Stat 710: Mathematical Statistics
Lecture 29

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Lecture 29: Chi-square tests and goodness of fit tests

Testing in multinomial distributions

Consider n independent trials with k possible outcomes for each trial.
Let $p_j > 0$ be the probability that the jth outcome occurs in a given trial and X_j be the number of occurrences of the jth outcome in n trials.
Then $X = (X_1, ..., X_k)$ has the multinomial distribution (Example 2.7) with the parameter $\mathbf{p} = (p_1, ..., p_k)$.
Let $\xi_i = (0, ..., 0, 1, 0, ..., 0)$, where the single nonzero component 1 is located in the jth position if the ith trial yields the jth outcome.
Then $\xi_1, ..., \xi_n$ are i.i.d. and $X/n = \bar{\xi} = \sum_{i=1}^n \xi_i / n$.
X/n is an unbiased estimator of $\mathbf{p}$ and, by the CLT,

$$Z_n(\mathbf{p}) = \sqrt{n}\left(\frac{X}{n} - \mathbf{p}\right) = \sqrt{n}(\bar{\xi} - \mathbf{p}) \to_d N_k(0, \Sigma),$$

where $\Sigma = \mathrm{Var}(X/\sqrt{n})$ is a symmetric $k \times k$ matrix whose ith diagonal element is $p_i(1 - p_i)$ and (i,j)th off-diagonal element is $-p_i p_j$.
We first consider the problem of testing

$$H_0 : \mathbf{p} = \mathbf{p}_0 \qquad \text{versus} \qquad H_1 : \mathbf{p} \neq \mathbf{p}_0,$$

where $\mathbf{p}_0 = (p_{01}, ..., p_{0k})$ is a known vector of cell probabilities.

χ^2 tests

For testing $H: \mathbf{p} = \mathbf{p}_0$ vs $H_1: \mathbf{p} \neq \mathbf{p}_0$, a class of tests related to the asymptotic tests described in §6.4.2 is the class of χ^2-tests.

A popular test is based on the following χ^2-statistic:

$$\chi^2 = \sum_{j=1}^{k} \frac{(X_j - np_{0j})^2}{np_{0j}} = \|D(\mathbf{p}_0)Z_n(\mathbf{p}_0)\|^2,$$

where $D(c)$ with $c = (c_1, ..., c_k)$ is the $k \times k$ diagonal matrix whose jth diagonal element is $c_j^{-1/2}$.

Another popular test is based on the following modified χ^2-statistic:

$$\tilde{\chi}^2 = \sum_{j=1}^{k} \frac{(X_j - np_{0j})^2}{X_j} = \|D(X/n)Z_n(\mathbf{p}_0)\|^2.$$

The next result shows that a test of asymptotic significance level α rejects $H_0: \mathbf{p} = \mathbf{p}_0$ when $\chi^2 > \chi^2_{k-1,\alpha}$ (or $\tilde{\chi}^2 > \chi^2_{k-1,\alpha}$), where $\chi^2_{k-1,\alpha}$ is the $(1-\alpha)$th quantile of χ^2_{k-1}.

Thus, these tests are called (asymptotic) χ^2-tests.

Theorem 6.8

Let $\phi = (\sqrt{p_1}, ..., \sqrt{p_k})$ and Λ be a $k \times k$ projection matrix.
(i) If $\Lambda\phi = a\phi$, then

$$[Z_n(\mathbf{p})]^\tau D(\mathbf{p}) \Lambda D(\mathbf{p}) Z_n(\mathbf{p}) \to_d \chi_r^2,$$

where χ_r^2 has the chi-square distribution χ_r^2 with $r = \mathrm{tr}(\Lambda) - a$.
(ii) The same result holds if $D(\mathbf{p})$ in (i) is replaced by $D(X/n)$.

Remark

The χ^2-statistic and the modified χ^2-statistic are special cases of the statistics in Theorem 6.8(i) and (ii), respectively, with $\Lambda = I_k$ satisfying $\Lambda\phi = \phi$.

Proof

The result in (ii) follows from the result in (i) and $X/n \to_p \mathbf{p}$.
To prove (i), let $D = D(\mathbf{p})$, $Z_n = Z_n(\mathbf{p})$, and $Z = N_k(0, I_k)$.
From the asymptotic normality of Z_n and Theorem 1.10,

$$Z_n^\tau D \Lambda D Z_n \to_d Z^\tau A Z \qquad \text{with} \quad A = \Sigma^{1/2} D \Lambda D \Sigma^{1/2}.$$

Theorem 6.8

Let $\phi = (\sqrt{p_1}, ..., \sqrt{p_k})$ and Λ be a $k \times k$ projection matrix.
(i) If $\Lambda\phi = a\phi$, then

$$[Z_n(\mathbf{p})]^\tau D(\mathbf{p})\Lambda D(\mathbf{p}) Z_n(\mathbf{p}) \to_d \chi_r^2,$$

where χ_r^2 has the chi-square distribution χ_r^2 with $r = \text{tr}(\Lambda) - a$.
(ii) The same result holds if $D(\mathbf{p})$ in (i) is replaced by $D(X/n)$.

Remark

The χ^2-statistic and the modified χ^2-statistic are special cases of the statistics in Theorem 6.8(i) and (ii), respectively, with $\Lambda = I_k$ satisfying $\Lambda\phi = \phi$.

Proof

The result in (ii) follows from the result in (i) and $X/n \to_p \mathbf{p}$.
To prove (i), let $D = D(\mathbf{p})$, $Z_n = Z_n(\mathbf{p})$, and $Z = N_k(0, I_k)$.
From the asymptotic normality of Z_n and Theorem 1.10,

$$Z_n^\tau D\Lambda D Z_n \to_d Z^\tau A Z \qquad \text{with} \quad A = \Sigma^{1/2} D\Lambda D \Sigma^{1/2}.$$

Theorem 6.8

Let $\phi = (\sqrt{p_1}, ..., \sqrt{p_k})$ and Λ be a $k \times k$ projection matrix.
(i) If $\Lambda\phi = a\phi$, then

$$[Z_n(\mathbf{p})]^\tau D(\mathbf{p})\Lambda D(\mathbf{p}) Z_n(\mathbf{p}) \to_d \chi_r^2,$$

where χ_r^2 has the chi-square distribution χ_r^2 with $r = \mathrm{tr}(\Lambda) - a$.
(ii) The same result holds if $D(\mathbf{p})$ in (i) is replaced by $D(X/n)$.

Remark

The χ^2-statistic and the modified χ^2-statistic are special cases of the statistics in Theorem 6.8(i) and (ii), respectively, with $\Lambda = I_k$ satisfying $\Lambda\phi = \phi$.

Proof

The result in (ii) follows from the result in (i) and $X/n \to_p \mathbf{p}$.
To prove (i), let $D = D(\mathbf{p})$, $Z_n = Z_n(\mathbf{p})$, and $Z = N_k(0, I_k)$.
From the asymptotic normality of Z_n and Theorem 1.10,

$$Z_n^\tau D\Lambda D Z_n \to_d Z^\tau A Z \qquad \text{with} \quad A = \Sigma^{1/2} D\Lambda D \Sigma^{1/2}.$$

Proof (continued)

From Exercise 51 in §1.6, the result in (i) follows if we can show that $A^2 = A$ (i.e., A is a projection matrix) and $\mathrm{tr}(A) = \mathrm{tr}(\Lambda) - a$.
Since Λ is a projection matrix and $\Lambda\phi = a\phi$, a must be either 0 or 1.
Note that $D\Sigma D = I_k - \phi\phi^\tau$.
Then

$$\begin{aligned}
A^3 &= \Sigma^{1/2} D\Lambda D\Sigma D\Lambda D\Sigma D\Lambda D\Sigma^{1/2} \\
&= \Sigma^{1/2} D(\Lambda - a\phi\phi^\tau)(\Lambda - a\phi\phi^\tau)\Lambda D\Sigma^{1/2} \\
&= \Sigma^{1/2} D(\Lambda - 2a\phi\phi^\tau + a^2\phi\phi^\tau)\Lambda D\Sigma^{1/2} \\
&= \Sigma^{1/2} D(\Lambda - a\phi\phi^\tau)\Lambda D\Sigma^{1/2} \\
&= \Sigma^{1/2} D\Lambda D\Sigma D\Lambda D\Sigma^{1/2} \\
&= A^2,
\end{aligned}$$

which implies that the eigenvalues of A must be 0 or 1.
Therefore, $A^2 = A$.
Also,

$$\mathrm{tr}(A) = \mathrm{tr}[\Lambda(D\Sigma D)] = \mathrm{tr}(\Lambda - a\phi\phi^\tau) = \mathrm{tr}(\Lambda) - a.$$

Example 6.23 (Goodness of fit tests)

Let $Y_1, ..., Y_n$ be i.i.d. from F. Consider the problem of testing

$$H_0 : F = F_0 \qquad \text{versus} \qquad H_1 : F \neq F_0,$$

where F_0 is a known c.d.f. (For instance, $F_0 = N(0,1)$.)
One way to test $H_0 : F = F_0$ is to partition the range of Y_1 into k disjoint events $A_1, ..., A_k$ and test $H_0 : \mathbf{p} = \mathbf{p}_0$ with $p_j = P_F(A_j)$ and $p_{0j} = P_{F_0}(A_j)$, $j = 1, ..., k$.
Let X_j be the number of Y_i's in A_j, $j = 1, ..., k$.
Based on X_j's, the χ^2-tests discussed previously can be applied.
They are called *goodness of fit* tests.

In the goodness of fit tests discussed in Example 6.23, F_0 in H_0 is known so that p_{0j}'s can be computed.
In some cases, we need to test the following hypotheses:

$$H_0 : F = F_\theta \qquad \text{versus} \qquad H_1 : F \neq F_\theta,$$

where θ is an unknown parameter in $\Theta \subset \mathscr{R}^s$.
For example, $F_\theta = N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$.

Example 6.23 (Goodness of fit tests)

Let $Y_1,...,Y_n$ be i.i.d. from F. Consider the problem of testing

$$H_0: F = F_0 \qquad \text{versus} \qquad H_1: F \neq F_0,$$

where F_0 is a known c.d.f. (For instance, $F_0 = N(0,1)$.)
One way to test $H_0: F = F_0$ is to partition the range of Y_1 into k disjoint events $A_1,...,A_k$ and test $H_0: \mathbf{p} = \mathbf{p}_0$ with $p_j = P_F(A_j)$ and $p_{0j} = P_{F_0}(A_j)$, $j = 1,...,k$.
Let X_j be the number of Y_i's in A_j, $j = 1,...,k$.
Based on X_j's, the χ^2-tests discussed previously can be applied.
They are called *goodness of fit* tests.

In the goodness of fit tests discussed in Example 6.23, F_0 in H_0 is known so that p_{0j}'s can be computed.
In some cases, we need to test the following hypotheses:

$$H_0: F = F_\theta \qquad \text{versus} \qquad H_1: F \neq F_\theta,$$

where θ is an unknown parameter in $\Theta \subset \mathscr{R}^s$.
For example, $F_\theta = N(\mu,\sigma^2)$, $\theta = (\mu,\sigma^2)$.

If we still try to test $H_0 : \mathbf{p} = \mathbf{p}_0$ with $p_j = P_{F_\theta}(A_j)$, $j = 1, ..., k$, the result in Example 6.23 is not applicable since $\mathbf{p}$ is unknown under H_0.
A generalized χ^2-test can be obtained using the following result.
Let $\mathbf{p}(\theta) = (p_1(\theta), ..., p_k(\theta))$ be a k-vector of known functions of $\theta \in \Theta \subset \mathscr{R}^s$, where $s < k$.
Consider the testing problem

$$H_0 : \mathbf{p} = \mathbf{p}(\theta) \qquad \text{versus} \qquad H_1 : \mathbf{p} \neq \mathbf{p}(\theta).$$

Note that $H_0 : \mathbf{p} = \mathbf{p}_0$ is the special case of $H_0 : \mathbf{p} = \mathbf{p}(\theta)$ with $s = 0$.
Let $\widehat{\theta}$ be an MLE of θ under H_0.
By Theorem 6.5, the LR test that rejects H_0 when $-2\log\lambda_n > \chi^2_{k-s-1,\alpha}$ has asymptotic significance level α, where $\chi^2_{k-s-1,\alpha}$ is the $(1-\alpha)$th quantile of χ^2_{k-s-1} and

$$\lambda_n = \prod_{j=1}^{k} [p_j(\widehat{\theta})]^{X_j} \Big/ (X_j/n)^{X_j}.$$

Using the fact that $p_j(\widehat{\theta})/(X_j/n) \to_p 1$ under H_0 and

$$\log(1+x) = x - x^2/2 + o(|x|^2) \qquad \text{as } |x| \to 0,$$

we obtain that

$$
\begin{aligned}
-2\log\lambda_n &= -2\sum_{j=1}^{k} X_j \log\left(1+\frac{p_j(\widehat{\theta})}{X_j/n}-1\right)\\
&= -2\sum_{j=1}^{k} X_j\left(\frac{p_j(\widehat{\theta})}{X_j/n}-1\right)+\sum_{j=1}^{k} X_j\left(\frac{p_j(\widehat{\theta})}{X_j/n}-1\right)^2+o_p(1)\\
&= \sum_{j=1}^{k}\frac{[X_j-np_j(\widehat{\theta})]^2}{X_j}+o_p(1)\\
&= \sum_{j=1}^{k}\frac{[X_j-np_j(\widehat{\theta})]^2}{np_j(\widehat{\theta})}+o_p(1),
\end{aligned}
$$

where the third equality follows from $\sum_{j=1}^{k} p_j(\widehat{\theta}) = \sum_{j=1}^{k} X_j/n = 1$.

Generalized χ^2-statistics

The generalized χ^2-statistics χ^2 and $\tilde{\chi}^2$ are defined to be the previously defined χ^2-statistics with p_{0j}'s replaced by $p_j(\widehat{\theta})$'s.

we obtain that

$$
\begin{aligned}
-2\log\lambda_n &= -2\sum_{j=1}^{k} X_j \log\left(1+\frac{p_j(\widehat{\theta})}{X_j/n}-1\right) \\
&= -2\sum_{j=1}^{k} X_j \left(\frac{p_j(\widehat{\theta})}{X_j/n}-1\right) + \sum_{j=1}^{k} X_j \left(\frac{p_j(\widehat{\theta})}{X_j/n}-1\right)^2 + o_p(1) \\
&= \sum_{j=1}^{k} \frac{[X_j - np_j(\widehat{\theta})]^2}{X_j} + o_p(1) \\
&= \sum_{j=1}^{k} \frac{[X_j - np_j(\widehat{\theta})]^2}{np_j(\widehat{\theta})} + o_p(1),
\end{aligned}
$$

where the third equality follows from $\sum_{j=1}^{k} p_j(\widehat{\theta}) = \sum_{j=1}^{k} X_j/n = 1$.

Generalized χ^2-statistics

The generalized χ^2-statistics χ^2 and $\tilde{\chi}^2$ are defined to be the previously defined χ^2-statistics with p_{0j}'s replaced by $p_j(\widehat{\theta})$'s.

Theorem 6.9

Under $H_0 : \mathbf{p} = \mathbf{p}(\theta)$, the generalized χ^2-statistics converge in distribution to χ^2_{k-s-1}.
The χ^2-test with rejection region $\chi^2 > \chi^2_{k-s-1,\alpha}$ (or $\tilde{\chi}^2 > \chi^2_{k-s-1,\alpha}$) has asymptotic significance level α, where $\chi^2_{k-s-1,\alpha}$ is the $(1-\alpha)$th quantile of χ^2_{k-s-1}.

Discussion

Theorem 6.9 can be applied to derive a goodness of fit test for $H_0 : \mathbf{p} = \mathbf{p}(\theta)$ vs $H_1 : \mathbf{p} = \mathbf{p}(\theta)$.
However, one has to compute an MLE of θ under $H_0 : \mathbf{p} = \mathbf{p}(\theta)$, which is different from an MLE under $H_0 : F = F_\theta$ unless $F = F_\theta$ and $\mathbf{p} = \mathbf{p}(\theta)$ are the same; see Moore and Spruill (1975).
Many elementary textbooks, however, use an MLE under $H_0 : F = F_\theta$, which is wrong.

Theorem 6.9

Under $H_0 : \mathbf{p} = \mathbf{p}(\theta)$, the generalized χ^2-statistics converge in distribution to χ^2_{k-s-1}.
The χ^2-test with rejection region $\chi^2 > \chi^2_{k-s-1,\alpha}$ (or $\tilde{\chi}^2 > \chi^2_{k-s-1,\alpha}$) has asymptotic significance level α, where $\chi^2_{k-s-1,\alpha}$ is the $(1-\alpha)$th quantile of χ^2_{k-s-1}.

Discussion

Theorem 6.9 can be applied to derive a goodness of fit test for $H_0 : \mathbf{p} = \mathbf{p}(\theta)$ vs $H_1 : \mathbf{p} = \mathbf{p}(\theta)$.
However, one has to compute an MLE of θ under $H_0 : \mathbf{p} = \mathbf{p}(\theta)$, which is different from an MLE under $H_0 : F = F_\theta$ unless $F = F_\theta$ and $\mathbf{p} = \mathbf{p}(\theta)$ are the same; see Moore and Spruill (1975).
Many elementary textbooks, however, use an MLE under $H_0 : F = F_\theta$, which is wrong.

MLE under $\mathbf{p} = \mathbf{p}(\theta)$

From the multinomial distribution, the MLE $\widehat{\theta}$ in the generalized χ^2 test should maximize the likelihood

$$\ell(\theta) = \frac{n!}{x_1! \cdots x_k!}[p_1(\theta)]^{x_1} \cdots [p_k(\theta)]^{x_k} I_{x_1+\cdots+x_k=1}$$

This MLE $\widehat{\theta}$ is different from the MLE maximizing the likelihood based on the family $\{F_\theta\}$

For testing $H_0 : F = N(\mu, \sigma^2)$, for example,

$$p_j(\theta) = \Phi\left(\frac{a_{j+1} - \mu}{\sigma}\right) - \Phi\left(\frac{a_j - \mu}{\sigma}\right), \quad j = 1, ..., k$$

where $-\infty = a_1 < a_2 < \cdots < a_k < a_{k+1} = \infty$ and a_j's are fixed constants
This MLE $\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma}^2)$ is certainly different from $\widehat{\mu}$ = the sample mean and $\widehat{\sigma}^2 = (n-1)/n$ times the sample variance, which is the MLE under the normal model $N(\mu, \sigma^2)$.