5. **Approximate GLR Tests**

a) Rejecting for small values of the GLR test statistic Λ is equivalent to rejecting for large values of $-2\log(\Lambda)$. Theorem A on p. 341 states that for large samples from Exponential family distributions, the distribution of $-2\log\Lambda$ is approximately $\chi^2_{(\nu)}$, where ν is the difference in the number of parameters that need to be estimated overall, and under H_o. Show (as in Example A on p. 339) that this is exactly true for testing a Normal mean with known σ, and approximately true when σ is unknown (using the result of 4d).

Transition from GLR test to the approximate GLR: Rejecting small values of the GLR test statistic Λ is equivalent to rejecting large values of $-2\log(\Lambda)$

Theorem A. Under smoothness conditions on the probability density or frequency functions involved, the null distribution of $-2\log\Lambda$ tends to a chi-square distribution with degrees of freedom equal to $\dim\Omega - \dim\omega_0$ as the sample size tends to infinity □

* $\dim\Omega$ and $\dim\omega_0$ → numbers of free parameters under Ω and ω_0 (capital and lowercase omega)

→ Phil adjusts this theorem by transitioning the smoothness condition to exponential family distributions.

Phil's version: for large samples from Exponential family distributions, the distribution of $-2\log\Lambda$ is approximately $\chi^2_{(\nu)}$ (with ν degrees of freedom), where ν is the difference in the number of parameters that need to be estimated overall, and under Ho.

Part (A) asks us to show that this is exactly true when testing a Normal mean with known σ, and approximately true when σ is unknown.

The distribution of $-2\log\Lambda$ is exactly $\chi^2_{(\nu)}$ when testing a normal mean with known σ

Let $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ $\quad \sigma^2$ known

likelihood function: $L(\mu) = \prod \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}\right]$

$$\propto e^{-n(\bar{x}-\mu)^2/2\sigma^2}$$

* proportional as the rest is constant and can be cancelled out
* $\bar{x}$ from the expansion of $(x_i-\mu)^2$

$H_o: \mu = \mu_0$ vs. $H_a: \mu \neq \mu_0$

as previously shown → GLR: $\Lambda = \dfrac{\max_{H_0} L(\mu)}{\max L(\mu)} = \dfrac{L(\mu_0)}{L(\bar{x})}$ → MLE

In this case, we reject Ho for small Λ as likelihood for the MLE would be larger than likelihood for the null.

We can now use this test statistic Λ and the approximate GLR to show this is exactly $\chi^2_{(\nu)}$

approximate GLR: $-2\log(\Lambda) = -2(\ell(\mu_0) - \ell(\bar{x}))$

↳ reject for large $(-2\log(\Lambda))$

$$= 2(\ell(\bar{x}) - \ell(\mu_0)) \quad \text{* distribute the negative sign}$$

$$= 2\left(-\underbrace{\frac{n(\bar{x}-\bar{x})^2}{2\sigma^2}}_{\emptyset} + \frac{n(\bar{x}-\mu_0)^2}{2\sigma^2}\right)$$

$$= \frac{(\bar{x}-\mu_0)^2}{\sigma^2/n} = \left(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right)^2 = z^2 \quad \text{→ z-score squared}$$

under Ho, $z = \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

$\rightarrow z^2 \sim \chi^2_{(1)}$

ν = difference in number of parameters to be estimated

exactly chi-squared $\rightarrow -2\log(\Lambda) \dot\sim \chi^2_{(\nu)}$

$\boxed{\chi^2_{(1)}}$

→ In this case, σ^2 is known so only estimating for μ.

σ^2 unknown $\rightarrow$ approximately chi squared

same normal distribution as before

$H_0 : \mu = \mu_0$ $\rightarrow$ σ unknown, but μ known so only 1 unknown parameter

$H_a : \mu \neq \mu_0$ $\rightarrow$ σ and μ unknown, so 2 unknown parameters

$\rightarrow \chi^2_{(v)} = v = 2-1 = 1$ degree of freedom $\rightarrow$ our goal is to show the approximate GLR is approximate chi-square with 1 degree of freedom

GLR: $\Lambda = \dfrac{L(\mu_0, \hat{\sigma}_0^2)}{L(\bar{x}, \hat{\sigma}_1^2)}$ $\hat{\sigma}_0^2 = \dfrac{\sum(x_i - \mu_0)^2}{n}$

$\hat{\sigma}_1^2 = \dfrac{\sum(x_i - \bar{x})^2}{n}$

Previously shown: $\Lambda = \left(\dfrac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2}\right)^{-n/2} = \left(\dfrac{\sum(x_i - \mu_0)^2}{\sum(x_i - \bar{x})^2}\right)^{-n/2} = \left(\dfrac{\sum(x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2}{\sum(x_i - \bar{x})^2}\right)^{-n/2} = \left(1 + \dfrac{n(\bar{x} - \mu_0)^2}{(n-1)s^2}\right)^{-n/2} = \left(1 + \dfrac{T^2}{n-1}\right)^{-n/2}$

It was previously shown how to find Λ with 2 unknown parameters

my problem: $\Lambda = \dfrac{L(\mu_0, \hat{\sigma}_0^2)}{L(\bar{x}, \hat{\sigma}_1^2)} = \left(1 + \dfrac{T^2}{n-1}\right)^{-n/2}$, $T = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ $\sim t(n-1)$ under H_0

now we can apply the approximate GLR test to see this is approximately $\chi^2_{(1)}$

approximate GLR :

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

note : for large n, $\log(1 + \frac{a}{n}) \approx \frac{a}{n}$

so, for large n, $= \dfrac{nT^2}{-1} \approx T^2 \approx z^2$ (t distribution is approximately normal with large n)

df $\rightarrow \infty$ $t \rightarrow N(0,1)$

$\boxed{\text{Therefore, } T^2 \sim \chi^2_{(1)}}$

b) Show how to use Lagrange multipliers to find the maximum likelihood estimates for multinomial cell probabilities.

we run into difficulty finding the MLE when we have multinomial cell probabilities. So, we can use the lagrange multiplier to help us with this.

multinomial cell probabilities:

counts $X_1, \ldots, X_m$ with probabilities $\theta_1, \theta_2, \ldots, \theta_m$ $\quad \sum \theta_j = 1$ → sum of the probabilities is 1

Rolling dice where each outcome has a certain probability

note: j = category (such as $1, \ldots, 6$ in a dice example)

$$L(\theta_1, \ldots, \theta_m) \propto \prod_{j=1}^{m} \theta_j^{x_j} \, I(\Sigma \theta_j = 1)$$

log likelihood $\ell(\theta_1, \ldots, \theta_m) = \Sigma x_i \log \theta_j + \text{constant}$

to find the MLE → $\frac{\partial}{\partial \theta_1} = \frac{x_i}{\theta}$ → this is a dead end

↳ θ is bounded by $(0,1)$, so derivatives in this case do not account for this restraint

To fix this, we can apply the

Lagrange multiplier.

$$\ell(\theta_1, \ldots, \theta_m) = \Sigma x_j \log \theta_j + \lambda(\Sigma \theta_j - 1)$$

→ note: $(\Sigma \theta_j - 1) = 0$ so this does not change overall equation

→ note: there is also a term that is constant with no θ's (as we had before we added the lagrange multiplier)

what this allows us to do is have useful/informative partial derivatives

Partial derivatives:

$$\frac{\partial}{\partial \theta_j} = \frac{x_j}{\theta_j} + \lambda \longrightarrow \hat{\theta}_j = \frac{-x_j}{\lambda} = \frac{x_j}{n}$$

$$\frac{\partial}{\partial \lambda} = \sum_{j=1}^{m} \theta_j - 1 \longrightarrow 1 = \Sigma \hat{\theta}_j = \frac{-\Sigma x_j}{\lambda} = \frac{-n}{\lambda}$$

$$\hat{\lambda} = -n$$

$$\boxed{\text{MLEs: } \quad \hat{\theta}_j = \frac{x_j}{n} \quad \text{and} \quad \hat{\lambda} = -n}$$

joint MLE for $\theta_1, \ldots, \theta_m$

$$\hat{\theta}_1 = \frac{x_1}{n}, \quad \hat{\theta}_2 = \frac{x_2}{n}, \quad \ldots, \quad \hat{\theta}_m = \frac{x_m}{n}$$

also, we can see that the sum of these probabilities equals 1

$$\Sigma \hat{\theta}_j = \frac{\Sigma x_j}{n} = \frac{n}{n} = 1$$

c) Show how to test multinomial probabilities using the approximate GLR test. Show that the GLR test statistic is asymptotically equivalent to the Pearson Chi-square statistic $\sum \frac{(O-E)^2}{E}$, where O and E are observed and expected counts. As an example, suppose $n = 30$ rolls of a 6-sided die result in counts of (10,5,5,5,5,0) for the outcomes $1, 2, ..., 6$ (this is what I would expect for my biased die that has two 1's and no 6). Compute the P-value of a test for $H_o : p_1 = p_2 = \ldots = p_6 = 1/6$ against a general alternative using the Chi-square approximation. Also use simulation to compute an exact P-value (up to simulation error). Simulate many sets of 30 fair dice rolls and compute the GLR stat for each replicate data set. Estimate the P-value as the proportion of times you get a statistic as large or larger than the observed value.

test multinomial probabilities using the approximate GLR test:

for multinomial cell probabilities: $\Lambda = \dfrac{\max_{H_0} L(\theta_1,\ldots,\theta_m)}{\max L(\theta_1,\ldots,\theta_m)}$ $\rightarrow$ max under H_0 : $\sum \log (P_j)^{x_j} = n\log(P_j)$ $\rightarrow$ max overall : $\sum \log\left(\left(\frac{x_j}{n}\right)^{x_j}\right)$ $\rightarrow$ we do this so that in the case $X_j = 0$, we still get a value $\rightarrow 0^0 \rightarrow 1$

for multinomial cell probabilities, we can utilize various chi-square statistics for hypothesis testing.

Chi-square statistics:

GLR stat (likelihood chi-square) : $2\sum O_j \log\left(\frac{O_j}{E_j}\right)$
(Neyman - Pearson Lemma)

E_j = expected count
O_j = observed count
j = category's

Pearson stat (pearson chi-square) : $\sum_{j=1}^{m} \frac{(O_j)-E_j)^2}{E_j}$
(Karl Pearson)

These two chi-square stats are similar but not equivalent (will get different test statistic values). $\rightarrow$ so we say they are asymptotically equivalent
This may influence if the statistic provides evidence or not to reject the null hypothesis.

Example: $n = 30$ rolls of a 6-sided die results in counts (10, 5, 5, 5, 5, 0) for the outcomes 1, 2, 3, 4, 5, 6 $\rightarrow m = 6$

$H_0 : \theta_j = 1/6$ vs. $H_a : \theta_j \neq 1/6$
(fair dice) (not fair dice)

approximate GLR:

$$-2\log(\Lambda) = -2n\log(1/6) + \sum X_j \log(X_j/n)$$
$$= \sum \log\left(\left(\tfrac{x_j}{n}\right)^{x_j}\right) - 2n\log(1/6)$$

use R to get this test statistic value:

what this gives us is the cut off for approximate GLR test of $H_0 : p_1 = p_2 = \ldots = p_6 = \frac{1}{6}$ vs H_a : not H_0

qchisq(0.95, 5) $\rightarrow$ reject for $-2\log(\Lambda) > 11.0705$

GLR stat : $2\left(\sum \log\left(\frac{x_j}{n}\right)^{x_j}\right) - n\log(1/6))$

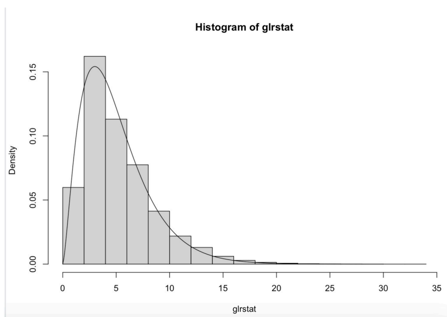
$L_0 = \sum \log\left(\frac{1}{6}\right)^{x_j}$ x_j = observed counts
$L_1 = \sum \log\left(\frac{x_j}{n}\right)^{x_j}$

observed GLR $= -2(L_0 - L_1)$ for the $n = 30$ and observing this outcome

test statistic of 13.86294 > 11.0705 $\rightarrow$ so significant at 0.05 (reject the null hypothesis that this is a fair die)

p-value: 1 - pchisq(13.86294, 5) = 0.01650388
leans on approximation

more accurate: see how often simulate n = 100,000



when the code runs the simulation with n = 100,000, it generates a a table for cutoff that can be visualized by this histogram.
This is the null sample distribution the GLR stats in each trial
We can see how often we get values as extreme and determine the rejection regions.

estimate exact P-value:

glrstat = rep(0, nsim)

p-value : 0.02582 $\rightarrow$ so still significant at $\alpha = 0.05$ level but larger p-value

chi square approximation $\rightarrow$ histogram

$\hookrightarrow$ better for continuous distributions, so in this case, the approximate GLR is actually a best test.

d) (If time) Consider a test of $H_o : \theta = \theta_o$ vs. $H_a : \theta \neq \theta_o$, based on $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Unif}(0, \theta)$. Explain why this distribution is not in the Exponential family. Show the null distribution is exactly Chi-square, but with 2 df, not the 1 df prescribed by Theorem A.

Now, lets consider a hypothesis test based on a uniform variable

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

$X_1, \ldots, X_n \sim Unif(0, \theta)$

$L(\theta) = \frac{1}{\theta^n} I(\theta > X_{(n)})$

note: MLE = $X_{(n)}$

This likelihood function shows us that this distribution is not part of the exponential family as the density of the data in this distribution depends on θ (which is what we need to be able to do with an exponential family distribution)

$\rightarrow f(x_i|\theta) = \frac{1}{\theta} I(0 < x_i < \theta)$

can't separate θ

show the null distribution is exactly chi-square, but with 2 degrees of freedom, and not the 1 as prescribed in theorem A.

GLR: $\Lambda = \frac{(1/\theta_0)^n}{(1/x_{(n)})^n} = \left(\frac{X_{(n)}}{\theta}\right)^n$

approximate GLR: $-2\log(\Lambda) = -2n\log\left(\frac{X_{(n)}}{\theta}\right) = -2n\log(U_{(n)})$

one way to think about this is through the cumulative density function (CDF).

Let $U_i = \left(\frac{X_i}{\theta}\right) \sim Unif(0,1)$

this implies the largest U_n $\rightarrow U_n = \frac{X_n}{\theta} \sim Beta(n, 1)$ $\rightarrow$ Beta(1,1) is exactly the uniform distribution

start w/ PDF: pdf: $f_{U_{(n)}}(u) = nu^{n-1} \quad 0 < u < 1$

cdf: $F_{U_n}(u) = u^n \quad 0 < u < 1$

Let $Y = -2\log(U_{(n)})$ $\rightarrow$ we can recognize this as the approximate GLR stat

CDF method: for Y

$F_Y(y) = P(-2n\log(U_{(n)}) \leq y) = P(-\log(U_{(n)}) \leq \frac{y}{2n}) = P(U_n > e^{-y/2n})$

$F_Y(y) = P(U_n > e^{-y/2n}) = 1 - e^{-y/2}$

* note: this is the CDF for Expo$(\frac{1}{2})$, or Gamma$(1, 1/2)$ or $\chi^2_{(2)}$

Therefore,

Approximate GLR test rejects for large values of $Y = -2\log(\Lambda)$ $\longrightarrow$ $-2\log(\Lambda) = -2n\log\left(\frac{X_{(n)}}{\theta}\right)$

As just shown, under H_0, this is $\chi^2_{(2)}$

prescription says we should only be using 1 degree of freedom $(df = 1 - 0 = 1)$ as we are just estimating the mean. but the math works out to be estimating χ^2_2. This will be a tighter chi-square than expected.