We'll next discuss some tests when the data are multinomial.

Aside: The Multinomial Distribution

Suppose $Z \in \{1, ..., k\}$ and let $p_j = P(Z = j)$. The parameter $p = (p_1, \ldots, p_k)$ is really only $k - 1$ dimensional, since $\sum_{j=1}^{k} p_j = 1$. Suppose we observe an iid sample $Z_1, \ldots, Z_n$. Let $X_j = \#\{Z_i : Z_i = j\}$. Then we say $X = (X_1, \ldots, X_k)$ has $Multinomial(n, p)$ distribution.

The PDF is

$$f(x_1, \ldots, x_k; p) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

Note that the labels $1, \ldots, k$ for the Z's are arbitrary, and that $Binomial(n, p)$ distribution is just a special case.

The MLE is $(\hat{p}_1, \ldots, \hat{p}_k) = (X_1/n, \ldots, X_k/n)$.

Consider testing $H_0 : (p_1, \ldots, p_k) = (p_{01}, \ldots, p_{ok})$ versus the alternative that they are not equal. The LRT rejects when

$$T(X) = \frac{\mathcal{L}_n(\hat{p})}{\mathcal{L}(p_0)} = \prod_{j=1}^{k} \left(\frac{\hat{p}_j}{p_{0j}} \right)^{X_j}$$

is large. Since I don't know how to calculate the exact probability of this, I'll use the limiting χ^2. That is,

$$\lambda(X) = 2 \log T(X) = 2 \sum_{j=1}^{k} X_j \log \left(\frac{\hat{p}_j}{p_{0j}} \right) \xrightarrow{D} \chi^2_{k-1}$$

The degrees of freedom is $k-1$ because the dimension of Θ is $k-1$ and the dimension of Θ_0 is zero (a point). The approximate size α LRT rejects H_0 when $\lambda(X) \geq \chi^2_{k-1,\alpha}$.

Example: Consider the following data on 2009 freshman admissions at Berkeley, taken from `http://students.berkeley.edu/admissions/freshmen.asp`.

	California Residents	Non-Residents	International Students
Applicants	38,082	6,309	4,259
Admitted	11,252	1,110	666
(% Admitted)	(29.5%)	(17.6%)	(15.6%)
Enrolled	4,262	216	301

Treat the enrolled students as a sample from the Multinomial distribution, and test the hypothesis that the proportion of the three groups among the enrolled students is the same as it was for admitted students, i.e., that

$$p = \left(\frac{11252}{13028}, \frac{1110}{13028}, \frac{666}{13028}\right)$$

Another popular test for this situation is called Pearson's χ^2 test. The statistic is defined as

$$T = \sum_{j=1}^{k} \frac{(X_j - np_{0j})^2}{np_{0j}} = \sum_{j=1}^{k} \frac{(X_j - E_j)^2}{E_j}$$

Here $E_j = E[X_j] = np_{0j}$ is the expected value of X_j under H_0.

This statistic also has a limiting χ^2_{k-1} distribution under H_0.

Example: Test the Berkeley data again using Pearson's χ^2 test.

The LRT and Pearson's χ^2 are asymptotically equivalent, so they give similar answers for large n. However, Pearson's χ^2 statistic tends to converge to χ^2_{k-1} in distribution faster, so it is preferable for small n.

The LRT and Pearson's χ^2 also arise in tests of independence. Consider a simple case first, that two r.v.'s Y and Z are binary. The data are shown in the left table and the corresponding probabilities in the right table.

	$Y=0$	$Y=1$	
$Z=0$	X_{00}	X_{01}	$X_{0.}$
$Z=1$	X_{10}	X_{11}	$X_{1.}$
	$X_{.0}$	$X_{.1}$	n

	$Y=0$	$Y=1$	
$Z=0$	p_{00}	p_{01}	$p_{0.}$
$Z=1$	p_{10}	p_{11}	$p_{1.}$
	$p_{.0}$	$p_{.1}$	1

We treat $X = (X_{00}, X_{01}, X_{1,0}, X_{11})$ as a sample from a multinomial distribution. Under the null hypothesis that Y and Z are independent, the cell probabilities are the product of the row and column probabilities:

$$p_{ij} = p_{i.}p_{.j}$$

We can use this to construct either a LRT (doing a constrained maximization) or Pearson's χ^2.

Consider now a table with I rows and J columns. The unconstrained MLEs are $\hat{p}_{ij} = X_{ij}/n$, and under H_0, the constrained MLEs are

$$\hat{p}_{0ij} = \hat{p}_{0i.}\hat{p}_{0.j} = \frac{X_{i.}}{n}\frac{X_{.j}}{n}$$

Therefore for the LRT we have $\lambda = 2\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}\log\left(\frac{nX_{ij}}{X_{i.}X_{j.}}\right)$ and for Pearson's χ^2 we have

$$T = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(X_{ij}-E_{ij})^2}{E_{ij}} = \sum_{i-1}^{I}\sum_{j=1}^{J}\frac{(X_{ij}-n\hat{p}_{0ij})^2}{n\hat{p}_{0ij}}$$

Both test statistics have a limiting χ^2_ν distribution, where $\nu = (I-1)(J-1)$.

Finally, we can adapt these ideas to form a test of goodness of fit. Here the null hypothesis is that the data come from an assumed parametric model. The idea is to "discretize" both the data and the model.

First, define k disjoint intervals $I_1, \ldots, I_k$. Define

$$p_j(\theta) = P_\theta(X \in I_j) = \int_{I_j} f(x;\theta)dx$$

Let $N_j = \#\{X_i \in I_j\}$, the number of observations that fall into I_j. Treat $N = (N_1, \ldots, N_k)$ as a sample from a multinomial distribution with $p(\theta) = (p_1(\theta), \ldots, p_k(\theta))$, and maximize the likelihood to get $\tilde{\theta}$.

Then under H_0 that the data are iid draws from $\mathcal{F} = \{f(x;\theta) : \theta \in \Theta\}$, the test statistic $Q = \sum_{j=1}^{k} \frac{(N_j - np_j(\tilde{\theta}))^2}{np_j(\tilde{\theta})} \xrightarrow{D} \chi^2_{k-1-s}$, where s is the dimension of θ.

"Multiple testing" refers to the problem of testing $m > 1$ hypotheses, and wanting to control something more than the error rate for each test.

The Bonferroni Method controls the probability of having at least one false rejection. If α is the upper bound placed on this probability, the method achieves this by using level α/m for each of the tests.

$$\begin{aligned} P(\text{at least one Type I error}) &= P(\bigcup_{i=1}^{m} \text{Type I error in the } i^{th} \text{ test}) \\ &\leq \sum_{i=1}^{m} P(\text{Type I error in the } i^{th} \text{ test}) \\ &= \sum_{i=1}^{m} \alpha/m = \alpha \end{aligned}$$

In many cases, Bonferroni is too conservative. Another option is to control the False Discovery Rate (FDR), which is

$$FDR = E\left(\frac{\text{Number of false rejections}}{\text{Total number of rejections}}\right)$$

Benjamini and Hochberg suggested the following procedure, which guarantees $FDR \leq \alpha$:

1. For each test, compute the $p - value$. Let $P_{(1)} < \cdots < P_{(m)}$ denote the ordered p-values.

2. Select $R = \max\{i : P_{(i)} < \frac{i\alpha}{C_m m}\}$, where C_m is 1 if the p-values are independent and $C_m = \sum_{i=1}^{m}(1/i)$ otherwise.

3. Reject all null hypotheses for which the p-value $\leq P_{(R)}$.