Hypothesis Testing, Econ 500

Helle Bunzel

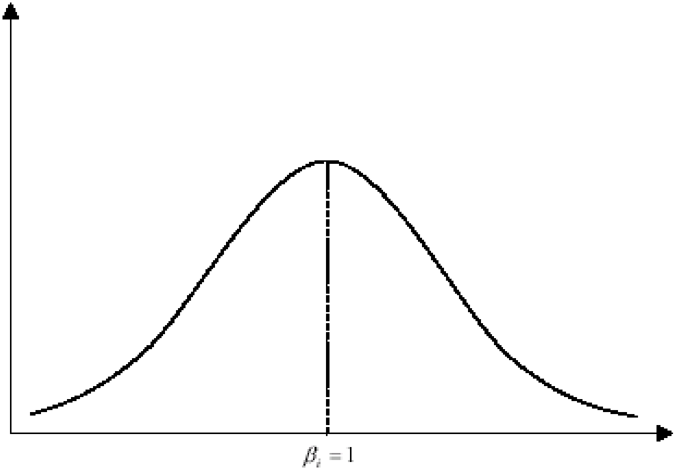
Fall 2006

The Basic Idea of Interval Estimates

- An interval rather than a point estimate is often of interest.
- Suppose we have an estimator $\hat{\beta}$ and we have determined the distribution, finite sample or asymptotic.
- We can use our knowledge about the distribution of $\hat{\beta}$ for testing.
- General principle:
  - I want to know (have a hypothesis): Is $\beta_i = 1$?
  - Suppose we know $\hat{\beta}_i \sim N\left(\beta_i, \sigma^2\right)$.

The Basic Idea of Interval Estimates

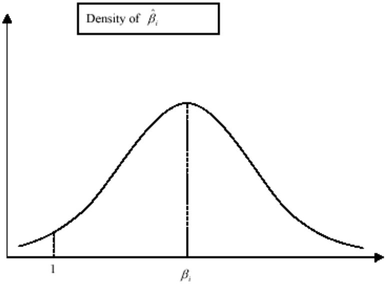
- Suppose β_i really is 1. Then the distribution of $\hat{\beta}_i$ looks as follows:



- When you have a given data set, you get a draw from this distribution.

The Basic Idea of Interval Estimates

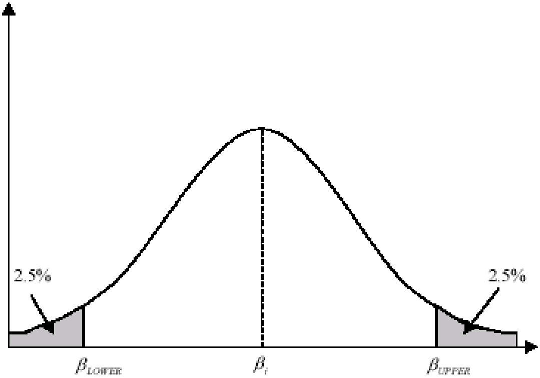
- You are more likely to get values of $\hat{\beta}_i$ that are close to β_i.
  - If your hypothesis is correct, you are very likely to have a value of $\hat{\beta}_i$ which is close to 1.
- Now, suppose that β_i is actually very far from 1:



- Given the density of $\hat{\beta}_i$, it is actually very unlikely that we get an estimate from the data which is close to 1.
- This is the principle which we'll use to create confidence intervals and tests for hypotheses.

The Basic Idea of Interval Estimates

- Confidence intervals:
  - Shade 2.5% of the area under the density function on each side of the mean: (We're defining "close to" and "far away from")



  - Now, there is a 95% chance of getting a draw in-between β_{LOWER} and β_{UPPER}. (Repeated data set)

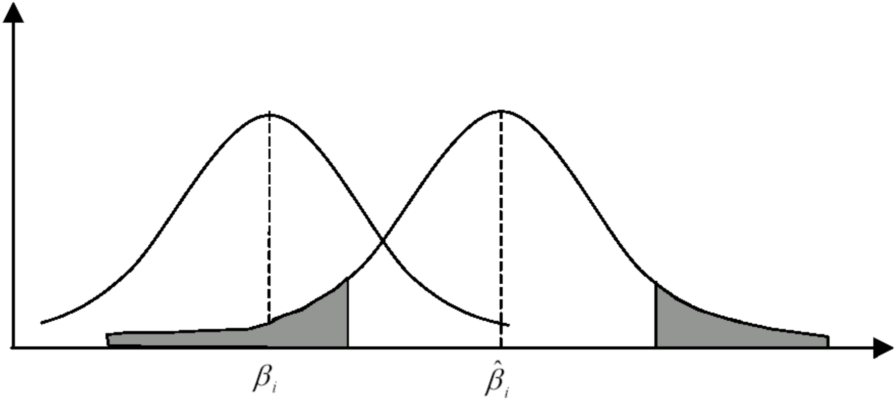
The Basic Idea of Interval Estimates

- We'd like to be able to calculate this interval.
  - Why can't we?
  - We don't know β and we don't know σ^2.
  - Assume for now that we know σ^2. Then:
    - Estimate β_i with $\hat{\beta}_i$.
    - Assuming that $\hat{\beta}_i$ is the mean of the distribution, we can calculate the intervals.
    - It is simply the 2.5 and the 97.5 percentiles of the distribution

$$N\left(\hat{\beta}_i, \sigma^2\right)$$

  - This interval we call the confidence interval.
  - The problem is that if $\hat{\beta}_i$ is very different from β_i, it doesn't say much, but it is the best we can do

The Basic Idea of Interval Estimates



- Here the true value is not even in the confidence interval.

The Basic Idea of Interval Estimates

- Saving grace:
  - When will β_i be outside the confidence interval?
  - What is the probability of getting a confidence interval like the one above?
- When we have calculated a confidence interval from the data, would it be correct to say that there is a 95% chance that β_i is inside the confidence interval?
  - NO.
  - Why?

Hypothesis Testing

- Testing a hypothesis is similar to creating a confidence interval. It is the same thing we're trying to do: We want to know which parameter values are likely.
- Our hypothesis is that $\beta_i = 1$.
- We think that $\hat{\beta}_i \sim N\left(1, \sigma^2\right)$.
- If that is the case, there is a 95% chance that we'd get a $\hat{\beta}_i$ between β_{LOWER} and β_{UPPER}. There's only a 5% chance that we'd get a value outside of that area.

$$IF\ \beta_i = 1\ THEN\ P(\hat{\beta}_i \in [\beta_{LOWER}, \beta_{UPPER}]) = 95\%$$

- This holds in a strict mathematical sense in repeated data sets, because $\hat{\beta}_i$ is a stochastic variable.
- For hypothesis testing, we turn this on its head:

$$IF\ \hat{\beta}_i \notin [\beta_{LOWER}, \beta_{UPPER}]\ \text{THEN}\ \beta_i = 1 \text{ is higly unlikely}$$

Hypothesis Testing

- This does not hold water in a strict mathematical sense.
  - Same problem as with the confidence intervals.
  - Note: An $N\left(1, \sigma^2\right)$ COULD generate $\hat{\beta}_i = 9825$
- So: we reject the hypothesis if $\hat{\beta}_i \notin [\beta_{LOWER}, \beta_{UPPER}]$.

Standardized Normal Variables and Confidence Intervals for the Mean with Known Variance

- If X is a normal random variable with mean μ and variance σ^2 then

$$Z = \left(\frac{X-\mu}{\sigma}\right) \sim N\left(0, \sigma^2\right)$$

- We estimate μ by $\hat{\mu} = \bar{X}$.
- This then implies

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (1)$$

$$z = \frac{\bar{X}-\mu}{\left[\frac{\sigma}{\sqrt{n}}\right]} \sim N(0, 1)$$

Standardized Normal Variables and Confidence Intervals for the Mean with Known Variance

- Define γ_1 as the upper $\alpha/2$ percent critical value of a standard normal variable.
- then

$$\begin{aligned}
1-\alpha &= \int_{-\gamma_1}^{\gamma_1} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2}\, dz = F(\gamma_1) - F(-\gamma_1) \\
&= \Pr\left[-\gamma_1 \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \gamma_1\right] \\
&= \Pr\left[-\gamma_1 \frac{\sigma}{\sqrt{n}} \leq \bar{x}-\mu \leq \gamma_1 \frac{\sigma}{\sqrt{n}}\right] \\
&= \Pr\left[\gamma_1 \frac{\sigma}{\sqrt{n}} \geq -\bar{x}+\mu \geq -\gamma_1 \frac{\sigma}{\sqrt{n}}\right] \\
&= \Pr\left[\bar{x}-\gamma_1 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}+\gamma_1 \frac{\sigma}{\sqrt{n}}\right]
\end{aligned}$$

Standardized Normal Variables and Confidence Intervals for the Mean with Known Variance

- Therefore, with σ known, the $(1-\alpha)$ 100% confidence interval for μ is:

$$\left[\bar{x} - \gamma_1 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \gamma_1 \frac{\sigma}{\sqrt{n}}\right]$$

Confidence Intervals for the Mean with Known Variance

Example

- Consider the income data for carpenters and house painters:

	carpenters	painters
sample size	$n_c = 12$	$n_p = 15$
mean income	$\bar{c} = \$6000$	$\bar{p} = \$5400$
estimated variance	$s_c^2 = \$565\,000$	$s_p^2 = \$362\,500$

- Assume:
  - the data for both groups of individuals is distributed normally.
  - σ^2 is known and equal to \$600,000.
- Consider a 95% confidence interval for the mean, μ_c.
- For the normal distribution with $\frac{\alpha}{2} = 0.025$, $\gamma_1 = 1.96$.

- Then

$$
\begin{aligned}
& 1-\alpha \\
= \quad & \Pr\left[6000-\frac{\sqrt{600\,000}}{\sqrt{12}}(1.96) \leq \mu_c \leq 6000+\frac{\sqrt{600\,000}}{\sqrt{12}}(1.96)\right] \\
= \quad & \Pr\left[6000-(223.607)(1.96) \leq \mu_c \leq 6000+(223.607)(1.96)\right] \\
= \quad & \Pr\left[5561.73 \leq \mu_c \leq 6438.27\right]
\end{aligned}
$$

Confidence Intervals for the Mean with Unknown Variance

- Recall that the t random variable is defined as

$$t = \frac{z}{\sqrt{\frac{\chi^2(\nu)}{\nu}}}$$

- where:
  - z is a standard normal
  - $\chi^2(\nu)$ is a χ^2 random variable with ν degrees of freedom
  - z and $\chi^2(\nu)$ are independent.
- Also recall that if $X_i \sim N(\mu, \sigma^2)$ then

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

and
$$\sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi^2(n-1)$$

Confidence Intervals for the Mean with Unknown Variance

- If $X \sim N(\beta, \sigma^2)$ and $\bar{X} = \hat{\mu}$, then we would have

$$\sum_{i=1}^{n} \left(\frac{X_i - \hat{\mu}}{\sigma} \right)^2 \sim \chi^2(n-1)$$

- Now find the distribution of S^2:

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \\ &= \frac{\sigma^2}{n-1} \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \end{aligned}$$

- So

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi^2(n-1)$$

Confidence Intervals for the Mean with Unknown Variance

- It can be shown that $\frac{\bar{X}-\mu}{\left[\frac{\sigma}{\sqrt{n}}\right]}$ and $\frac{(n-1)S^2}{\sigma^2}$ are independent, so

$$\frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\frac{1}{\sigma}\sqrt{S^2}}$$

$$= \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\sqrt{n}\,(\bar{X}-\mu)}{S} \sim t(n-1)$$

Confidence Intervals for the Mean with Unknown Variance

- If γ_1 is the upper $\alpha/2$ percent critical value of a t random variable then

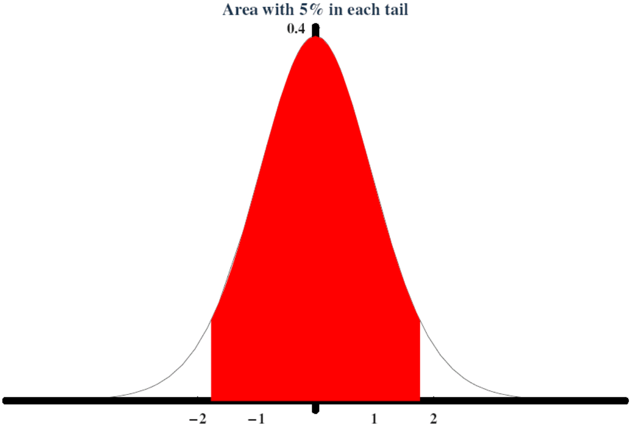
$$1-\alpha = F(\gamma_1) - F(-\gamma_1) = \Pr\left[-\gamma_1 \leq \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} \leq \gamma_1\right]$$

$$= \Pr\left[-\gamma_1 \leq \frac{\sqrt{n}(\bar{X}-\mu)}{S} \leq \gamma_1\right] = \Pr\left[\frac{-\gamma_1 S}{\sqrt{n}} \leq \bar{X}-\mu \leq \frac{\gamma_1 S}{\sqrt{n}}\right]$$

$$= \Pr\left[\bar{X} - \frac{\gamma_1 S}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{\gamma_1 S}{\sqrt{n}}\right]$$

Confidence Intervals for the Mean with Unknown Variance

- This is the $(1-\alpha)(100\%)$ confidence interval for μ:



- Consider the income data for carpenters and house painters.
- We can construct the 95% confidence interval as follows:

$$
\begin{aligned}
& 1-\alpha \\
& = \Pr\left[6000 - \frac{\sqrt{565\,000}}{\sqrt{12}}(2.201) \leq \mu_c \leq 6000 + \frac{\sqrt{565\,000}}{\sqrt{12}}(2.201)\right] \\
& = \Pr\left[6000 - (216.99)(2.201) \leq \mu_x \leq 6000 + (216.99)(2.201)\right] \\
& = \Pr\left[5522.4 \leq \mu_c \leq 6477.6\right]
\end{aligned}
$$

- So, for example, 7000 is not in the 95% confidence interval for the mean.

Confidence Intervals for the Variance

- Recall that
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$
- Now if γ_1 and γ_2 are such that
$$\begin{aligned} \Pr(\chi^2(\nu) \leq \gamma_1) &= \frac{\alpha}{2} \\ \Pr(\chi^2(\nu) \geq \gamma_2) &= \frac{\alpha}{2} \end{aligned}$$

Confidence Intervals for the Variance

- Then

$$1-\alpha = F(\gamma_2, \nu) - F(\gamma_1, \nu) = \Pr\left[\gamma_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq \gamma_2\right]$$

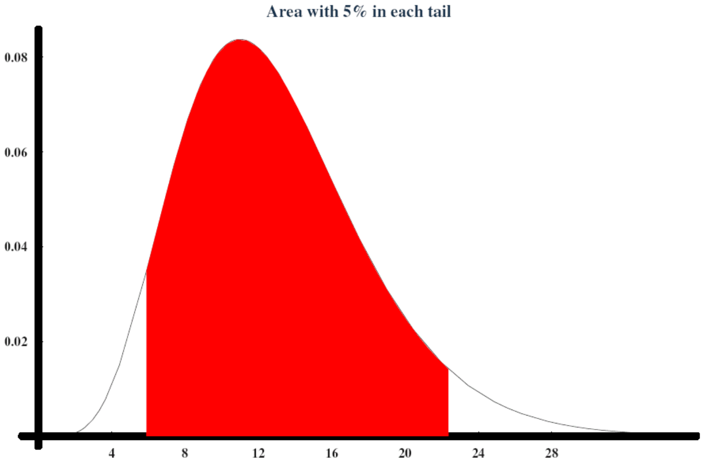
$$= \Pr\left[\frac{\gamma_1 \sigma^2}{n-1} \leq S^2 \leq \frac{\gamma_2 \sigma^2}{n-1}\right]$$

$$= \Pr\left[\frac{(n-1)}{\sigma^2}\gamma_1 \geq \frac{1}{S^2} \geq \frac{(n-1)}{\sigma^2 \gamma_2}\right]$$

$$= \Pr\left[\frac{(n-1)S^2}{\gamma_2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\gamma_1}\right]$$

- Thus $\left[\frac{(n-1)S^2}{\gamma_2}; \frac{(n-1)S^2}{\gamma_1}\right]$ is the $(1-\alpha)$ 100% confidence interval for the variance σ^2.

Confidence Intervals for the Variance



Confidence Intervals for the Variance

Example

- Consider the income data for carpenters and house painters
- We can construct the 95% confidence interval for the variance of the income of the painters as follows:

$$1-\alpha \;=\; \Pr\left[\frac{(n_p-1)S_p^2}{\gamma_2} \leq \sigma_p^2 \leq \frac{(n_p-1)S_p^2}{\gamma_1}\right]$$

- Recall the data:

	carpenters	painters
sample size	$n_c = 12$	$n_p = 15$
mean income	$\bar{c} = \$6000$	$\bar{p} = \$5400$
estimated variance	$s_c^2 = \$565\,000$	$s_p^2 = \$362\,500$

- For a χ^2 distribution with 14 degrees of freedom we obtain $\gamma_1 = 5.63$ and $\gamma_2 = 26.12$.

- We then get

$$
\begin{aligned}
1-\alpha &= \Pr\left[\frac{(14)(362\,500)}{26.12} \leq \sigma_p^2 \leq \frac{(14)(362\,500)}{5.63}\right] \\
&= \Pr\left[195\,042 \leq \sigma_p^2 \leq 901\,421\right]
\end{aligned}
$$

Two Samples and a Confidence Interval for the Variances

- Suppose that:
  - $x_1^1, x_2^1, \ldots, x_n^1$ is a random sample from a distribution $X_1 \sim N(\mu_1, \sigma_1^2)$
  - $x_1^2, x_2^2, \ldots, x_n^2$ is a random sample from a distribution $X_2 \sim N(\mu_2, \sigma_2^2)$

Two Samples and a Confidence Interval for the Variances

- Recall that if

$$\frac{(n_1 - 1) S_1^2}{\sigma_1^2} \sim \chi^2 (n_1 - 1)$$
$$\frac{(n_2 - 1) S_2^2}{\sigma_2^2} \sim \chi^2 (n_2 - 1)$$

- Then

$$\frac{\frac{\left(\frac{(n_1 - 1) S_1^2}{\sigma_1^2}\right)}{(n_1 - 1)}}{\frac{\frac{(n_2 - 1) S_2^2}{\sigma_2^2}}{(n_2 - 1)}} = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F\left(n_1 - 1,\ n_2 - 1\right)$$

Two Samples and a Confidence Interval for the Variances

- F distributions are normally tabled giving the area in the upper tail

$$1-\alpha = \Pr\left(F_{n_1-1,\,n_2-1} \le \gamma\right) \quad \text{or} \quad \alpha = \Pr\left(F_{n_1-1,\,n_2-1} \ge \gamma\right)$$

- Now let γ_1 and γ_2 be such that

$$\begin{aligned}
\Pr\left(F_{n_1-1,\,n_2-1} \le \gamma_1\right) &= \frac{\alpha}{2} \\
\Pr\left(F_{n_1-1,\,n_2-1} \ge \gamma_2\right) &= \frac{\alpha}{2} \\
\Pr\left(F_{n_1-1,\,n_2-1} \le \gamma_2\right) &= 1-\frac{\alpha}{2} \\
\Pr\left(F_{n_1-1,\,n_2-1} \le \gamma_2\right) - \Pr\left(F_{n_1-1,\,n_2-1} \le \gamma_1\right) &= 1-\frac{\alpha}{2}-\frac{\alpha}{2} = 1-\alpha
\end{aligned}$$

Two Samples and a Confidence Interval for the Variances

- We can now construct a confidence interval as follows:

$$\begin{aligned} & 1-\alpha \\ & = \Pr\left[\gamma_1 \le \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} \le \gamma_2\right] = \Pr\left[\frac{S_2^2}{S_1^2}\gamma_1 \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{S_2^2}{S_1^2}\gamma_2\right] \\ & = \Pr\left[\frac{S_1^2}{S_2^2}\frac{1}{\gamma_1} \ge \frac{\sigma_1^2}{\sigma_2^2} \ge \frac{S_1^2}{S_2^2}\frac{1}{\gamma_2}\right] = \Pr\left[\frac{S_1^2}{S_2^2}\frac{1}{\gamma_2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2}\frac{1}{\gamma_1}\right] \end{aligned}$$

- $\left[\frac{S_1^2}{S_2^2}\frac{1}{\gamma_2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2}\frac{1}{\gamma_1}\right]$ is then the $(1-\alpha)$ 100% confidence interval for ratio of the variances.

- Consider the income data for carpenters and house painters
- Write

$$1-\alpha = \Pr\left[\frac{S_c^2}{S_p^2}\frac{1}{\gamma_2} \leq \frac{\sigma_c^2}{\sigma_p^2} \leq \frac{S_c^2}{S_p^2}\frac{1}{\gamma_1}\right]$$

- The upper critical level is $F(11,\ 14, : 0.025) = 3.10$.
- Since the tables don't contain the lower tail, we obtain the critical value $\gamma_1(11,\ 14)$ as $(1/\gamma_1(14,\ 11)$.
- This critical value is given by

$$\frac{1}{F(14,\ 11 : 0.025)} = \frac{1}{3.36} = 0.297$$

Two Samples and a Confidence Interval for the Variances
Example

- The confidence interval is then given by

$$
\begin{aligned}
1-\alpha &= \Pr\left[\frac{565\,000}{362\,500}\frac{1}{3.1} \leq \frac{\sigma_c^2}{\sigma_p^2} \leq \frac{565\,000}{362\,500}\frac{1}{.297}\right] \\
&= \Pr\left[.502\,78 \leq \frac{\sigma_c^2}{\sigma_p^2} \leq 5.248\right]
\end{aligned}
$$

Two Samples and a Confidence Interval for the Variances

Example

- Consider the following 18 exam scores:

46	58	87	97.5	82.5	68
83.25	99.5	66.5	75.5	62.5	67
78	32	74.5	47	99.5	26

- The mean of the data is 69.4583.
- The variance is 466.899
- The standard deviation is 21.6078.
- We are interested in the null hypothesis $\mu = 80$.
- From a table: The $t\,(17)$ 0.025 critical value is $\gamma_1 = 2.110$.

Two Samples and a Confidence Interval for the Variances

- This means then that

$$1-\alpha = 0.95 = F(\gamma_1) - F(-\gamma_1) = \Pr\left[-\gamma_1 \leq \frac{\sqrt{n}(\bar{y}-\mu)}{S} \leq \gamma_1\right]$$

$$= \Pr\left[-2.110 \leq \frac{4.2426(69.4583-\mu)}{21.6078} \leq 2.110\right]$$

$$= \Pr\left[-10.746\,35 \leq 69.4583-\mu \leq 10.746\,35\right]$$

$$= \Pr\left[10.746\,35 \geq -69.4583+\mu \geq -10.746\,35\right]$$

$$= \Pr\left[80.204\,65 \geq 69.4583+\mu \geq -58.7119\right]$$

$$= \Pr\left[58.7119 \leq \mu \leq 80.204\,65\right]$$

Hypothesis Testing

Definition

A statistical hypothesis is a conjecture about the distribution of one or more random variables.

Definitions

If a statistical hypothesis completely specifies the distribution, it is referred to as a **simple hypothesis**; if not, it is referred to as a **composite hypothesis**.

- A simple hypothesis must specify not only the functional form of the underlying distribution, but also the values of all parameters.
- A statement regarding a parameter θ, such as $\theta \in \omega \subset \Omega$, is called a statistical hypothesis about θ and is usually referred to by H_0.
- The decision on accepting or rejecting the hypothesis is based on the value of a test statistic, the distribution P_θ of which is known to belong to a class $P = \{P_\theta, , \theta \in \Omega\}$.

Hypothesis Testing

- The distributions of P can then be classified into those for which the hypothesis is true and those for which it is false.
- The resulting two mutually exclusive classes are denoted H and K (or H_0 and H_1), and the corresponding subsets of Ω by Ω_H and Ω_K
- Mathematically the hypothesis is equivalent to the statement that P_θ is an element of H.

- To test statistical hypotheses it is necessary to formulate alternative hypotheses.
- The statement that $\theta \in \omega^c$ is also a statistical hypothesis about θ, which is called the *alternative* to H_0
  - This is usually denoted H_1 or H_A.
- Thus we have

$$H_0 : \theta \in \omega$$

$\mathrm{H}_1 : \theta \in \omega^c$

- If ω contains only one point, that is, $\omega = \{\theta_0\}$, then H_0 is called a *simple* hypothesis
- Otherwise it is called a *composite* hypothesis.

- Example of a composite hypothesis:

$$\mu \geq \mu_0$$

- with composite alternative

$$\mu < \mu_0$$

- We usually call the hypothesis we are testing the **null** hypothesis.

- Let the decision of not rejecting or rejecting H_0 be denoted by d_0 and d_1 respectively.
- A test procedure assigns to each possible outcome of X one of these two decisions
- It divides the sample space into two complementary regions S_0 and S_1.
- If X falls into S_0 the hypothesis is not rejected; otherwise it is rejected.
- The set S_0 is called the region of non-rejection, the set S_1 is called the region of rejection or **critical** region.
- Tests where H_0 is of the form $\theta = \theta_0$ and the alternative H_1 is of the two-sided form $\theta \neq \theta_0$ are called two-tailed tests.
- When H_0 is of the form $\theta = \theta_0$ but the alternative is $H_1\colon\ \theta < \theta_0$ or $H_1\colon\ \theta > \theta_0$, the test is called a one-tailed test.

- Suppose

$$H_0 \quad : \quad \theta = \theta_0$$
$$H_A \quad : \quad \theta = \theta_1$$

- The statistician can make two possible types of errors:
  - If $\theta = \theta_0$ and the test rejects $\theta = \theta_0$ and concludes $\theta = \theta_1$ then the error is of type I:

    $$\textbf{Rejection of the null hypothesis when it is true is a type I error}.$$

  - If $\theta = \theta_1$ and the test does not reject $\theta = \theta_0$ but then the error is of type II.

    $$\textbf{Non} - \textbf{rejection of the null hypothesis}$$
    $$\textbf{when it is false is a type II error}.$$

Example

Suppose that the null hypothesis is that a person tested for HIV does not have the disease. Then a false positive is a Type I error while a false negative is a Type II error.

- We can summarize as follows where S_1 is the region of rejection or critical region.

$$
\begin{aligned}
P_\theta\{X \in S_1\} &= \text{Probability of a Type I error if } \theta \in \Omega_H \\
P_\theta\{X \in S_1\} &= (\text{1 - Probability of a Type II error}) \text{ if } \theta \in \Omega_K \\
P_\theta\{X \in S_0\} &= \text{Probability of a Type II error if } \theta \in \Omega_K
\end{aligned}
$$

Size, Power and Significance Level

- Typically we bound to the probability of incorrectly rejecting H_0 when it is true.
- Taking this bound as given, we minimize the probability of accepting H_0 when it is false.
- In practice we select a number α between 0 and 1 such that:

$$P_\theta\{X \in S_1\} \;\leq\; \alpha, \text{ for all } \theta \in \Omega_H$$

  - α is called the level of significance.
- This says that the probability that $X \in S_1$ is less than α when θ is in the hypothesized part of its domain.
- Subject to the constraint, it is desired to either

$$\min_{\theta \in \Omega_K} P_\theta\{X \in S_0\}$$

or

$$\max_{\theta \in \Omega_K} P_\theta\{X \in S_1\}$$

Size, Power and Significance Level

- Note that these are identical since

$$\min_{\theta \in \Omega_K} P_\theta\{X \in S_0\} = 1 - \max_{\theta \in \Omega_K} P_\theta\{X \in S_1\}$$

- Usually

$$\sup_{\theta \in \Omega_H} P_\theta\{X \in S_1\} = \alpha$$

- But we have a seperate name for $\sup_{\theta \in \Omega_H} P_\theta\{X \in S_1\}$: Size of the test.
- The probability of rejection evaluated for a given $\theta \in \Omega_K$ is called the **power** of the test against the alternative θ.
- Considered as a function of θ for all $\theta \in \Omega$, this probability is called the power function of the test and is denoted by $\beta(\theta)$:

$$\text{Power Function} = \beta(\theta) = P_\theta\{X \in S_1\}$$

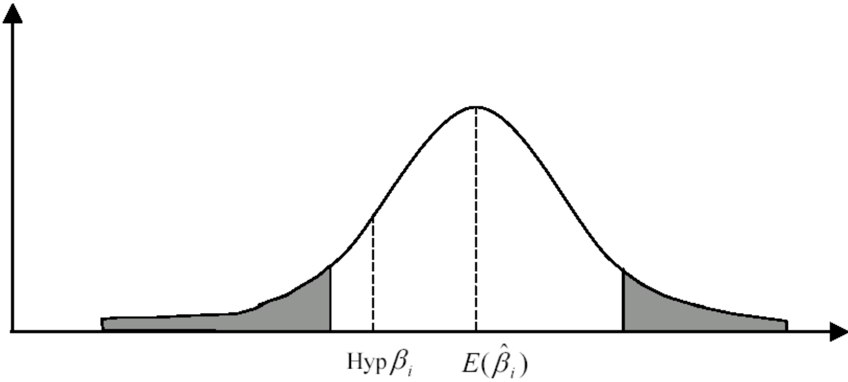
Size, Power and Significance Level

- The ideal power function is 0 for all $\theta \in \Omega_H$ and 1 for all $\theta \in \Omega_K$.
- Can you construct a test with power 1?
- The probabilities of committing the two types of error can be summarized as:

		Decision	
		Do not reject H_0	Reject H_0
$\mathbf{H_0}$	True	Correct Decision	Type I Error
	False	Type II Error	Correct Decision

- Consider repeated datasets to which we apply the same test statistic.
- The limiting proportion of trials in which H_0 is rejected when H_0 is true is the **probability of a type I error**
- The limiting proportion of trials in which H_0 is not rejected when H_1 is true is the **probability of a type II error**

Size, Power and Significance Level



Size, Power and Significance Level

- So:

$$\begin{aligned} \alpha &= P\{\text{Type I error}\} = P\{\text{Reject H}_0 \text{ when H}_0 \text{ is true}\} \\ &= P\{\text{Reject H}_0 | \text{H}_0\} \end{aligned}$$

  and

$$\begin{aligned} 1-\beta &= P\{\text{Type II error}\} = P\{\text{Fail to reject H}_0 \text{ when H}_1 \text{ is true}\} \\ &= P\{\text{Fail to reject H}_0 | \text{H}_1\} \end{aligned}$$

- and the power of the test is

$$\beta(\theta | \theta \in \Omega_K) = P\{\text{Reject H}_0 \mid H_1\}$$

- The **power** of the test, is the probability that the test correctly rejects H_0. A test with high power is preferred.

Size, Power and Significance Level

Definition

The *p*-value associated with a statistical test is the probability that we obtain the observed value of the test statistic or a value that is more extreme in the direction of the alternative hypothesis calculated when H_0 is true.

- The *p*-value of a test can be reported instead of reject or not reject.
- If T is a test statistic, the *p*-value, or *attained significance level*, is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.
- To compute a *p*-value:
  - compute the realized value of the test statistic
  - find the probability that the random variable represented by the test statistic is larger than this realized value in the given sample.

Likelihood ratio tests

- Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with pdf $f(x|\theta)$. T
- hen the likelihood function of a sample is given by

$$L(\theta|x_1, x_2, \ldots, x_n) = L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta)$$

- Let Ω denote the entire parameter space such that $\theta \in \Omega$.
- The *likelihood ratio test statistic* for testing $\{H_0 : \theta \in \Omega_0\}$ versus $\{H_1 : \theta \in \Omega_1\}$ is

$$\lambda(x) = \frac{\sup_{\Omega_0} L(\theta|x)}{\sup_{\Omega} L(\theta|x)}$$

- The *likelihood ratio test* (LRT) rejects on the set $\{x : \lambda(x) \leq c\}$ where c is any number satisfying $0 \leq c \leq 1$.
  - Why is c between 0 and 1?

Likelihood ratio tests

- Suppose $\hat{\theta}$, is an MLE of θ obtained by maximizing $L(\theta|x)$ without any restrictions on Ω.
- Then consider the MLE of θ, call it $\hat{\theta}_0$, which is obtained by maximizing $L(\theta|x)$ subject to the constraint that $\theta \in \Omega_0$.
- Then the LRT statistic is given by

$$\lambda(x) = \frac{L(\hat{\theta}_0|x)}{L(\hat{\theta}|x)}$$

- We reject the null hypothesis that $H_0 : \theta = \theta_0$ if

$$\lambda(x) = \frac{L(\hat{\theta}_0|x)}{L(\hat{\theta}|x)} < k$$

  where k is some constant.
- Common practice is to work with $-\log \lambda(x)$.

Likelihood ratio tests

- Then the critical region becomes

$$
\begin{aligned}
& \log\left[\lambda(x)\right] \\
= \quad & \log\left[L(\hat{\theta}_0|x)\right] \; - \; \log\left[L(\hat{\theta}|x)\right] \; < \; \log k \\
\Rightarrow \quad & -\log\left[\lambda(x)\right] \; = \; \log\left[L(\hat{\theta}|x)\right] \; - \; \log\left[L(\hat{\theta}_0|x)\right] \; > \; -\log k \\
\Rightarrow \quad & -\log\left[\lambda(x)\right] \; > \; c, \quad c \; = \; -\log k \; > \; 0 \text{ because } k \leq 1
\end{aligned}
$$

- Asymptotically, $-2\log\left[\lambda(x)\right]$ is distributed as a $\chi^2(\upsilon)$.
  - υ is equal to the difference in the dimension of Ω and Ω_H
  - This is often equal to the number of restrictions imposed by the null hypothesis.

- Consider a random sample $X_1, X_2, \ldots, X_n$ from a $N(\mu, \sigma^2)$ population.
- Consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.
- Assume that σ^2 is known.
- We know that the MLE estimator of μ is $\bar{X}$. Thus the denominator of $\lambda(x)$ is $L(\bar{x}|x_1, x_2, \ldots, x_n)$.
- So the LRT statistic is

$$\lambda(x) = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{\frac{-1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu_0)^2}}{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{\frac{-1}{2\sigma^2}\sum_{i=1}^n (x_i-\bar{x})^2}} = \frac{e^{\frac{-1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu_0)^2}}{e^{\frac{-1}{2\sigma^2}\sum_{i=1}^n (x_i-\bar{x})^2}} = e^{\frac{-1}{2\sigma^2}\left(\sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\bar{x})^2\right)}$$

Likelihood ratio tests

- Rewrite $\sum_{i=1}^{n}(x_i - \mu_0)^2$ as follows

$$\begin{aligned}
&\sum_{i=1}^{n}(x_i - \mu_0)^2 \\
= \quad & \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \mu_0)^2 \\
= \quad & \sum_{i=1}^{n}(x_i - \bar{x})^2 + 2(\bar{x} - \mu_0)\sum_{i=1}^{n}(x_i - \bar{x}) + n(\bar{x} - \mu_0)^2 \\
= \quad & \sum_{i=1}^{n}(x_i - \bar{x})^2 + 2(\bar{x} - \mu_0)(n\bar{x} - n\bar{x}) + n(\bar{x} - \mu_0)^2 \\
= \quad & \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2
\end{aligned}$$

- Now substitute this into the expression of the LR test:

$$\begin{aligned}\lambda(x) &= e^{\frac{-1}{2\sigma^2}\left(\sum_{i=1}^n (x_i-\mu_0)^2 \;-\; \sum_{i=1}^n (x_i-\bar{x})^2\right)} \\ &= e^{\frac{-1}{2\sigma^2}\left(\sum_{i=1}^n (x_i \;-\; \bar{x})^2 \;+\; n(\bar{x}-\mu_0)^2 \;-\; \sum_{i=1}^n (x_i-\bar{x})^2\right)} \\ &= e^{\frac{-n}{2\sigma^2}(\bar{x}-\mu_0)^2}\end{aligned}$$

- We reject $H_0 : \mu = \mu_0$ if

$$\begin{aligned}\lambda(x) &= e^{\frac{-n}{2\sigma^2}(\bar{x}-\mu_0)^2} \leq k \\ &\Rightarrow \frac{-n}{2\sigma^2}(\bar{x} - \mu_0)^2 \leq \log k \\ &\Rightarrow (\bar{x} - \mu_0)^2 \geq -\frac{2\sigma^2}{n}\log k\end{aligned}$$

$$
\begin{aligned}
\lambda(x) &\leq k \\
\Rightarrow |\bar{x} - \mu_0| &\geq \sqrt{\frac{-2\sigma^2}{n} \log k} \\
\Rightarrow \frac{|\bar{x} - \mu_0|}{\frac{\sigma}{n}} &\geq \frac{\sqrt{\frac{-2\sigma^2}{n} \log k}}{\frac{\sigma}{n}} \\
\Rightarrow \frac{|\bar{x} - \mu_0|}{\frac{\sigma}{n}} &\geq \gamma
\end{aligned}
$$

- γ will be determined so that the critical region is of the appropriate size (5% typically).
- Because $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$ under the null hypothesis, we can use the standard normal table.

- Consider a random sample $X_1, X_2, \ldots, X_n$ from a $N(\mu, \sigma^2)$ population.
- Consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.
- Assume that σ^2 is NOT known.
- Recall that the MLE estimator of μ is $\bar{X}$ and the MLE estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

- The restricted estimators are $\hat{\mu}_0 = \mu_0$ and

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2.$$

- So the LRT statistic is

$$\lambda(x) = \frac{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^n e^{\frac{-1}{2\hat{\sigma}_0^2}\sum_{i=1}^n (x_i-\mu_0)^2}}{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n e^{\frac{-1}{2\hat{\sigma}^2}\sum_{i=1}^n (x_i-\bar{x})^2}}$$

- Now rewrite the numerator as follows

$$\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^n e^{\frac{-1}{2\hat{\sigma}_0^2}\sum_{i=1}^n (x_i-\mu_0)^2} = \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^n e^{\frac{-n\hat{\sigma}_0^2}{2\hat{\sigma}_0^2}} = \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^n e^{\frac{-n}{2}}$$

Likelihood ratio tests

Example

- Then rewrite the denominator as follows

$$\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n e^{\frac{-1}{2\hat{\sigma}^2}\sum_{i=1}^n (x_i-\bar{x})^2} = \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n e^{\frac{-n\hat{\sigma}^2}{2\hat{\sigma}^2}}$$
$$= \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n e^{\frac{-n}{2}}$$

- Now rewrite the LR test using these:

$$\lambda(x) = \frac{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^n e^{\frac{-1}{2\hat{\sigma}_0^2}\sum_{i=1}^n (x_i-\mu_0)^2}}{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n e^{\frac{-1}{2\hat{\sigma}^2}\sum_{i=1}^n (x_i-\bar{x})^2}}$$
$$= \frac{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^n}{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n} = \left(\frac{\frac{1}{\hat{\sigma}_0}}{\frac{1}{\hat{\sigma}}}\right)^n = \left(\frac{\hat{\sigma}_0}{\hat{\sigma}}\right)^{-n}$$

- Then:

$$\lambda(x) = \left(\frac{\hat{\sigma}_0}{\hat{\sigma}}\right)^{-n} = \left(\frac{\sum_{i=1}^{n}(x_i - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^{\frac{-n}{2}}$$

- Now make the substitution for $\sum_{i=1}^{n}(x_i - \mu_0)^2$ as before:

$$\begin{aligned}\lambda(x) &= \left(\frac{\sum_{i=1}^{n}(x_i - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^{\frac{-n}{2}} \\ &= \left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^{\frac{-n}{2}} \\ &= \left(1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^{\frac{-n}{2}}\end{aligned}$$

- We reject the hypothesis if:

$$
\begin{aligned}
\lambda(x) &= \left(1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^{-\frac{n}{2}} \leq c \\
&\Rightarrow \left(1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) \leq c^{-\frac{2}{n}} \\
&\Rightarrow \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \leq c^{-\frac{2}{n}} - 1 \\
&\Rightarrow \left(\frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right)^2 \leq c^{-\frac{2}{n}} - 1
\end{aligned}
$$

- Continuing:

$$\left(\frac{\sqrt{n}\left(\bar{x}-\mu_0\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}}\right)^2 \leq c^{-\frac{2}{n}}-1$$

$$\Rightarrow \frac{1}{n-1}\left(\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}}\right)^2 \geq c^{\frac{-2}{n}}-1$$

$$\Rightarrow \frac{1}{n-1}\left(\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\right)^2 \geq c^{\frac{-2}{n}}-1$$

$$\Rightarrow \left(\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\right)^2 \geq (n-1)\left(c^{\frac{-2}{n}}-1\right)$$

- And finally:

$$\left(\frac{\sqrt{n}(\bar{x} - \mu_0)}{s}\right)^2 \geq (n-1)\left(c^{\frac{-2}{n}} - 1\right)$$

$$\Rightarrow \left(\frac{\sqrt{n}(\bar{x} - \mu_0)}{s}\right) \geq \sqrt{(n-1)\left(c^{\frac{-2}{n}} - 1\right)} = \gamma$$

- γ will be determined so that the critical region is of the appropriate size.
- Note that the left hand side is distributed as a t random variable.
- Specifically, we reject $H_0 : \mu = \mu_0$ if

$$\left|\frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}\right| \geq t_{\alpha/2}(n-1)$$

Likelihood ratio tests

Example

- Recall the following example:
- Consider the following 18 exam scores:

46	58	87	97.5	82.5	68
83.25	99.5	66.5	75.5	62.5	67
78	32	74.5	47	99.5	26

- The mean of the data is 69.4583.
- The variance is 466.899
- The standard deviation is 21.6078.
- We are interested in the null hypothesis $\mu = 80$.
- We can compute the t-statistic as follows.

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} = \frac{69.4583 - 80}{\frac{21.6078}{\sqrt{18}}}$$

$$= -2.069\,83$$

- Consider an α level of $\alpha = 0.05$.
- The value of the t statistic is $-2.069\,83$.
- We reject the null hypothesis if $|-2.069\,83| > t_{n-1,\,.025}$.
- From the t-tables $t_{17,\,.025} = 2.110$.
- We cannot reject the null hypothesis that $\mu = 80$.

Most Powerful Tests

- A test with level α that has the highest possible power is called a **most powerful** (MP) test.
- A test statistic is MP if it is associated with an MP test.
- Consider a test statistic X.
- Consider an indicator function $\varphi_k(x)$ where $0 \leq k \leq \infty$ and x is the observation vector.
- We define $\varphi_k(x)$ as follows

$$\varphi_k(x) = \begin{cases} 1 & \text{if } X > k \\ 0 & \text{if } X < k \end{cases}$$

- The value of k is such that $X > k \Rightarrow X \in S_1$.
- For the likelihood ratio test this would yield

$$\varphi_k(x) = \begin{cases} 1 & \text{if } \lambda(x, \theta_0, \theta_1) > k \\ 0 & \text{if } \lambda(x, \theta_0, \theta_1) < k \end{cases}$$

Most Powerful Tests

- The lemma below is shortened compared to the notes.

Theorem (Neyman-Pearson Lemma)

1. *If* $\alpha > 0$ *and* φ_k *is a size* α *likelihood ratio test, then* φ_k *is MP in the class of level* α *tests.*
2. *If* φ *is an MP level* α *test, then it must be a level* α *likelihood ratio test; that is there exists* k *such that*

$$P_\theta\left[\varphi(\cdot) \neq \varphi_k(\cdot), \lambda(\cdot, \theta_0, \theta_1) \neq k\right] = 0 \tag{3}$$

for $\theta = \theta_0$ *and* $\theta = \theta_1$.