

## Lecture 22 — The generalized likelihood ratio test

In the next two lectures, we revisit the problem of hypothesis testing in the context of parametric models. We'll introduce the generalized likelihood ratio test and explore applications to the analysis of categorical data.

## 22.1 GLRT for a simple null hypothesis

Let  $\{f(x|\theta) : \theta \in \Omega\}$  be a parametric model, and let  $\theta_0 \in \Omega$  be a particular parameter value. For testing

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &\neq \theta_0 \end{aligned}$$

the **generalized likelihood ratio test (GLRT)** rejects for small values of the test statistic

$$\Lambda = \frac{\text{lik}(\theta_0)}{\max_{\theta \in \Omega} \text{lik}(\theta)},$$

where  $\text{lik}(\theta)$  is the likelihood function. (In the case of IID samples  $X_1, \dots, X_n \stackrel{IID}{\sim} f(x|\theta)$ ,  $\text{lik}(\theta) = \prod_{i=1}^n f(X_i|\theta)$ .) The numerator is the value of the likelihood at  $\theta_0$ , and the denominator is the value of the likelihood at the MLE  $\hat{\theta}$ . The level- $\alpha$  GLRT rejects  $H_0$  when  $\Lambda \leq c$ , where (as usual)  $c$  is chosen so that  $\mathbb{P}_{H_0}[\Lambda \leq c]$  equals (or approximately equals)  $\alpha$ .

Note that the GLRT differs from the likelihood ratio test discussed previously in the context of the Neyman-Pearson lemma, where the denominator was instead given by  $\text{lik}(\theta_1)$  for a simple alternative  $\theta = \theta_1$ . The alternative  $H_1$  above is not simple, and the GLRT replaces the denominator with the maximum value of the likelihood over all values of  $\theta$ .

**Example 22.1.** Let  $X_1, \dots, X_n \stackrel{IID}{\sim} \mathcal{N}(\theta, 1)$  and consider the problem of testing

$$\begin{aligned} H_0 : \theta &= 0 \\ H_1 : \theta &\neq 0 \end{aligned}$$

The MLE for  $\theta$  is  $\hat{\theta} = \bar{X}$ . We compute

$$\begin{aligned} \text{lik}(0) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \\ \max_{\theta \in \mathbb{R}} \text{lik}(\theta) &= \text{lik}(\hat{\theta}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \bar{x})^2}{2}}. \end{aligned}$$

Then

$$\begin{aligned}\Lambda &= \frac{\text{lik}(0)}{\max_{\theta \in \mathbb{R}} \text{lik}(\theta)} = \exp \left( -\sum_{i=1}^n \frac{X_i^2}{2} + \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{2} \right) \\ &= \exp \left( -\sum_{i=1}^n \frac{X_i^2}{2} + \sum_{i=1}^n \frac{X_i^2 - 2X_i\bar{X} + \bar{X}^2}{2} \right) = \exp \left( -\frac{n}{2} \bar{X}^2 \right).\end{aligned}$$

Rejecting for small values of  $\Lambda$  is the same as rejecting for large values of  $-2 \log \Lambda = n\bar{X}^2$ . Under  $H_0$ ,  $\sqrt{n}\bar{X} \sim \mathcal{N}(0, 1)$ , so  $n\bar{X}^2 \sim \chi_1^2$ . Then the GLRT rejects  $H_0$  when  $n\bar{X}^2 > \chi_1^2(\alpha)$ , the upper- $\alpha$  point of the  $\chi_1^2$  distribution. (This is the same as rejecting when  $|\bar{X}| > z(\alpha/2)/\sqrt{n}$ , so the GLRT is equivalent to usual two-sided  $z$ -test based on  $\bar{X}$ .)

In general, the exact sampling distribution of  $-2 \log \Lambda$  under  $H_0$  may not have a simple form as in the above example, but it may be approximated by a chi-squared distribution for large  $n$ :

**Theorem 22.2.** *Let  $\{f(x|\theta) : \theta \in \Omega\}$  be a parametric model and let  $X_1, \dots, X_n \stackrel{IID}{\sim} f(x|\theta_0)$ . Suppose  $\theta_0$  is an interior point of  $\Omega$ , and the regularity conditions of Theorems 14.1 and 15.2 (for consistency and asymptotic normality of the MLE) hold. Then*

$$-2 \log \Lambda \rightarrow \chi_k^2$$

*in distribution as  $n \rightarrow \infty$ , where  $k = \dim \Omega$  is the dimension of  $\Omega$ .*

*Sketch of proof.* For simplicity, we consider only the case  $k = 1$ , so  $\theta$  is a single parameter. Letting  $l(\theta)$  denote the log-likelihood function and  $\hat{\theta}$  denote the MLE,

$$-2 \log \Lambda = -2l(\theta_0) + 2l(\hat{\theta}).$$

Applying a Taylor expansion of  $l(\theta_0)$  around  $\theta_0 = \hat{\theta}$ ,

$$l(\theta_0) \approx l(\hat{\theta}) + (\theta_0 - \hat{\theta})l'(\hat{\theta}) + \frac{1}{2}(\theta_0 - \hat{\theta})^2 l''(\hat{\theta}) \approx l(\hat{\theta}) - \frac{1}{2}nI(\theta_0)(\theta_0 - \hat{\theta})^2,$$

where the second approximation uses  $l'(\hat{\theta}) = 0$  and  $l''(\hat{\theta}) \approx -nI(\hat{\theta}) \approx -nI(\theta_0)$ . Then

$$-2 \log \Lambda \approx nI(\theta_0)(\theta_0 - \hat{\theta})^2.$$

$\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \rightarrow \mathcal{N}(0, 1)$  in distribution by asymptotic normality of the MLE, so the continuous mapping theorem implies  $-2 \log \Lambda \approx nI(\theta_0)(\theta_0 - \hat{\theta})^2 \rightarrow \chi_1^2$  as desired.  $\square$

This theorem implies that an approximate level- $\alpha$  test is given by rejecting  $H_0$  when  $-2 \log \Lambda > \chi_k^2(\alpha)$ , the upper- $\alpha$  point of the  $\chi_k^2$  distribution. The “dimension”  $k$  of  $\Omega$  is the number of free parameters in the model, or the number of parameters minus the number of independent constraints. For instance, in Example 22.1, there is a single parameter  $\theta$ , so the dimension is 1. For a multinomial model with parameters  $(p_1, \dots, p_k)$ , there are  $k$  parameters but they are constrained to sum to 1, so the dimension is  $k - 1$ .

## 22.2 GLRT for testing a sub-model

More generally, let  $\Omega_0 \subset \Omega$  be a subset of the parameter space  $\Omega$ , corresponding to a lower-dimensional sub-model. For testing

$$\begin{aligned} H_0 &: \theta \in \Omega_0 \\ H_1 &: \theta \notin \Omega_0 \end{aligned}$$

the generalized likelihood ratio statistic is defined as

$$\Lambda = \frac{\max_{\theta \in \Omega_0} \text{lik}(\theta)}{\max_{\theta \in \Omega} \text{lik}(\theta)}.$$

In other words,  $\Lambda$  is the ratio of the values of the likelihood function evaluated at the MLE in the sub-model and at the MLE in the full-model.

For large  $n$ , under any  $\theta_0 \in \Omega_0$ ,  $-2 \log \Lambda$  is approximately distributed as  $\chi_k^2$  where  $k$  is the difference in dimensionality between  $\Omega_0$  and  $\Omega$ , and an approximate level- $\alpha$  test rejects  $H_0$  when  $-2 \log \Lambda > \chi_k^2(\alpha)$ :

**Theorem 22.3.** *Let  $\{f(x|\theta) : \theta \in \Omega\}$  be a parametric model, and let  $X_1, \dots, X_n \stackrel{iid}{\sim} f(x|\theta_0)$  where  $\theta_0 \in \Omega_0$ . Suppose  $\theta_0$  is an interior point of both  $\Omega_0$  and  $\Omega$ , and the regularity conditions of Theorems 14.1 and 15.2 hold for both the full model  $\{f(x|\theta) : \theta \in \Omega\}$  and the sub-model  $\{f(x|\theta) : \theta \in \Omega_0\}$ . Then*

$$-2 \log \Lambda \rightarrow \chi_k^2$$

*in distribution as  $n \rightarrow \infty$ , where  $k = \dim \Omega - \dim \Omega_0$ .*

**Example 22.4** (Hardy-Weinberg equilibrium). At a single diallelic locus in the genome with two possible alleles A and a, any individual can have genotype AA, Aa, or aa. If we randomly select  $n$  individuals from a population, we may model the numbers of individuals with these genotypes as  $(N_{AA}, N_{Aa}, N_{aa}) \sim \text{Multinomial}(n, (p_{AA}, p_{Aa}, p_{aa}))$ .

When the alleles A and a are present in the population with proportions  $\theta$  and  $1 - \theta$ , then under an assumption of random mating, quantitative genetics theory predicts that  $p_{AA}$ ,  $p_{Aa}$ , and  $p_{aa}$  should be given by  $p_{AA} = \theta^2$ ,  $p_{Aa} = 2\theta(1 - \theta)$ , and  $p_{aa} = (1 - \theta)^2$ —this is called the Hardy-Weinberg equilibrium. In practice we do not know  $\theta$ , but we may still test the null hypothesis that Hardy-Weinberg equilibrium holds for some  $\theta$ :

$$H_0 : p_{AA} = \theta^2, p_{Aa} = 2\theta(1 - \theta), p_{aa} = (1 - \theta)^2 \text{ for some } \theta \in (0, 1).$$

This null hypothesis corresponds to a 1-dimensional sub-model (with a single free parameter  $\theta$ ) inside the 2-dimensional multinomial model (specified by general parameters  $p_{AA}, p_{Aa}, p_{aa}$  summing to 1). We may test  $H_0$  using the GLRT:

The multinomial likelihood is given by

$$l(p_{AA}, p_{Aa}, p_{aa}) = \binom{n}{N_{AA}, N_{Aa}, N_{aa}} p_{AA}^{N_{AA}} p_{Aa}^{N_{Aa}} p_{aa}^{N_{aa}}.$$

Letting  $\hat{p}_{AA}, \hat{p}_{Aa}, \hat{p}_{aa}$  denote the full-model MLEs and  $\hat{p}_{0,AA}, \hat{p}_{0,Aa}, \hat{p}_{0,aa}$  denote the sub-model MLEs, the generalized likelihood ratio is

$$\Lambda = \left( \frac{\hat{p}_{0,AA}}{\hat{p}_{AA}} \right)^{N_{AA}} \left( \frac{\hat{p}_{0,Aa}}{\hat{p}_{Aa}} \right)^{N_{Aa}} \left( \frac{\hat{p}_{0,aa}}{\hat{p}_{aa}} \right)^{N_{aa}},$$

so

$$-2 \log \Lambda = 2N_{AA} \log \frac{\hat{p}_{AA}}{\hat{p}_{0,AA}} + 2N_{Aa} \log \frac{\hat{p}_{Aa}}{\hat{p}_{0,Aa}} + 2N_{aa} \log \frac{\hat{p}_{aa}}{\hat{p}_{0,aa}}. \quad (22.1)$$

The full-model MLEs are given by  $\hat{p}_{AA} = N_{AA}/n$ ,  $\hat{p}_{Aa} = N_{Aa}/n$ , and  $\hat{p}_{aa} = N_{aa}/n$ , by Example 13.4 from Lecture 13. To find the sub-model MLEs, note that under  $H_0$ , the multinomial likelihood as a function of  $\theta$  is

$$\begin{aligned} \text{lik}(\theta) &= \binom{n}{N_{AA}, N_{Aa}, N_{aa}} (\theta^2)^{N_{AA}} (2\theta(1-\theta))^{N_{Aa}} ((1-\theta)^2)^{N_{aa}} \\ &= \binom{n}{N_{AA}, N_{Aa}, N_{aa}} 2^{N_{Aa}} \theta^{2N_{AA}+N_{Aa}} (1-\theta)^{N_{Aa}+2N_{aa}}. \end{aligned}$$

Maximizing the likelihood over parameters  $(p_{AA}, p_{Aa}, p_{aa})$  belonging to the sub-model is equivalent to maximizing the above over  $\theta$ . Differentiating the logarithm of the above likelihood and setting it equal to 0, we obtain the MLE

$$\hat{\theta} = \frac{2N_{AA} + N_{Aa}}{2N_{AA} + 2N_{Aa} + 2N_{aa}} = \frac{2N_{AA} + N_{Aa}}{2n}$$

for  $\theta$ , which yields the sub-model MLEs

$$\begin{aligned} \hat{p}_{0,AA} &= \left( \frac{2N_{AA} + N_{Aa}}{2n} \right)^2 \\ \hat{p}_{0,Aa} &= 2 \left( \frac{2N_{AA} + N_{Aa}}{2n} \right) \left( \frac{N_{Aa} + 2N_{aa}}{2n} \right) \\ \hat{p}_{0,aa} &= \left( \frac{N_{Aa} + 2N_{aa}}{2n} \right)^2. \end{aligned}$$

Substituting these expressions into equation (22.1) yields the formula for  $-2 \log \Lambda$  in terms of the observed counts  $N_{AA}, N_{Aa}, N_{aa}$ . The difference in dimensionality of the two models is  $2 - 1 = 1$ , so an approximate level- $\alpha$  test would reject  $H_0$  when  $-2 \log \Lambda$  exceeds  $\chi_1^2(\alpha)$ .

Rice provides an example (Example 8.5.1A) of genotype data from a population of  $n = 1029$  individuals in Hong Kong, in which the alleles determine the presence of an antigen in the red blood cell. In this example,  $N_{AA} = 342$ ,  $N_{Aa} = 500$ ,  $N_{aa} = 187$ , and we may calculate  $-2 \log \Lambda = 0.0325$ . Letting  $F$  denote the  $\chi_1^2$  CDF, the  $p$ -value of our test is  $1 - F(0.0325) = 0.86$ , so there is no significant evidence of deviation from the Hardy-Weinberg equilibrium.