9.4 Optimal Tests for simple hypotheses

- Null hypothesis $\mathcal{H}_0 : f = f_0$
- Alternate hypothesis $\mathcal{H}_A : f = f_1$

We want to find a rejection region $R$ such that the error of both types are as small as possible.

$$\int_R f_0(x) \, dx = \alpha \quad \text{and} \quad 1 - \beta = \int_R f_1 \, dx$$

Neyman-Pearson Lemma

**Theorem 9.1** For testing $f_0(x)$ against $f_1(x)$ a critical region of the form

$$\Lambda(x) = \frac{f_0(x)}{f_1(x)} < K$$

where $K$ is constant has the greatest power (smallest $\beta$) in the class of tests with the same $\alpha$

$$\beta_R - \beta_S = \left(\int_R - \int_S\right) f = \int_{R \cap S^c} f - \int_{S \cap R^c} f$$

Since in $R$, $f_1 \geq f_0/K$ and in $R^c$, $-f_1 \geq -f_0/K$ we have:

$$\beta_S - \beta_R \geq \frac{1}{K} \left(\int_{R \cap S^c} f_0 - \int_{S \cap R^c} f_0\right)$$

$$= \frac{1}{K} \left(\int_R f_0 - \int_S f_0\right) = \frac{1}{K} \left(\alpha_R - \alpha_S\right)$$

Example: Two simple hypotheses about Bernouilli parameter:

$\mathcal{H}_0 : p = 0.5 \quad \mathcal{H}_1 : p = 0.25$

<table>
<thead>
<tr>
<th>$y$</th>
<th>$f_0(y)$</th>
<th>$f_1(y)$</th>
<th>$\Lambda(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.0312</td>
<td>.2373</td>
<td>.13</td>
</tr>
<tr>
<td>1</td>
<td>.1562</td>
<td>.3955</td>
<td>.40</td>
</tr>
<tr>
<td>2</td>
<td>.3125</td>
<td>.2637</td>
<td>1.2</td>
</tr>
<tr>
<td>3</td>
<td>.3125</td>
<td>.0879</td>
<td>3.6</td>
</tr>
<tr>
<td>4</td>
<td>.1562</td>
<td>.0146</td>
<td>10.6</td>
</tr>
<tr>
<td>5</td>
<td>.0312</td>
<td>.0010</td>
<td>31.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$R$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>${}$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>${0}$</td>
<td>.031</td>
<td>.763</td>
<td>.237</td>
</tr>
<tr>
<td>${0,1}$</td>
<td>.187</td>
<td>.367</td>
<td>.633</td>
</tr>
<tr>
<td>${0,1,2}$</td>
<td>.500</td>
<td>.103</td>
<td>.897</td>
</tr>
</tbody>
</table>
9.4.1 Neyman Pearson Test for $\mathcal{N}(0, \theta)$

Given a random sample from $X$ the likelihood is

$$L(\theta) = \theta^{-n/2} \exp \left(-\frac{1}{2\theta} \sum X_i^2\right)$$

For the NP likelihood ratio test the rejection region for testing $\theta_1$ against $\theta_2$ is of the form:

$$\Lambda = \left(\frac{\theta_1}{\theta_2}\right)^{-n/2} \exp \left(-\frac{1}{2\theta_1} - \frac{1}{2\theta_2} \right) \sum X_i^2 < K'$$

If $\theta_1 < \theta_2$ this inequality holds for a given $K'$ if and only if, for some $K$, $\sum X_i^2 > K$

The distribution of $\sum X_i^2/\theta$ is $\chi^2_n$ and we can thus find the error sizes in this case:

$$\alpha = P(\sum X_i^2 > K|\theta_1) = 1 - F_{\chi^2_n}(\frac{K}{\theta_1})$$

$$\beta = P(\sum X_i^2 < K|\theta_2) = F_{\chi^2_n}(\frac{K}{\theta_2})$$

9.5 Composite Hypotheses

Typically one or both hypotheses are composite, suppose $\mathcal{H}_A$ is composite, a test that is most powerful for all simple hypotheses contained in $\mathcal{H}_A$ is called uniformly most powerful.

Example:

Suppose we test $\mathcal{H}_0 : p = 0.5$ based on $Y$, the number of 1's in a random sample of size $n = 5$. We saw that rejecting $\mathcal{H}_0$ for $Y < K$. We saw that rejecting $\mathcal{H}_0$ for $Y < K$ is most powerful against $p = 0.25$ in the class of tests with $\alpha = 0.187$. But the same is true if we were to use an arbitrary $p < 0.5$. The likelihood ratio is:

$$\Lambda = \frac{.5^Y.5^{5-Y}}{p^Y(1-p)^{5-Y}} \propto \left(\frac{1}{p} - 1\right)^Y$$

An increasing function of $Y$ when $p < \frac{1}{2}$. So, $\Lambda$ is less than a constant when $Y$ is less than a constant. Thus $Y < K$ is most powerful against every $p < .5$; it is UMP against $\mathcal{H}_A : p < .5$.

Note:

The rejection region $\{0, 1\}$ for $Y$ is UMP for $p = 0.5$ against $p < 0.5$, in the class of tests with $\alpha = 0.187$. The power function is

$$\pi(p) = P(Y = 0 \text{ or } 1|p) = (1 - p)^5 + 5p(1 - p)^4 = (1 - p)^4(1 + 4p)$$

This is a decreasing function of $p$ on $(0,1)$, so its largest value on $p \geq .5$ is at $p = 0.5$. So $\alpha^* = \pi(.5) = \max \alpha$'s.

Since $Y \leq 1$ has the greatest power at any alternative $p < .5$ among tests with $\alpha \leq .187$ it has greatest power at any alternative among tests with $\alpha^* = .187$, it is UMP for $p \geq .5$ against $p < .5$. 
9.5.1 Generalized Maximum Likelihood Ratio Tests

The simple hypothesis likelihood ration, orders the sample points according to the value of $\Lambda$, the regions $\Lambda(x)$ are most powerful.

Suppose we want to test the null $\lambda = 1$ in an $\text{Exp}(\lambda)$ against $\lambda \neq 1$, given a set of $n$ observations, the likelihood can be written as a function of the sufficient statistic $\bar{X}$ as

$$L(\lambda) = \lambda^n \exp(-n\bar{X})$$

The likelihood of $H_0$ is $L(1) = \exp(-n\bar{X})$ but there are values of $\lambda$ with greater likelihoods, the maximum being attained at $\hat{\lambda} = \frac{1}{\bar{X}}: L(\hat{\lambda}) = (e\bar{X})^{-n}$, the ratio of these likelihoods is

$$\frac{L(1)}{L(\hat{\lambda})} = (e\bar{X} e^{-\bar{X}})^n$$

For extreme values of this ratios we will reject the null hypothesis.

For a family of distributions indexed by $\theta$, take as $\Omega(H_0)$ a subset of parameter space, the alternative is the complement. Take $\hat{\theta}_0$ that maximises the likelihood within $H_0$ and $\hat{\theta}$ the ordinary MLE.

$$L(\hat{\theta}_0) = \sup_{\Omega(H_0)} L(\theta), \quad L(\hat{\theta}) = \sup_{\Omega(H_0) \cup \Omega(H_A)} L(\theta)$$

The generalized likelihood ratio is then

$$\Lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$$

$\Lambda < 1$ and $\Lambda = 1$ if $\hat{\theta}_0 = \hat{\theta}$, when the ratio is small the explanation of data given by $H_0$ is much worse than the best available solution, so regions of the form $\Lambda < K$ will be our rejection regions.

**Theorem 9.2** Under smoothness conditions on the pdf, the null distribution of $-2\log \Lambda$ tends to a $\chi^2_k$ distribution, with $k = \text{number of free parameters overall} - \text{number of free parameters in } H_0$.

$$k = \text{dim} (\Omega(H_A) \cap \Omega(H_0)) - \text{dim} (\Omega(H_0))$$

Example: Multinomial

$H_0 : p = p(\theta) \in \omega_0$

$H_A : \Omega = \{p_1, p_2, \ldots, p_n = 1, p_1 + p_2 + \cdots + p_n = 1\}$

Numerator:

$$\max_{p \in \omega_0} \left( \frac{n!}{x_1! \cdots x_m!} \right) p_{x_1}^{x_1}((\theta) p_2^{x_2}((\theta) \cdots p_m^{x_m}(\theta)$$

, we saw that the maximum will be attained at $(p_1, p_2, \ldots, p_n)_{i}(\hat{\theta})$, and the MLE is $\hat{p}_i = \frac{x_i}{n}$.

$$\Lambda = \prod_{i=1}^m \left( \frac{p_i(\hat{\theta})}{\hat{p}_i} \right)^{x_i}$$

$$-2\log \Lambda = 2 \sum x_i \log \left( \frac{\hat{p}_i}{p_i(\hat{\theta})} \right) = 2 \sum O_i \log \left( \frac{O_i}{E_i} \right)$$

For large samples, we have $-2\log \Lambda \sim \chi^2_{m-1-k}$, where $k$ is the dimension of $\omega_o$. 

3
9.6 Duality with Confidence Intervals

If we know the distribution of a statistic does’t depend on the parameter \( \theta \), we can use it to build confidence intervals, it’s called a pivotal, and comes useful for confidence interval building.

Definition:
A pivotal statistic is a quantity which is a function of the statistics and parameters of interest whose distribution does not depend on the parameter of interest.

**Theorem 9.3** If we have a test for every \( \theta_0 \) in \( \Theta \) at level \( \alpha \) of \( H_0 : \theta = \theta_0 \), call the acceptance region of the test \( A(\theta_0) \), then the set \( C(X) = \{ \theta : X \in A(\theta) \} \) it is a 100(1 - \( \alpha \)) % CI for \( \theta \).

A 100(1 - \( \alpha \)) % CI for \( \theta \) consists of all the values of \( \theta_0 \) for which the hypothesis \( \theta = \theta_0 \) will not be rejected at level \( \alpha \).

Conversely the hypothesis \( \theta = \theta_0 \) is accepted if \( \theta_0 \) lies in the CI.

9.7 Bayesian Testing

\( \ell \) is defined as the loss function, no loss for the correct action, \( A = \ell(H_0, \text{rej.} H_0) \), \( B = \ell(H_A, \text{acc.} H_0) \), the Bayes losses are the expected values with regards to the posterior distributions.

\[
g_0 = P(H_0) \quad g_1 = P(H_A) = 1 - g_0
\]

Posterior probabilities:

\[
h(H_0 | X) = \frac{g_0 f_0(X)}{g_0 f_0(X) + g_1 f_1(X)} \quad h(H_A | X) = \frac{g_1 f_1(X)}{g_0 f_0(X) + g_1 f_1(X)}
\]

Bayes losses:

\[
B(\text{rej}|X) = A \times h(H_0 | X) + 0 \times h(H_A | X)
\]

\[
B(\text{acc}|X) = 0 \times h(H_0 | X) + B \times h(H_A | X)
\]

The ratio of the Bayes loss for rejecting to the Bayes loss for accepting is:

\[
\frac{B(\text{rej} H_0 | X)}{B(\text{acc} H_0 | X)} = \frac{A g_0 f_0(X)}{B g_1 f_1(X)}
\]

Rejecting \( H_0 \) is the better action when the ratio is less than 1:

\[
\Lambda = \frac{f_0(X)}{f_1(X)} < K = \frac{B g_1}{A g_0}
\]