# 8 Generalized likelihood ratio tests

*An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem.*

## 8.1 The $\chi^2$ distribution

This distribution plays a huge role in Statistics. For $n$ a positive integer, the $\chi^2_n$ distribution is the distribution of

$$X_1^2 + \cdots + X_n^2, \text{ where } X_1, \ldots, X_n \text{ are IID each } N(0,1).$$

It is not difficult to show that

$$\chi^2_n = \text{gamma}\left(\tfrac{1}{2}n, \tfrac{1}{2}\right),$$

with pdf

$$f(t) = \left(\tfrac{1}{2}\right)^{n/2} t^{n/2-1} e^{-t/2} \Big/ \Gamma(n/2), \quad t > 0.$$

We speak of the 'chi-squared distribution with (or on) $n$ **degrees of freedom**'. If $X \sim \chi^2_n$, $\mathbb{E}(X) = n$, $\text{var}(X) = 2n$.

## 8.2 Generalised likelihood ratio tests

Tests with critical regions of the form $C = \{x : L_x(H_0, H_1) > k\}$ are intuitively sensible, and in certain cases optimal. But such a critical region may not reduce to a region depending only on the value of a simple function $T(x)$. Even if it does, the distribution of the statistic $T(X)$, required to determine $k$, may not be simple. However, the following theorem allows us to use a likelihood ratio test even in this case. We must take care in describing the circumstances for which the theorem is valid.

So far we have considered **disjoint** alternatives, but if our real interest is in testing $H_0$ and we are not interested in any specific alternative it is simpler to take (in the parametric framework) $\Theta_1 = \Theta$, rather than $\Theta_1 = \Theta \setminus \Theta_0$.

So we now suppose we are testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta$, i.e., a null hypothesis which restricts $\theta$ against a general alternative.

Suppose that with this formulation $\Theta_0$ imposes $p$ independent restrictions on $\theta$, so that, for example, we have $\Theta = \{\theta : \theta = (\theta_1, \ldots, \theta_k)\}$ and

$$\begin{aligned}
&H_0 : \theta_{i_1} = \alpha_1, \ldots, \theta_{i_p} = \alpha_p, \quad \text{for given } \alpha_j\text{; or}\\
&H_0 : A\theta = b, \quad \text{for given } A_{p\times k}, b_{p\times 1}\text{; or}\\
&H_0 : \theta_i = \theta_i(\phi_1, \ldots, \phi_{k-p}), \quad i = 1, \ldots, k, \quad \text{for given } \theta_1(\cdot), \ldots, \theta_k(\cdot)\\
&\qquad\qquad \text{and } \phi_1, \ldots, \phi_{k-p} \text{ to be estimated.}
\end{aligned}$$



$\Theta_1$ has $k$ free parameters and $\Theta_0$ has $k - p$ free parameters. We write $|\Theta_1| = k$ and $|\Theta_0| = k - p$. Then we have the following theorem (not to be proved.)

**Theorem 8.1** *Suppose $\Theta_0 \subset \Theta_1$ and $|\Theta_1| - |\Theta_0| = p$. Then under certain conditions, as $n \to \infty$ with $X = (X_1, \ldots, X_n)$ and $X_i$ IID,*

$$2 \log L_X(H_0, H_1) \sim \chi^2_p,$$

*if $H_0$ is true. If $H_0$ is not true, $2 \log L_X$ tends to be larger. We reject $H_0$ if $2 \log L_x > c$, where $\alpha = \mathbb{P}\left(\chi^2_p > c\right)$ to give a test of size approximately $\alpha$.*

We say that $2 \log L_X(H_0, H_1)$ is asymptotically distributed as $\chi^2_p$. The conditions required by the theorem hold in all the circumstances we shall meet in this course.

**Lemma 8.2** *Suppose $X_1, \ldots, X_n$ are IID $N(\mu, \sigma^2)$. Then*

(i) $\max_\mu f(x \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\sum_i (x_i - \bar{x})^2 / 2\sigma^2\right]$.

(ii) $\max_{\sigma^2} f(x \mid \mu, \sigma^2) = \left[2\pi \frac{\sum_i (x_i - \mu)^2}{n}\right]^{-n/2} \exp[-n/2]$.

(iii) $\max_{\mu,\sigma^2} f(x \mid \mu, \sigma^2) = \left[2\pi \frac{\sum_i (x_i - \bar{x})^2}{n}\right]^{-n/2} \exp[-n/2]$.

## 8.3 Single sample: testing a given mean, known variance ($z$-test)

Let $X_1, \ldots, X_n$ be IID $N(\mu, \sigma^2)$, where $\sigma^2$ is known. We wish to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. The generalized likelihood ratio test suggests that we should reject $H_0$ if $L_x(H_0, H_1)$ is large, where

$$\begin{aligned}
L_x(H_0, H_1) &= \frac{\sup_\mu f\left(x \mid \mu, \sigma^2\right)}{f\left(x \mid \mu_0, \sigma^2\right)}\\
&= \frac{\left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\sum_i (x_i - \bar{x})^2 / 2\sigma^2\right]}{\left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\sum_i (x_i - \mu_0)^2 / 2\sigma^2\right]}\\
&= \exp\left[(1/2\sigma^2) \sum_{i=1}^n \left\{(x_i - \mu_0)^2 - (x_i - \bar{x})^2\right\}\right]\\
&= \exp\left[(1/2\sigma^2) n (\bar{x} - \mu_0)^2\right]
\end{aligned}$$

That is, we should reject $H_0$ if $(\bar{x} - \mu_0)^2$ is large.

This is no surprise. For under $H_0$, $\bar{X} \sim N(\mu_0, \sigma^2/n)$, so that

$$Z = \sqrt{n}(\bar{X} - \mu_0)/\sigma \sim N(0, 1),$$

and a test of size $\alpha$ is to reject $H_0$ if $z > z_{\alpha/2}$ or if $z < -z_{\alpha/2}$, where $z_{\alpha/2}$ is the 'upper $\alpha/2$ point of $N(0,1)$' i.e., the point such that $\mathbb{P}(N(0,1) > z_{\alpha/2}) = \alpha/2$. This is an example of a **two-tailed test**.

Note that $2 \log L_X(H_0, H_1) = Z^2 \sim \chi_1^2$. In this example $H_0$ imposes $p = 1$ constraint on the parameter space and the approximation in Theorem 8.1 is exact.

### 8.4 Single sample: testing a given variance, known mean ($\chi^2$-test)

As above, let $X_1, \ldots, X_n$ be IID $N(\mu, \sigma^2)$, where $\mu$ is known. We wish to test $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$. The generalized likelihood ratio test suggests that we should reject $H_0$ if $L_x(H_0, H_1)$ is large, where

$$L_x(H_0, H_1) = \frac{\sup_{\sigma^2} f\left(x \mid \mu, \sigma^2\right)}{f\left(x \mid \mu, \sigma_0^2\right)} = \frac{\left[\frac{2\pi \sum_i (x_i - \mu)^2}{n}\right]^{-n/2} \exp\left[-n/2\right]}{(2\pi\sigma_0^2)^{-n/2} \exp\left[-\sum_i (x_i - \mu)^2 / 2\sigma_0^2\right]} .$$

If we let $t = \sum_i (x_i - \mu)^2 / n\sigma_0^2$ we find

$$2 \log L_x(H_0, H_1) = n(t - 1 - \log t) ,$$

which increases as $t$ increases from 1 and $t$ decreases from 1. Thus we should reject $H_0$ when the difference of $t$ and 1 is large.

Again, this is not surprising, for under $H_0$,

$$T = \sum_{i=1}^{n} (X_i - \mu)^2 / \sigma_0^2 \sim \chi_n^2 .$$

So a test of size $\alpha$ is the two-tailed test which rejects $H_0$ if $t > F_{\alpha/2}^{(n)}$ or $t < F_{1-\alpha/2}^{(n)}$ where $F_{1-\alpha/2}^{(n)}$ and $F_{\alpha/2}^{(n)}$ are the lower and upper $\alpha/2$ points of $\chi_n^2$, i.e., the points such that $\mathbb{P}\left(\chi_n^2 < F_{1-\alpha/2}^{(n)}\right) = \mathbb{P}\left(\chi_n^2 > F_{\alpha/2}^{(n)}\right) = \alpha/2$.

### 8.5 Two samples: testing equality of means, known common variance ($z$-test)

Let $X_1, \ldots, X_m$ be IID $N(\mu_1, \sigma^2)$ and let $Y_1, \ldots, Y_n$ be IID $N(\mu_2, \sigma^2)$, and suppose that the two samples are independent. It is required to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. The likelihood ratio test is based on

$$\begin{aligned}
L_x(H_0, H_1) &= \frac{\sup_{\mu_1, \mu_2} f\left(x \mid \mu_1, \sigma^2\right) f\left(y \mid \mu_2, \sigma^2\right)}{\sup_{\mu} f\left(x \mid \mu, \sigma^2\right) f\left(y \mid \mu, \sigma^2\right)} \\
&= \frac{(2\pi\sigma^2)^{-(m+n)/2} \exp\left[-\sum_i (x_i - \bar{x})^2 / 2\sigma^2\right] \exp\left[-\sum_i (y_i - \bar{y})^2 / 2\sigma^2\right]}{(2\pi\sigma^2)^{-(m+n)/2} \exp\left[-\sum_i \left(x_i - \frac{m\bar{x} + n\bar{y}}{m+n}\right)^2 / 2\sigma^2\right] \exp\left[-\sum_i \left(y_i - \frac{m\bar{x} + n\bar{y}}{m+n}\right)^2 / 2\sigma^2\right]} \\
&= \exp\left[\frac{m}{2\sigma^2} \left(\bar{x} - \frac{m\bar{x} + n\bar{y}}{m+n}\right)^2 + \frac{n}{2\sigma^2} \left(\bar{y} - \frac{m\bar{x} + n\bar{y}}{m+n}\right)^2\right] \\
&= \exp\left[\frac{1}{2\sigma^2} \frac{mn}{(m+n)} (\bar{x} - \bar{y})^2\right]
\end{aligned}$$



So we should reject $H_0$ if $|\bar{x} - \bar{y}|$ is large. Now, $\bar{X} \sim N(\mu_1, \sigma^2/m)$ and $\bar{Y} \sim N(\mu_2, \sigma^2/n)$, and the samples are independent, so that, on $H_0$,

$$\bar{X} - \bar{Y} \sim N\left(0, \sigma^2 \left(\frac{1}{m} + \frac{1}{n}\right)\right)$$

or

$$Z = (\bar{X} - \bar{Y}) \left(\frac{1}{m} + \frac{1}{n}\right)^{-\frac{1}{2}} \frac{1}{\sigma} \sim N(0, 1).$$

A size $\alpha$ test is the two-tailed test which rejects $H_0$ if $z > z_{\alpha/2}$ or if $z < -z_{\alpha/2}$, where $z_{\alpha/2}$ is, as in 8.3 the upper $\alpha/2$ point of $N(0, 1)$. Note that $2 \log L_X(H_0, H_1) = Z^2 \sim \chi_1^2$, so that for this case the approximation in Theorem 8.1 is again exact.

### 8.6 Goodness-of-fit tests

Suppose we observe $n$ independent trials and note the numbers of times that each of $k$ possible outcomes occurs, i.e., $(x_1, \ldots, x_k)$, with $\sum_j x_j = n$. Let $p_i$ be the probability of outcome $i$. On the basis of this data we want to test the hypothesis that $p_1, \ldots, p_k$ take particular values. We allow that these values might depend on some unknown parameter $\theta$ (or parameters if $\theta$ is a vector). I.e., we want to test

$$H_0 : \; p_i = p_i(\theta) \text{ for } \theta \in \Theta_0 \quad \text{against} \quad H_1 : \; p_i \text{ are unrestricted.}$$

For example, $H_0$ might be the hypothesis that the trials are samples from a binomial distribution $B(k, \theta)$, so that under $H_0$ we would have $p_i(\theta) = \binom{k}{i} \theta^i (1 - \theta)^{k-i}$.

This is called a **goodness-of-fit test**, because we are testing whether our data fit a particular distribution (in the above example the binomial distribution).

The distribution of $(x_1, \ldots, x_k)$ is the **multinomial distribution**

$$\mathbb{P}(x_1, \ldots, x_k \mid p) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} ,$$

for $(x_1, \ldots, x_k)$ s.t. $x_i \in \{0, \ldots, n\}$ and $\sum_{i=1}^k x_i = n$. Then we have

$$\sup_{H_1} \log f(x) = \text{const} + \sup \left\{ \sum_{i=1}^{k} x_i \log p_i \;\middle|\; 0 \leq p_i \leq 1, \; \sum_{i=1}^{k} p_i = 1 \right\} .$$

Now, $\sum_i x_i \log p_i$ may be maximised subject to $\sum_i p_i = 1$ by a Lagrangian technique and we get $\hat{p}_i = x_i / n$. Likewise,

$$\sup_{H_0} \log f(x) = \text{const} + \sup_{\theta} \left\{ \sum_{i=1}^{k} x_i \log p_i(\theta) \right\} .$$

The generalized likelihood tells us to reject $H_0$ if $2L_x(H_0, H_1)$ is large compared to the chi-squared distribution with d.f. $|\Theta_1| - |\Theta_0|$. Here $|\Theta_1| = k - 1$ and $|\Theta_0|$ is the number of independent parameters to be estimated under $H_0$.