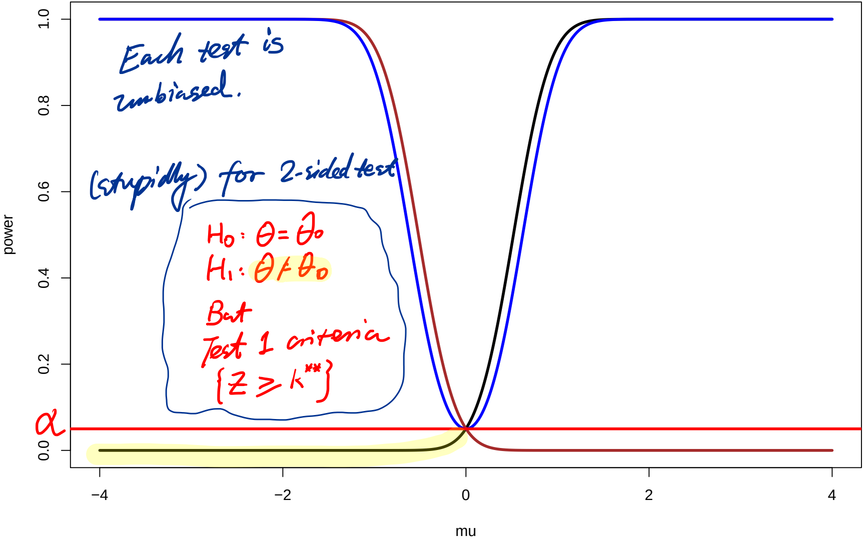
All Together

```
mu <- seq(-4,4, by=0.01)
n <- 10

##
power.1 <- 1 - pnorm(1.64 - sqrt(n)*mu)
power.2 <- pnorm(-1.64 - sqrt(n)*mu)
power.3 <- 1 - pnorm(1.96 - sqrt(n)*mu) +
  pnorm(-1.96 - sqrt(n)*mu)

##
plot(mu, power.1, lwd=3, type="l", ylab="power")
lines(mu, power.2, col="brown", lwd=3)
lines(mu, power.3, col="brown", lwd=3)

#
abline(h=0.05, lwd=3, col="red")
```



- Test 1 (black): $H_1 : \mu > 0$, Test 2 (brown): $H_1 : \mu < 0$, Test 3 (blue): $H_1 : \mu \neq 0$.

N-P Lemma

- From the plot, we see that Test 1 is UMP for $H_1 : \mu > 0$.
- From the plot, we see that Test 2 is UMP for $H_1 : \mu < 0$.

- Test 3 (maximum likelihood ratio test) is not UMP!
- Fortunately, it turns out that even when the maximum likelihood ratio test is not UMP (and many times it is), it typically has excellent properties (in particular, it can be shown to have nearly the largest possible power as the sample size increases towards infinity). As such, we tend to use the maximum likelihood ratio test in most complex testing situations where no other specific UMP test is available.