S&DS 242/542: Theory of Statistics Lecture 16: The generalized likelihood ratio test

GLRT for a simple null hypothesis

Let $f(x \mid \theta)$ be a parametric model, and θ_0 a fixed parameter. Given data $X_1, \ldots, X_n \stackrel{IID}{\sim} f(x \mid \theta)$, we consider testing

$$H_0: \theta = \theta_0$$
 vs. $H_1: \theta \neq \theta_0$

The generalized likelihood ratio test (GLRT) rejects H_0 for large values of the test statistic

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

where $lik(\theta) = \prod_{i=1}^{n} f(X_i \mid \theta)$ is the likelihood function and Ω is the parameter space.

The numerator is the maximum of the likelihood over all possible parameters $\theta \in \Omega$, i.e. $\max_{\theta \in \Omega} \text{lik}(\theta) = \text{lik}(\hat{\theta})$ where $\hat{\theta}$ is the MLE.

Rationale for the GLRT

$$\Lambda = \frac{\mathsf{max}_{\theta \in \Omega} \mathsf{lik}(\theta)}{\mathsf{lik}(\theta_0)} = \frac{\mathsf{lik}(\hat{\theta})}{\mathsf{lik}(\theta_0)}$$

- The GLRT is similar to the likelihood ratio test from the Neyman-Pearson lemma. However, the alternative hypothesis H₁ is *not* of the simple form H₁ : θ = θ₁, so there may not be a uniformly most powerful test. The GLRT replaces lik(θ₁) in the Neyman-Pearson likelihood ratio statistic by lik(θ̂).
- ▶ We always have $lik(\hat{\theta}) \ge lik(\theta_0)$. For large *n*, we expect $lik(\hat{\theta})$ to be closer to $lik(\theta_0)$ when the true parameter is θ_0 , than when the true parameter is a different value, because $\hat{\theta}$ is a consistent estimate of the true parameter. The GLRT tests whether $lik(\hat{\theta})$ exceeds $lik(\theta_0)$ by an amount that is larger than explainable by random chance, if θ_0 were the true parameter.

The rejection region of the GLRT

Equivalently, the GLRT rejects H_0 for large values of

$$2\log \Lambda = 2\ell_n(\hat{\theta}) - 2\ell_n(\theta_0)$$

where $\ell_n(\theta)$ is the log-likelihood,

$$\ell_n(\theta) = \log \operatorname{lik}(\theta) = \sum_{i=1}^n \log f(X_i \mid \theta)$$

Under appropriate conditions, we will see that the null distribution of $2 \log \Lambda$ for large *n* is approximately χ_k^2 , where *k* is the *dimension* of the parameter space. The level- α GLRT rejects H_0 when

$$2\log\Lambda > \chi_k^2(\alpha)$$

where $\chi_k^2(\alpha)$ is the upper- α point of χ_k^2 .

A normal model example

To understand this null distribution χ_k^2 , let us first consider a normal model with a single parameter, so that k = 1.

Let $X_1, \ldots, X_n \stackrel{IID}{\sim} \mathcal{N}(\theta, 1)$ and consider the problem of testing



A normal model example 2 log A = 2 ln (8) - 2 ln (00) $= Zl_{x}(\bar{x}) - Zl_{y}(o)$ $z \geq \sum_{i=1}^{n} \left(-\frac{(X_i \cdot \overline{X})^2}{2} + \frac{(X_i \cdot 0)^2}{2} \right)$ $=\sum_{i=1}^{n}\left(-\left(\chi_{i}^{2}-\zeta\chi_{i}\chi_{i}+\chi_{i}^{2}\right)+\chi_{i}^{2}\right)$ $= \sum_{i=1}^{n} (ZX_i \overline{X} - \overline{X}^{i}) = Zn \overline{X}^{i} - n \overline{X}^{i} = n \overline{X}^{i}$ Under Ho: X~N(0, 1) JAX~N(0,1) $\Rightarrow n \tilde{\chi}^{2}: (J_{n} \tilde{\chi})^{2} \sim \chi^{2}$

For more general parametric models, the following result shows that the sampling distribution of $2 \log \Lambda$ under H_0 is approximately χ_k^2 for large *n*.

Theorem

Let $f(x \mid \theta)$ be a parametric model, where the parameter space Ω has dimension k. Let $X_1, \ldots, X_n \stackrel{IID}{\sim} f(x \mid \theta)$, and let θ_0 be an interior point of Ω . Under regularity conditions, for testing

$$H_0: \theta = \theta_0$$
 vs. $H_1: \theta \neq \theta_0$,

the GLRT statistic $2 \log \Lambda$ converges in distribution to χ_k^2 under H_0 , as $n \to \infty$.

Proof sketch, for k = 1:

$$Z \log \mathcal{A} = Z l_n(\widehat{\Theta}) - Z l_n(\Theta)$$

$$Under H_0: \Theta = \Theta_0, \quad \widehat{\Theta} \to \Theta_0 \quad \text{in probability as } n \to \infty.$$

$$S_0 \quad \widehat{\Theta} - \Theta_0 \quad \text{should typically be small for lypen.}$$

$$Taylor expand \quad l_n(\Theta) \text{ and } \widehat{\Theta}:$$

$$l_n(\Theta_0) \approx l_n(\widehat{\Theta}) + l_n'(\widehat{\Theta}) (\Theta_0 - \widehat{\Theta}) + \frac{1}{2} l_n''(\widehat{\Theta}) (\Theta_0 - \widehat{\Theta})^2$$

$$= 0 \quad \text{ble } \widehat{\Theta} \text{ is the maximizer } \sigma \in l_n \qquad \approx -n \cdot I(\widehat{\Theta}) \approx -n \cdot I(\widehat{\Theta})$$

$$\approx \int_{n} (\hat{o}) - \frac{n}{2} I(\theta_{o}) (\theta_{o} - \hat{\theta})^{2}$$

> Zlay A = Zla(ô) - Zla(0,) $\approx n \cdot \mathcal{I}(\theta_{o}) \cdot (\hat{\Theta} - \theta_{o})^{\nu}$ $\operatorname{Recull}: \operatorname{Jn}(\widehat{O} - O_{0}) \to \mathcal{N}(O, \frac{1}{I(O_{0})}) \text{ under } H_{0}: \Theta = \Theta_{0}$ in distilution as atop. $\rightarrow \int n I(\theta_{0}) \left(\hat{\theta} - \theta_{0} \right) \rightarrow \mathcal{N}(0, 1)$ By cont. mapping that ZI3 N~ n I(0.) (ô-0.)2

Dimension of the parameter space

The "dimension" k of the parameter space Ω is the number of free parameters in the model, which is typically the number of parameters minus the number of distinct equality constraints.

Example: In the model $\mathcal{N}(\theta, 1)$, there is a single parameter $\theta \in \mathbb{R}$ with no constraints, so the dimension of Ω is k = 1.

Example: Let $(X_1, \ldots, X_k) \sim \text{Multinomial}(n, (p_1, \ldots, p_k))$. For a fixed probability vector $(p_{0,1}, \ldots, p_{0,k})$, consider testing

$$H_0: (p_1, \ldots, p_k) = (p_{0,1}, \ldots, p_{0,k})$$
$$H_1: (p_1, \ldots, p_k) \neq (p_{0,1}, \ldots, p_{0,k}).$$

GLRT for multinomial proportions $\mathcal{L}(X_{1,1}, X_{k} | p_{1,1}, p_{k}) = \binom{n}{X_{1} \dots X_{k}} p_{1} p_{2} \sum_{i=1}^{N} p_{k}$ => la(pipple) = log (x, x) + x, log pit + Xkly pic $MLE_{S}: (\hat{p}_{1}, \hat{p}_{k}) = \begin{pmatrix} X_{1} \\ n_{1} \\ \dots \\ n_{r} \end{pmatrix}$ >> Z log A = Zln(pin pu) - Zln(po,1,,po,u) = ZX, log Xiln + + ZXk log Xk/n Port Dimension of JL here is k-1 ble k parouters, I equality construint : pit 1 pk = 1. So reject Ho when US N>X h

Testing a composite null hypothesis

Let $f(x \mid \theta)$ be a parametric model, with parameter space Ω . Given data $X_1, \ldots, X_n \stackrel{ID}{\sim} f(x \mid \theta)$, consider now a test of

 $H_0: \theta \in \Omega_0$ vs. $H_1: \theta \notin \Omega_0$

where $\Omega_0 \subset \Omega$ is a lower dimensional space in the full parameter space Ω (representing a "sub-model" in the full model).

This is usually a setting where $\theta \in \Omega \subset \mathbb{R}^k$ has multiple parameters, some of which remain unspecified under H_0 .

Example: Let $X_1, \ldots, X_n \stackrel{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$. Consider testing

$$H_0: \mu = 0$$
 vs. $H_1: \mu \neq 0$

when the variance σ^2 is unknown. Here

$$\begin{split} \Omega &= \{(\mu,\sigma^2):\sigma^2>0\} \subset \mathbb{R}^2\\ \Omega_0 &= \{(\mu,\sigma^2):\mu=0,\sigma^2>0\} \end{split}$$

 Ω has dimension 2, and Ω_0 has dimension 1.



GLRT for a composite null hypothesis

The GLRT for testing

$$H_0: \theta \in \Omega_0$$
 vs. $H_1: \theta \notin \Omega_0$

rejects H_0 for large values of

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

A is the ratio of the likelihoods evaluated at the MLE $\hat{\theta}$ in the full model Ω and the MLE $\hat{\theta}_0$ in the sub-model Ω_0 .

Equivalently, the GLRT rejects H_0 for large values of

$$2\log \Lambda = 2\ell_n(\hat{\theta}) - 2\ell_n(\hat{\theta}_0)$$

Theorem

Let $f(x \mid \theta)$ be a parametric model, where the parameter space Ω has dimension k, and the parameter space of the sub-model Ω_0 has dimension k_0 . Let $X_1, \ldots, X_n \stackrel{\text{IID}}{\sim} f(x \mid \theta)$. Under regularity conditions, for testing

$$H_0: \theta \in \Omega_0$$
 vs $H_1: \theta \notin \Omega_0$,

the GLRT statistic $2 \log \Lambda$ under any true parameter θ_0 in the interior of Ω_0 converges in distribution to $\chi^2_{k-k_0}$, as $n \to \infty$. Thus the level- α GLRT rejects H_0 when

$$2\log\Lambda\geq\chi^2_{k-k_0}(\alpha)$$

where $k - k_0$ is the difference in dimensions between Ω and Ω_0 . (If $\Omega_0 = \{\theta_0\}$ is a single point, then $k_0 = 0$, reducing to the previous case of a simple null hypothesis.)

Let $X_1, \ldots, X_n \stackrel{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$. Consider testing

 $H_0: \mu = 0$ vs. $H_1: \mu \neq 0$

when $\sigma^2 > 0$ is unknown.

 $f(x|_{\mathcal{M},\sigma^2}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-w)^2}{2\sigma^2}}$ $= \int_{m} (\mu, \sigma^{2}) = \sum_{n=1}^{\infty} \left(-\frac{1}{2} \log 2\pi \sigma^{2} - \frac{(\chi, -\mu)^{2}}{2} \right)$ = - = 1 2 2 0 - - + = = (X. - w) MLEs in the full would: (in or) = (X, the Di (X, -X)2)

In submidd in=0: l (0, 02) = - " 6 2x0" - L 2 X2 Bil (0, v) - - 1 + 1 EX. Set O = 2 ln(0, 02) and some for or $\exists \left(\hat{\mu}_{o}, \hat{\sigma}_{o}^{2}\right) = \left(0, \pm \hat{\Sigma}_{u}^{2} X_{u}^{2}\right)$ So $Z \log \mathcal{A} = 2 l_n(\hat{\mu}, \hat{\sigma}^2) - 2 l_n(\hat{\mu}_0, \hat{\sigma}_0^2)$ = Zla(x, 22)-Zla(0, 22)

= Z(-2 1, 2.0 - 1, 2 (x-x) + 2 1, 2.0 + 1, 2x) = n log -= n log - z X. - + z (X. - X) Difference of dimensions: 2-1=1. So reject Ho when Zly A > Xila)

Relationship to the t-test

In this example, the GLRT is in fact equivalent to a two-sided t-test:

32= + = (X,-x)= += (X,-2X, x+x) $= (\underbrace{1}_{x} \underbrace{\widetilde{\Sigma}}_{x} \underbrace{\chi}_{y}) - \underbrace{2}_{x} \underbrace{\widetilde{\chi}}_{x} \underbrace{\chi}_{y}^{2} = (\underbrace{1}_{x} \underbrace{\widetilde{\Sigma}}_{x} \underbrace{\chi}_{y}) - \underbrace{\chi}_{z}^{2} = \widehat{\varphi}_{z}^{2} - \underbrace{\chi}_{z}^{2}$ $\Rightarrow Zl_{3} \Lambda^{z} n l_{3} \frac{\hat{\sigma}^{z}}{\hat{\sigma}^{z}} = n l_{y} \frac{\hat{\sigma}^{z} + \tilde{\chi}^{z}}{\hat{\sigma}^{z}} = n l_{y} \left(l + \frac{\tilde{\chi}^{z}}{\hat{\sigma}^{z}} \right)$ · Recall the t-statistic: T= Jnx = Jnx 18

Relationship to the t-test

Relationship to the t-test

Thus $2 \log \Lambda$ is an increasing function of |T|. Rejecting H_0 for large values of $2 \log \Lambda$ is equivalent to rejecting H_0 for large values of |T|.

A difference between the GLRT and t-test is that the GLRT threshold is asymptotic, whereas the t-test threshold is exact:

- The usual two-sided t-test would reject H₀ when |T| > t^(α/2)_{n-1}, the upper-^α/₂ point of the t-distribution. This achieves Type I error probability exactly α when the data are normal.
- The GLRT may be thought of as making the approximations

$$2\log \Lambda = n\log\left(1 + \frac{1}{n-1}T^2\right) \approx \frac{n}{n-1}T^2 \approx T^2$$

and also approximating the distribution of T^2 by the χ_1^2 distribution of Z^2 , where $Z \sim \mathcal{N}(0, 1)$. These approximations become more accurate as *n* increases.

Testing Hardy-Weinberg equilibrium

At a single locus in the genome with two possible alleles A and a, an 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

$$(X_{AA}, X_{Aa}, X_{aa}) \sim \mathsf{Multinomial}(n, (p_{AA}, p_{Aa}, p_{aa}))$$

When the alleles A and a are present in the population with proportions $1 - \theta$ and θ , then under an assumption of random mating, genetics theory predicts that

$$p_{AA} = (1 - \theta)^2, \quad p_{Aa} = 2\theta(1 - \theta), \quad p_{aa} = \theta^2$$

This is called the hypothesis of Hardy-Weinberg equilibrium.

Testing Hardy-Weinberg equilibrium

We may specify Hardy-Weinberg equilibrium as the null hypothesis

$$H_0: p_{AA} = (1- heta)^2, \ p_{Aa} = 2 heta(1- heta), \ p_{aa} = heta^2$$
 for some $heta \in [0,1]$

This corresponds to testing a 1-dimensional sub-model

$$\Omega_0 = \left\{ \left((1- heta)^2, 2 heta(1- heta), heta^2
ight) : heta \in [0,1]
ight\}$$

inside the 2-dimensional multinomial model

$$\Omega = \left\{ (p_{AA}, p_{Aa}, p_{aa}) : p_{AA}, p_{Aa}, p_{aa} \ge 0 \text{ and } p_{AA} + p_{Aa} + p_{aa} = 1 \right\}$$

Testing Hardy-Weinberg equilibrium

To carry out this test using the GLRT:

Qn (pAA, PA, Pan) = log (XAA, XAN, Xm) + XAA & PAA + XAn by PAn + Xan by Pan · Full-midd MLES: (pAN, pAn, pan)= (XM Xm Xm) · Sabordad MLE (For Lever 12): 0= ZXan + XAn => (po, AA, po, A., po, ...) = ((1-ô), 2ô(1-d), ô2) =) Zlo L= Zla (pA, pa, pa) - Zla (po, AA, po, A., po, a) Dillen in diversions: Z-1=1. So give the when ZGL > Xi lad

Example from Rice

Example 8.5.1A of Rice's text analyzes genotype data from n = 1029 individuals in Hong Kong, in which the alleles determine the presence of an antigen in the red blood cell. In this example,

$$X_{AA} = 342, X_{Aa} = 500, X_{aa} = 187$$

We may calculate

$$\hat{p}_{AA} = 0.332, \ \hat{p}_{Aa} = 0.486, \ \hat{p}_{aa} = 0.182$$

and $\hat{\theta} = \frac{2 \times 187 + 500}{2 \times 1029} = 0.425,$
 $\hat{p}_{0,AA} = 0.331, \ \hat{p}_{0,Aa} = 0.489, \ \hat{p}_{aa} = 0.180$

This gives $2 \log \Lambda = 0.0325$. Letting *F* be the χ_1^2 CDF, the *p*-value for the GLRT is 1 - F(0.0325) = 0.86, so there is no significant evidence of deviation from Hardy-Weinberg equilibrium.