Slides 6: Likelihood ratio tests

- Likelihood ratio
- Nested hypotheses
- Chi-square null distribution
- Chi-squared test of goodness of fit
- A case study

$$\text{Likelihood ratio test statistic} = -2\log\frac{\max_{\theta\in\Omega_0} L(\theta)}{\max_{\theta\in\Omega} L(\theta)}$$

has χ^2_{df} as an approximate null distribution, with

$$\mathrm{df} = \dim(\Omega) - \dim(\Omega_0)$$

A general method of finding asymptotically optimal tests (having the largest power for a given α) takes likelihood ratio as the test statistic.

Consider a parametric population distribution with a single parameter θ and a likelihood function $L(\theta) = L(\theta; x_1, \ldots, x_n)$. For testing

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta = \theta_1,$$

use the likelihood ratio

$$\lambda = \frac{L(\theta_0)}{L(\theta_1)}$$

as a test statistic. Large values of λ suggest that H_0 explains the data set better than H_1. Therefore, the likelihood ratio test rejects H_0 for small values of the likelihood ratio.

$$\boxed{\text{Likelihood ratio rejection rule is } \{\lambda \le \lambda_\alpha\}.}$$

Neyman-Pearson lemma: the likelihood ratio test is optimal in the case of two simple hypothesis.

Question. How do we find the critical value λ_α?

For example, consider $\mathrm{N}(\mu, \sigma)$ model with $\theta = (\mu, \sigma)$. Instead of a pair of two alternative hypotheses $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, one can think in terms of a pair of nested hypothesis

$$H_0 : \mu = \mu_0, \quad H : \mu \in (-\infty, \infty).$$

More generally, consider

$$H_0 : \theta \in \Omega_0, \quad H : \theta \in \Omega,$$

where parameter sets $\Omega_0 \subset \Omega$ are such that $\dim(\Omega) > \dim(\Omega_0)$.
Generalised likelihood ratio

$$\tilde{\lambda} = \tfrac{L(\hat{\theta}_0)}{L(\hat{\theta})},$$

is defined in terms of two maximum likelihood estimates

$$\begin{aligned} \hat{\theta}_0 &= \text{maximises the likelihood function } L(\theta) \text{ over } \theta \in \Omega_0, \\ \hat{\theta} &= \text{maximises the likelihood function } L(\theta) \text{ over } \theta \in \Omega. \end{aligned}$$

Question. What is $\mathrm{df} = \dim(\Omega) - \dim(\Omega_0)$ in the example above?

Generalised likelihood ratio test rejects H_0 for small values of $\tilde{\lambda}$ or equivalently for large values of

$$-\ln\tilde{\lambda} = \ln L(\hat{\theta}) - \ln L(\hat{\theta}_0).$$

It turns out that the test statistic $-2\ln\tilde{\Lambda}$ has a nice approximate null distribution

$$-2\ln\tilde{\Lambda} \overset{H_0}{\approx} \chi^2_{\mathrm{df}}, \quad \text{where df} = \dim(\Omega) - \dim(\Omega_0).$$

χ^2_k-distribution is the gamma distribution with $\alpha = \frac{k}{2}, \lambda = \frac{1}{2}$. If independent $Z_1, \ldots, Z_k$ have the same N(0,1) distribution, then

$$Z_1^2 + \ldots + Z_k^2 \sim \chi^2_k.$$

Question. Consider $\mathrm{N}(\mu, \sigma)$ model with $\theta = (\mu, \sigma)$. With $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, how would you connect the corresponding likelihood ratio test to the large sample test for the mean?

Suppose that the population distribution is discreet with probabilities $(p_1, \dots, p_J)$. A sample of size n is summarised by the vector of observed counts whose joint distribution is multinomial

$$(O_1, \dots, O_J) \sim \text{Mn}(n; p_1, \dots, p_J),$$

$$\mathrm{P}(O_1 = k_1, \dots, O_J = k_J) = \frac{n!}{k_1! \cdots k_J!} p_1^{k_1} \cdots p_J^{k_J}.$$

Consider a parametric model for the data

$$H_0 : (p_1, \dots, p_J) = (v_1(\delta), \dots, v_J(\delta))$$

with unknown r-dimensional parameter $\delta = (\delta_1, \dots, \delta_r)$.

To see if the proposed model fits the data, compute $\hat{\delta}$, the maximum likelihood estimate of δ, and then the expected cell counts

$$E_j = n \cdot v_j(\hat{\delta}),$$

where "expected" means expected under the null hypothesis model.

Question. What is Ω_0 and Ω in this setting?

In the current setting, the likelihood ratio test statistic $-2\log\tilde{\lambda}$ is approximated by the so-called chi-squared test statistic

$$\mathrm{X}^2 = \sum_{j=1}^{J} \frac{(O_j - E_j)^2}{E_j}.$$

The approximate null distribution of X^2 is χ^2_{df} with df$= J - 1 - r$, since

$$\dim(\Omega_0) = r \quad \text{and} \quad \dim(\Omega) = J - 1,$$

where dim stands for dimension or the number of independent parameters. A mnemonic rule for the number of degrees of freedom:

$$\begin{aligned}\mathrm{df} &= (\text{number of cells}) - 1 \\ &\quad - (\text{number of independent parameters estimated from the data}).\end{aligned}$$

Since the chi-squared test is approximate, all *expected* counts are recommended to be at least 5. If not, then you should combine small cells in larger cells and recalculate the number of degrees of freedom df.

A 1889 study made in Germany recorded the numbers of boys $(x_1, \ldots, x_n)$ for $n = 6115$ families with 12 children each. The general model is described by a vector $\theta = (p_0, p_1, \ldots, p_{12})$ such that

$$p_j = \mathrm{P}(X = j), \quad j = 0, 1, \ldots, 12.$$

We first test a simple null hypothesis claiming that $X \sim$ Bin(12, 0.5), or

$$H_0 : p_j = \binom{12}{j} \cdot 2^{-12}, \quad j = 0, 1, \ldots, 12.$$

The expected cell counts

$$E_j = 6115 \cdot \binom{12}{j} \cdot 2^{-12}, \quad j = 0, 1, \ldots, 12,$$

are summarised in the table below. The chi-squared test statistic

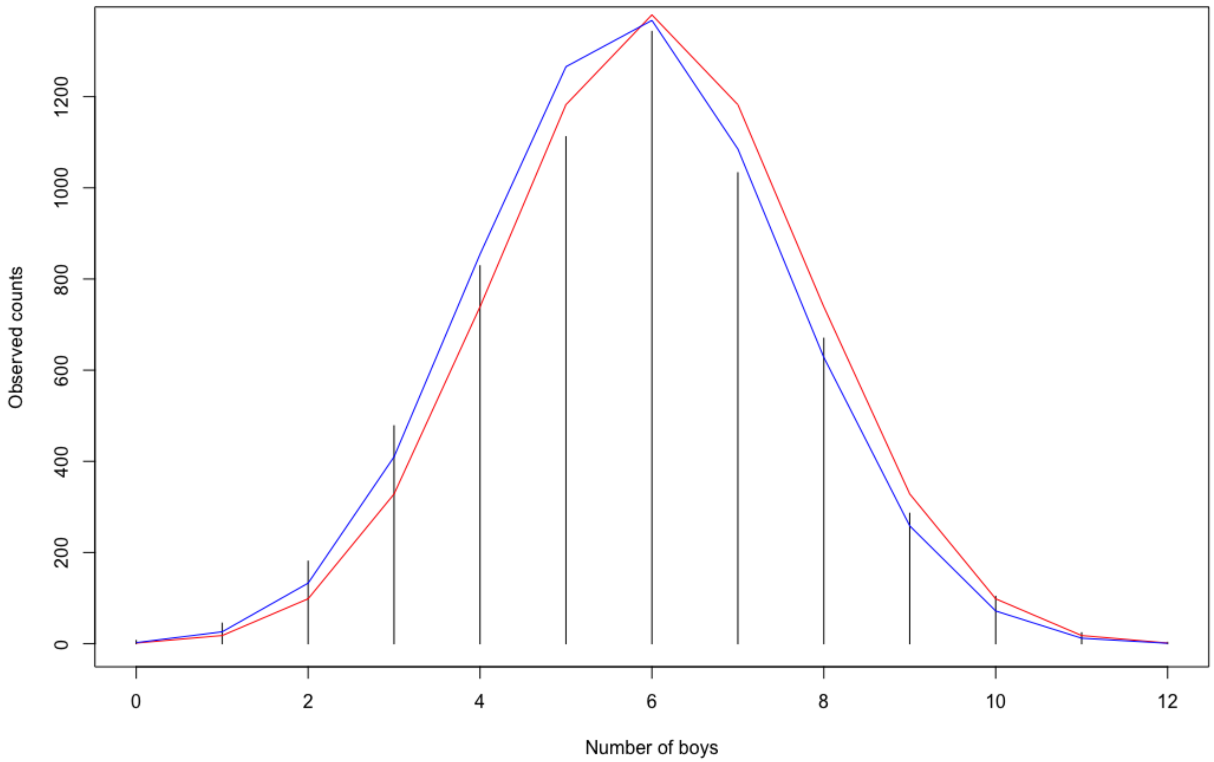
$$\mathrm{X}^2 = \sum_{j=0}^{12} \frac{(O_j - E_j)^2}{E_j}.$$

has the observed value $\mathrm{X}^2 = 249.2$. We have $\mathrm{df} = 13 - 1 - 0 = 12$. Since $\chi^2_{12}(0.005) = 28.3$, we reject H_0 at 0.5% level.

cell j	O_j	Model 1: E_j	and $\frac{(O_j-E_j)^2}{E_j}$	Model 2: E_j	and $\frac{(O_j-E_j)^2}{E_j}$
0	7	1.5	20.2	2.3	9.6
1	45	17.9	41.0	26.1	13.7
2	181	98.5	69.1	132.8	17.5
3	478	328.4	68.1	410.0	11.3
4	829	739.0	11.0	854.2	0.7
5	1112	1182.4	4.2	1265.6	18.6
6	1343	1379.5	1.0	1367.3	0.4
7	1033	1182.4	18.9	1085.2	2.5
8	670	739.0	6.4	628.1	2.8
9	286	328.4	5.5	258.5	2.9
10	104	98.5	0.3	71.8	14.4
11	24	17.9	2.1	12.1	11.7
12	3	1.5	1.5	0.9	4.9
Total	6115	6115	$\mathrm{X}^2 = 249.2$	6115	$\mathrm{X}^2 = 110.5$

Consider next a more flexible model $X \sim \mathrm{Bin}(12, \delta)$. Model 2 leads to a composite null hypothesis

$$H_0 : p_j = \binom{12}{j} \cdot \delta^j (1-\delta)^{12-j}, \quad j = 0, \ldots, 12, \quad 0 \le \delta \le 1.$$



Estimate δ using the maximum likelihood estimate of the proportion of boys in a family

$$\hat{\delta} = \frac{\text{number of boys}}{\text{number of children}} = \frac{1 \cdot 45 + 2 \cdot 181 + \ldots + 12 \cdot 3}{6115 \cdot 12} = 0.481$$

The expected cell counts

$$E_j = 6115 \cdot \binom{12}{j} \cdot \hat{\delta}^j \cdot (1 - \hat{\delta})^{12-j}$$

are given in the table and the graph above.

The observed chi-squared test statistic for Model 2

$$\mathrm{X}^2 = 110.5$$

is much smaller than that for Model 1. However, with $r = 1$, $\mathrm{df} = 11$, and the table value $\chi^2_{11}(0.005) = 26.76$, we reject even Model 2 at 0.5% level.

We see that what is needed is an even more flexible model addressing large variation in the observed cell counts.

Suggestion for Model 3: allow the probability of a male child δ to differ from family to family. Namely, assume that for each family the value δ is generated by a beta-distribution $\mathrm{Beta}(a, b)$.

Question. What is dimension r for the suggested Model 3?