# Topic 16: Composite Hypotheses

November, 2008

For a **composite hypothesis**, the parameter space $\Theta$ is divided into two disjoint regions, $\Theta_0$ and $\Theta_1$. The test is written

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1$$

with $H_0$ is called the null hypothesis and $H_1$ the alternative hypothesis.

Rejection and failure to reject the null hypothesis, critical regions, $C$, and type I and type II errors have the same meaning for a composite hypotheses as it does with a simple hypothesis.

## 1 Power

**Power** is now a function

$$\pi(\theta) = P_\theta\{X \in C\}.$$

that gives the probability of rejecting the null hypothesis for a given value of the parameter. Consequently, the ideal test function has

$$\pi(\theta) \approx 0 \text{ for all } \theta \in \Theta_0 \text{ and } \pi(\theta) \approx 1 \text{ for all } \theta \in \Theta_1$$

and the test function yields the correct decision with probability nealry 1.

In reality, incorrect decisions are made. For $\theta \in \Theta_0$,

$$\pi(\theta) \ \text{ is the probability of making a type I error}$$

and for $\theta \in \Theta_1$,

$$1 - \pi(\theta) \ \text{ is the probability of making a type II error.}$$

The goal is to make the chance for error small. The traditional method is the same as that employed in the Neyman-Pearson lemma. Fix a **level** $\alpha$, defined to be the largest value of $\pi(\theta)$ in the region $\Theta_0$ defined by the null hypothesis and look for a critical region that makes the power function large for $\theta \in \Theta_1$

**Example 1.** *Let $X_1, X_2, \ldots, X_n$ be independent $N(\mu, \sigma^2)$ random variables with $\sigma^2$ known and $\mu$ unknown. For the composite hypothesis for the* **one-sided test**

$$H_0 : \mu \leq \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0.$$

*We use the test statistic from the likelihood ratio test and reject $H_0$ if $\bar{X}$ is too large. The power function*

$$\pi(\mu) = P_\mu\{\bar{X} \geq k(\mu_0)\}.$$

*To obtain level $\alpha$, note that $\pi(\mu)$ increases with $\mu$ and so we want $\alpha = \pi(\mu_0)$. Then*

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = z_\alpha.$$

*where $\Phi(z_\alpha) = 1 - \alpha$ and $\Phi$ is the distribution function for the standard normal, thus $k(\mu_0) = \mu_0 + (\sigma/\sqrt{n})z_\alpha$.*

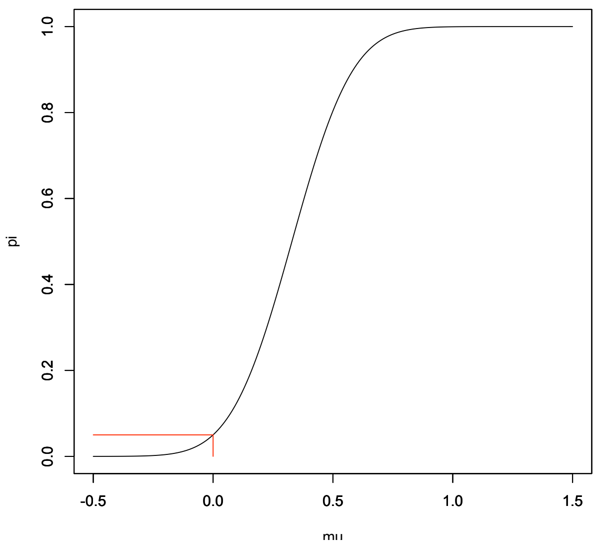


**Figure 1:** Power function for the one-sided test above. Note that $\pi(0) = 0.05$ as shown by the red lines is the level of the test.

*The power function for this test*

$$\pi(\mu) = P_\mu\{\bar{X} \geq \frac{\sigma}{\sqrt{n}}z_\alpha + \mu_0\} = P_\mu\{\bar{X} - \mu \geq \frac{\sigma}{\sqrt{n}}z_\alpha - (\mu - \mu_0)\}$$
$$= P_\mu\left\{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq z_\alpha - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right\} = 1 - \Phi\left(z_\alpha - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$$

*We plot the power function with* $\mu_0 = 0$, $\sigma = 1$, *and* $n = 25$,

```
> zalpha=qnorm(.95)
> mu<-(-50:150)/100
> z=zalpha-5*mu
> pi=1-pnorm(z)
> plot(mu,pi,type="l")
```

*For a* **two-sided test**

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0.$$

*We reject* $H_0$ *if* $|\bar{X} - \mu_0|$ *is too large. Again, to obtain level* $\alpha$,

$$|Z| = \left|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right| = z_{\alpha/2}.$$

*The power function for the test*

$$\pi(\mu) = 1 - P_\mu\left\{-z_{\alpha/2} \leq \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right\} = 1 - P_\mu\left\{-z_{\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right\}$$
$$= 1 - \Phi\left(z_{\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$$

```
> zalpha = qnorm(.975)
> mu=(-200:200)/100
> pi = 1 - pnorm(zalpha-5*mu)+pnorm(-zalpha-5*mu)
> plot(mu,pi,type="l")
```

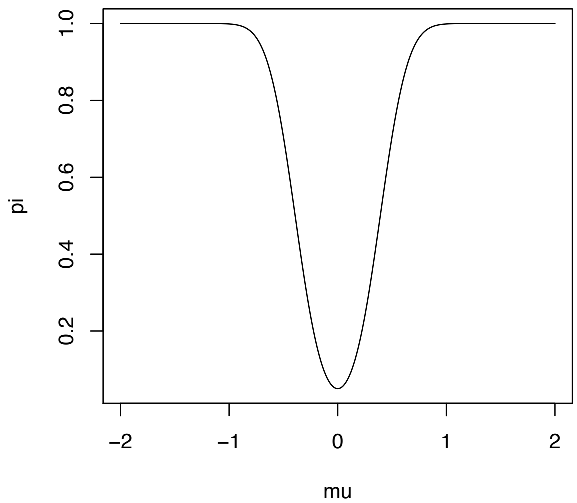


**Figure 2:** Power function for the two-sided test above

## 2 The $p$-value

The report of *reject* the null hypothesis does not describe the strength of the evidence because it fails to give us the sense of whether or not a small change in the values in the data could have resulted in a different decision. Consequently, the common method is not to choose, in advance, a significance level $\alpha$ of the test and then report "reject" or "fail to reject", but rather to report the value of the test statistic and to give all the values for $\alpha$ that would lead to the rejection of $H_0$. The $p$-value is the probability of obtaining a result at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.

If the $p$-value is below a given significance level $\alpha$, then we say that the result is **statistically significant** at the level $\alpha$.

For example, if the test is based on having a test statistic $S(X)$ exceed a level $k$, i.e., we have decision

$$\text{reject if and only if } S(X) \geq k.$$

and if the value $S(X) = k_0$ is observed, then the $p$-value equals

$$\max\{\pi(\theta); \theta \in \Theta_0\} = \max\{P_\theta\{S(X) \geq k_0\}; \theta \in \Theta_0\}.$$

In the one-sided test above, if $\bar{X} = 1$, then

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{1-0}{1/\sqrt{25}} = 5.$$

```
> pvalue = 1 - pnorm(5)
> pvalue
[1] 2.866516e-07
```

In this case, the $p$-value is $2.87 \times 10^{-7}$.

**Example 2.** *For $X_1, X_2, \ldots, X_n$ independent $U(0,\theta)$ random variables, $\theta \in \Theta \in [0,\infty)$. Take*

$$H_0 : \theta_L \leq \theta \leq \theta_R \quad \text{versus} \quad H_1 : \theta < \theta_L \text{ or } \theta > \theta_R.$$

*We will try to base a test based on the statistic $X_{(n)} = \max_{1 \leq i \leq n} X_i$ and reject $H_0$ if $X_{(n)} > \theta_R$ and too much smaller that $\theta_L$, say $\tilde{\theta}$. Then, the power function*

$$\pi(\theta) = P_\theta\{X_{(n)} \leq \tilde{\theta}\} + P_\theta\{X_{(n)} \geq \theta_R\}$$

*We compute the power function in three cases.*

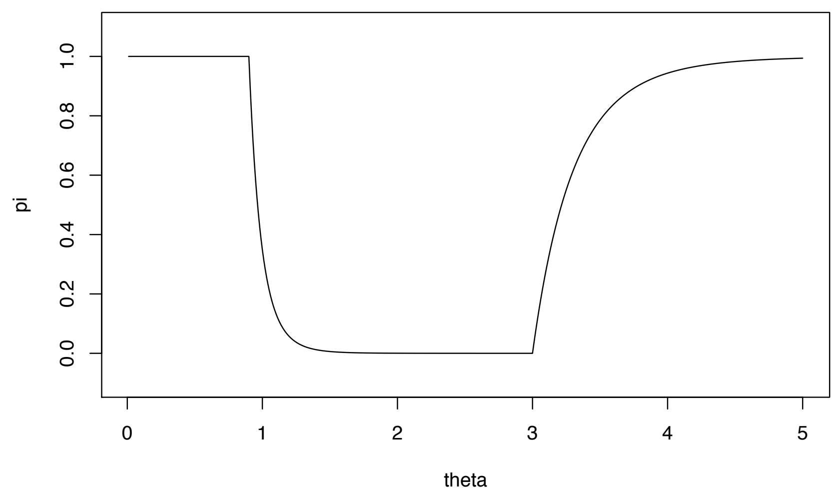


**Figure 3:** Power function for the test above with $\theta_L = 1, \theta_R = 3, \tilde{\theta} = 0.9$, and $n = 10$. The power of the test is $\pi(1) = 0.3487$.

**Case 1**. $\theta \leq \tilde{\theta}$.

$$P_\theta\{X_{(n)} \leq \tilde{\theta}\} = 1 \textit{ and } P_\theta\{X_{(n)} \geq \theta_R\} = 0$$

and therefore $\pi(\theta) = 1$.

**Case 2**. $\tilde{\theta} < \theta \leq \theta_R$.

$$P_\theta\{X_{(n)} \leq \tilde{\theta}\} = \left(\frac{\tilde{\theta}}{\theta}\right)^n \textit{ and } P_\theta\{X_{(n)} \geq \theta_R\} = 0$$

and therefore $\pi(\theta) = (\tilde{\theta}/\theta)^n$.

**Case 3**. $\theta > \theta_R$.

$$P_\theta\{X_{(n)} \leq \tilde{\theta}\} = \left(\frac{\tilde{\theta}}{\theta}\right)^n \textit{ and } P_\theta\{X_{(n)} \geq \theta_R\} = 1 - \left(\frac{\theta_R}{\theta}\right)^n$$

and therefore $\pi(\theta) = (\tilde{\theta}/\theta)^n + 1 - (\theta_R/\theta)^n$.

The size of the test

$$\alpha = \max\left\{\left(\frac{\tilde{\theta}}{\theta}\right)^n; \theta_L \le \theta \le \theta_R\right\} = \left(\frac{\tilde{\theta}}{\theta_L}\right)^n.$$

To achieve this level, choose $\tilde{\theta} = \theta_L \sqrt[n]{\alpha}$.

# 3 Confidence Intervals

We have seen three methods for estimation of a population parameter, unbiased estimation, method of moments, and maximum moment estimation, are called **point estimates**. Often, we augment this point estimate with an **interval estimate** called a **confidence interval**. This interval comes with a probability $\gamma$ that the interval contains the population parameter. This value is called the **confidence** level, generally stated as a percent. Thus, a "95% confidence interval". Consequently, for a given estimation procedure in a given situation, the higher the confidence level, the wider the confidence interval will be.

The calculation of a confidence interval generally requires assumptions about the nature of the estimation process it is primarily a parametric method for example, it may depend on an assumption that the distribution of the population from which the sample came is normal or it may invoke the central limit theorem if the sample size is sufficiently large.

In practice, we choose a number $\gamma$ between 0 and 1. From data $X$, we compute two statistics $L(X)$ and $R(X)$ so that **irrespective** of the value of the parameter,

$$P_\theta\{L(X) < \theta < R(X)\} \ge \gamma.$$

If $\ell = L(X)$ and $r = R(X)$ are the observed values based on the data $X$, then the interval $(\ell, r)$ is the confidence interval for $\theta$ with confidence level $\gamma$.

For example, if $\hat{\theta}$ is a maximum likelihood estimator for $\theta$ based on a random sample $X_1, X_2 \cdots, X_n$ and $n$ is large, then $\hat{\theta}$ is approximately normally distributed and

$$Z_n = \frac{\hat{\theta} - \theta}{1/\sqrt{nI(\theta)}}$$

is approximately a standard normal. We approximate $I(\theta)$ the Fisher information for one observation by the observed information $I(\hat{\theta})$.

Write $\alpha = 1 - \gamma$ and pick $z_{\alpha/2}$ to be the upper tail probability. Then, we can choose

$$\ell = \hat{\theta} - \frac{z_{\alpha/2}}{\sqrt{nI(\hat{\theta})}} \quad \text{and} \quad r = \hat{\theta} + \frac{z_{\alpha/2}}{\sqrt{nI(\hat{\theta})}}.$$

Often $\gamma$-level confidence interval are complementary to $\alpha = 1 - \gamma$ level two-sided hypothesis tests. For example, for the two sided test above, we reject fail to reject $H_0$ if and only if

$$|Z| = \left|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}.$$

we can rewrite this to have

$$\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \mu_0 < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}.$$

Notice that no matter what value $\mu$ is the true state of nature, we have

$$P_\mu\{\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\} = 1 - \alpha.$$

Then given the observed value $\bar{x}$ from the data, then we call the interval

$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \mu < \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$$

a **level $\gamma$ confidence interval** for the parameter $\mu$. This interval is symmetric about $\mu$ and the distance

$$m = \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$$

from the center is called the **margin of error**. Note that we fail to reject $H_0$ at the significance level $\alpha$ if and only if $\mu_0$ is in the $\gamma = 1 - \alpha$ confidence interval.

An $\alpha$ level test for the hypothesis

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0$$

generates a $\gamma = 1 - \alpha$ confidence interval.

Given the data $X$, let $\omega(X)$ denote those parameter values for which the test fails to reject this hypothesis. Thus,

$$\theta_0 \in \omega(X) \text{ if and only if fail to reject } H_0.$$

and

$$P_{\theta_0}\{\theta_0 \in \omega(X)\} = P_{\theta_0}\{\text{fail to reject } H_0\} = 1 - P_{\theta_0}\{\text{reject } H_0\} = 1 - \alpha = \gamma.$$

Now let $L(X)$ and $R(X)$ be the end points of the interval $\omega(X)$.

**Example 3.** *For Bernoulli trials, several refinements have been made in the margin the nature of confidence intervals. The latest is due to Agresti and Cluil in 1998.*

*Let $x$ be the number of successes in $n$ Bernoulli trials. Consider the two statistics*

$$\hat{p} = \frac{x}{n} \quad \textit{and} \quad \hat{p} = \frac{x+2}{n+4}.$$

*The first is both the maximum likelihood and the unbiased estimator for $p$. The second is an adjustment that comes from adding 4 additional trials that have 2 successes.*

*In this case, the endpoints of the $\gamma$ level confidence interval are*

$$\hat{p} \pm z_{(1-\gamma)/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}.$$