Another broadly applicable class of tests is the likelihood ratio test (LRT). Let

$$T(X) = \frac{\sup_{\theta \in \Theta} \mathcal{L}_n(\theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}_n(\theta)}$$

If $T(X)$ is large, it means there are values of $\theta$ in $\Theta_1$ which are larger than for any in $\Theta_0$. A likelihood ratio test is a test for which

$$R = \{x : T(x) > c\}$$

If $\hat{\theta}_n$ is the MLE and $\hat{\theta}_{n,0}$ is the MLE restricting $\theta \in \Theta_0$, then

$$T(X) = \frac{\mathcal{L}_n(\hat{\theta}_n)}{\mathcal{L}_n(\hat{\theta}_{n,0})}$$

Sometimes we can calculate the power function for the LRT exactly.

Example: Suppose $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$. Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Find $T(X)$ and find a simplified expression for the form of the rejection region. Use it to find the size $\alpha$ LRT.

When the power function can not be calculated exactly, and $\Theta_0$ consists of fixing certain elements of $\theta$ (e.g., as in a point-null hypothesis), we can use the limiting distribution

$$\lambda(X) = 2 \log T(X) \xrightarrow{D} \chi^2_{r-q}$$

where $r$ is the dimension of $\theta$ and $q$ is the number of restricted elements.

Aside: The $\chi^2_k$ distribution (read "chi squared with $k$ degrees of freedom") is the distribution of the sum of squares of $k$ independent standard normal random variables. That is, if $Z_1, \ldots, Z_k \overset{iid}{\sim} N(0,1)$, then

$$Y = \sum_{i=1}^{k} Z_i^2 \sim \chi^2_k$$

We can use this approximation to find an appropriate critical value.

Example: Suppose $X_1, \ldots, X_n \overset{iid}{\sim} Poisson(\theta)$. Let $\hat{\theta}_n = \sum_{i=1}^n X_i/n$ be the MLE for $\theta$. For testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, we have

$$
\begin{aligned}
\lambda &= 2 \log \frac{\mathcal{L}(\hat{\theta}_n)}{\mathcal{L}(\theta_0)} \\
&= 2 \log \frac{e^{-n\hat{\theta}_n} \hat{\theta}_n^{\sum x_i}}{e^{-n\theta_0} \theta_0^{\sum x_i}} \\
&= 2n[(\theta_0 - \hat{\theta}_n) - \hat{\theta}_n \log(\theta_0/\hat{\theta}_n)]
\end{aligned}
$$

Since for large $n$, $\lambda \overset{D}{\approx} \chi_1^2$, to construct an approximate size $\alpha$ LRT, we find $\chi_{1,\alpha}^2$ s.t. $P(\chi_1^2 < \chi_{1,\alpha}^2) = 1 - \alpha$ and reject $H_0$ if $\lambda > \chi_\alpha^2$.

Suppose that for every $\alpha \in (0, 1)$ we have a size $\alpha$ test with rejection region $R_\alpha$. Then

$$\text{p-value} = \inf\{\alpha : T(X) \in R_\alpha\}$$

That is, the p-value is the smallest level at which we can reject $H_0$.

When $R_\alpha = \{x : T(x) \geq c_\alpha\}$,

$$\text{p-value} = \sup_{\theta \in \Theta_0} P_\theta(T(X) \geq T(x))$$

where $x$ is the observed data.

Therefore, the p-value is the probability under $H_0$ of observing a value $T(X)$ the same as or more extreme than what was actually observed.

In the case of the Wald test, the (approximate) p-value is

$$\text{p-value} = P_{\theta_0}(|W| > |w|) \approx P(|Z| > |w|) = 2\Phi(-|w|)$$

where $w$ is the observed value of the statistic and $Z \sim N(0, 1)$.

In the case of the LRT with point null hypothesis and limiting $\chi^2_{r-q}$ distribution, the (approximate) p-value is

$$\text{p-value} = P_{\theta_0}(\lambda(X) > \lambda(x)) \approx P(\chi^2_{r-q} > \lambda(x))$$

Theorem: If the test statistic has a continuous distribution, then under $H_0 : \theta = \theta_0$, the p-value has a $Unif(0,1)$ distribution. Therefore, if we reject $H_0$ when the p-value is less than $\alpha$, the probability of a Type I error is $\alpha$.

Note! It is very tempting to think that $P(H_0|Data)$, but this is not the case. We have calculated the p-value *assuming* $H_0$ *is true.* Moreover, this kind of quantity doesn't make sense in frequentist statistics, in which we think of the parameters (determining $H_0$) as being fixed. However, we will see soon that this quantity does make sense (and can be calculated) in a Bayesian framework.