

# STAT 801=830

## Likelihood Ratio Tests

Richard Lockhart

Simon Fraser University

STAT 801 — Fall 2012



# Purposes of These Notes

- Describe likelihood ratio tests
- Discuss large sample  $\chi^2$  approximation.
- Discuss level and power



# Likelihood Ratio Tests

- For general composite hypotheses optimality theory is not usually successful in producing an optimal test.
- Instead we look for heuristics to guide our choices.
- The simplest approach is to consider the likelihood ratio

$$\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}$$

and choose values of  $\theta_1 \in \Theta_1$  and  $\theta_0 \in \Theta_0$  which are reasonable estimates of  $\theta$  assuming respectively the alternative or null hypothesis is true.

- The simplest method is to make each  $\theta_i$  a maximum likelihood estimate, but maximized only over  $\Theta_i$ .



## Example 1: $N(\mu, 1)$

- Test  $\mu \leq 0$  against  $\mu > 0$ . (Remember UMP test. )
- Log likelihood is

$$-n(\bar{X} - \mu)^2/2$$

- If  $\bar{X} > 0$  then global maximum in  $\Theta_1$  at  $\bar{X}$ .
- If  $\bar{X} \leq 0$  global maximum in  $\Theta_1$  at 0.
- Thus  $\hat{\mu}_1$  which Max  $\ell(\mu)$  subject to  $\mu > 0$  at  $\hat{\mu}_1 = \bar{X}1(\bar{X} > 0)$ .
- Similarly,  $\hat{\mu}_0$  is  $\bar{X}$  if  $\bar{X} \leq 0$  and 0 if  $\bar{X} > 0$ .
- Hence

$$\frac{f_{\hat{\theta}_1}(X)}{f_{\hat{\theta}_0}(X)} = \exp\{\ell(\hat{\mu}_1) - \ell(\hat{\mu}_0)\} = \exp\{n\bar{X}|\bar{X}|/2\}$$

- Monotone increasing function of  $\bar{X}$  so rejection region has form  $\bar{X} > K$ .
- To get level  $\alpha$  reject if  $n^{1/2}\bar{X} > z_\alpha$ .
- Notice simpler statistic is *log likelihood ratio*

$$\lambda \equiv 2 \log \left( \frac{f_{\hat{\mu}_1}(X)}{f_{\hat{\mu}_0}(X)} \right) = n\bar{X}|\bar{X}|$$



## Example 2: $H_o : \mu = 0$ in $N(\mu, 1)$

- Value of  $\hat{\mu}_0$  is 0
- Maximum of log-likelihood over alternative  $\mu \neq 0$  occurs at  $\bar{X}$ .
- This gives

$$\lambda = n\bar{X}^2$$

which has a  $\chi_1^2$  distribution.

- This test leads to the rejection region  $\lambda > (z_{\alpha/2})^2$  which is the usual (UMPU) z-test.



### Example 3: $N(\mu, \sigma^2)$ model, $\mu = 0$ against $\mu \neq 0$

- Must find two estimates of  $\mu, \sigma^2$ .
- Maximum likelihood over alternative occurs at global mle  $\bar{X}, \hat{\sigma}^2$ .
- We find

$$\ell(\hat{\mu}, \hat{\sigma}^2) = -n/2 - n \log(\hat{\sigma})$$

- Maximize  $\ell$  over null hypothesis.
- Recall

$$\ell(\mu, \sigma) = -\frac{1}{2\sigma^2} \sum (X_i - \mu)^2 - n \log(\sigma)$$

- On null  $\mu = 0$  so find  $\hat{\sigma}_0$  by maximizing

$$\ell(0, \sigma) = -\frac{1}{2\sigma^2} \sum X_i^2 - n \log(\sigma)$$



## LRT – general description

- This leads to

$$\hat{\sigma}_0^2 = \sum X_i^2 / n$$

and

$$\ell(0, \hat{\sigma}_0) = -n/2 - n \log(\hat{\sigma}_0)$$

- This gives

$$\lambda = -n \log(\hat{\sigma}^2 / \hat{\sigma}_0^2)$$

- Since

$$\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} = \frac{\sum (X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2 + n\bar{X}^2}$$

we can write

$$\lambda = n \log(1 + t^2 / (n - 1))$$

where

$$t = \frac{n^{1/2} \bar{X}}{s}$$

is the usual  $t$  statistic.

- LRT rejects for large values of  $|t|$  — the usual test.



# LRT – general description

- Notice that if  $n$  is large we have

$$\lambda \approx n[1 + t^2/(n-1) + O(n^{-2})] \approx t^2.$$

- Since  $t$  statistic is approximately standard normal if  $n$  large we see

$$\lambda = 2[\ell(\hat{\theta}_1) - \ell(\hat{\theta}_0)]$$

has nearly a  $\chi_1^2$  distribution.

- General phenomenon when null hypothesis has form  $\phi = 0$ .
- Here is the general theory.
- Suppose vector  $\theta$  of  $p + q$  parameters partitioned into  $\theta = (\phi, \gamma)$  with  $\phi$  a vector of  $p$  parameters and  $\gamma$  a vector of  $q$  parameters.
- To test  $\phi = \phi_0$  we find two mles of  $\theta$ .
- First: global mle  $\hat{\theta} = (\hat{\phi}, \hat{\gamma})$  maximizes likelihood over  $\Theta_1 = \{\theta : \phi \neq \phi_0\}$  (typically  $P_\theta(\hat{\phi} = \phi_0) = 0$ ).





# LRT – general description

- Maximize likelihood over null hypothesis, that is find  $\hat{\theta}_0 = (\phi_0, \hat{\gamma}_0)$  to maximize

$$\ell(\phi_0, \gamma)$$

- The log-likelihood ratio statistic is

$$2[\ell(\hat{\theta}) - \ell(\hat{\theta}_0)]$$

- Now suppose that the true value of  $\theta$  is  $\phi_0, \gamma_0$  (so that the null hypothesis is true).
- The score function is a vector of length  $p + q$  and can be partitioned as  $U = (U_\phi, U_\gamma)$ .
- The Fisher information matrix can be partitioned as

$$\begin{bmatrix} \mathcal{I}_{\phi\phi} & \mathcal{I}_{\phi\gamma} \\ \mathcal{I}_{\gamma\phi} & \mathcal{I}_{\gamma\gamma} \end{bmatrix}.$$



# Large sample theory for LRT

- According to our large sample theory for the mle we have

$$\hat{\theta} \approx \theta + \mathcal{I}^{-1}U$$

and

$$\hat{\gamma}_0 \approx \gamma_0 + \mathcal{I}_{\gamma\gamma}^{-1}U_\gamma$$

- Two term Taylor expansions of both  $\ell(\hat{\theta})$  and  $\ell(\hat{\theta}_0)$  around  $\theta_0$  give

$$\ell(\hat{\theta}) \approx \ell(\theta_0) + U^t \mathcal{I}^{-1}U + \frac{1}{2}U^t \mathcal{I}^{-1}V(\theta)\mathcal{I}^{-1}U$$

where  $V$  is the second derivative matrix of  $\ell$ .



# Large sample theory for LRT

- Remember that  $V \approx -\mathcal{I}$  and you get

$$2[\ell(\hat{\theta}) - \ell(\theta_0)] \approx U^t \mathcal{I}^{-1} U.$$

- A similar expansion for  $\hat{\theta}_0$  gives

$$2[\ell(\hat{\theta}_0) - \ell(\theta_0)] \approx U_{\gamma}^t \mathcal{I}_{\gamma\gamma}^{-1} U_{\gamma}.$$

- If you subtract these you find that

$$2[\ell(\hat{\theta}) - \ell(\hat{\theta}_0)]$$

can be written in the approximate form

$$U^t M U$$

for a suitable matrix  $M$ .

- Now use general theory of distribution of  $X^t M X$  where  $X$  is  $MVN(0, \Sigma)$ .



# The theorem: large sample theory of LRT

The ideas above lead to a proof of the following theorem.

*The log-likelihood ratio statistic*

$$\lambda = 2[\ell(\hat{\theta}) - \ell(\hat{\theta}_0)]$$

*has, under the null hypothesis, approximately a  $\chi_p^2$  distribution.*



# Quadratic forms and $\chi^2$

In proving the main theorem we need some facts about quadratic forms.

Suppose  $X \sim MVN(0, \Sigma)$  with  $\Sigma$  non-singular and  $M$  is a symmetric matrix. If  $\Sigma M \Sigma M \Sigma = \Sigma M \Sigma$  then  $X^t M X$  has a  $\chi^2_\nu$  distribution with  $df$   $\nu = \text{trace}(M \Sigma)$ . The condition simplifies to  $M \Sigma M = M$



## Proof

- We have  $X = AZ$  where  $AA^t = \Sigma$  and  $Z$  is standard multivariate normal.
- So  $X^tMX = Z^tA^tMAZ$ .
- Let  $Q = A^tMA$ .
- Since  $AA^t = \Sigma$  condition in the theorem is

$$AQQA^t = AQA^t$$

- Since  $\Sigma$  is non-singular so is  $A$ .
- Multiply by  $A^{-1}$  on left and  $(A^t)^{-1}$  on right; get  $QQ = Q$ .
- $Q$  is symmetric so  $Q = P\Lambda P^t$  where  $\Lambda$  is diagonal matrix containing the eigenvalues of  $Q$  and  $P$  is orthogonal matrix whose columns are the corresponding orthonormal eigenvectors.
- So rewrite

$$Z^tQZ = (P^tZ)^t\Lambda(PZ).$$



## More proof

- $W = P^t Z$  is  $MVN(0, P^t P = I)$ ; i.e.  $W$  is standard multivariate normal.
- Now

$$W^t \Lambda W = \sum \lambda_i W_i^2$$

- We have established that the general distribution of any quadratic form  $X^t M X$  is a linear combination of  $\chi^2$  variables.
- Now go back to the condition  $Q Q = Q$ .
- If  $\lambda$  is an eigenvalue of  $Q$  and  $v \neq 0$  is a corresponding eigenvector then  $Q Q v = Q(\lambda v) = \lambda Q v = \lambda^2 v$  but also  $Q Q v = Q v = \lambda v$ .
- Thus  $\lambda(1 - \lambda)v = 0$ .
- It follows that either  $\lambda = 0$  or  $\lambda = 1$ .



## End of proof

- This means that the weights in the linear combination are all 1 or 0 and that  $X^tMX$  has a  $\chi^2$  distribution with degrees of freedom,  $\nu$ , equal to the number of  $\lambda_i$  which are equal to 1.
- This is the same as the sum of the  $\lambda_i$  so

$$\nu = \text{trace}(\Lambda)$$

- But

$$\begin{aligned}\text{trace}(M\Sigma) &= \text{trace}(MAA^t) \\ &= \text{trace}(A^tMA) \\ &= \text{trace}(Q) \\ &= \text{trace}(P\Lambda P^t) \\ &= \text{trace}(\Lambda P^tP) \\ &= \text{trace}(\Lambda)\end{aligned}$$





# Application to LRT

- In the application  $\Sigma$  is  $\mathcal{I}$  the Fisher information and  $M = \mathcal{I}^{-1} - J$  where

$$J = \begin{bmatrix} 0 & 0 \\ 0 & \mathcal{I}_{\gamma\gamma}^{-1} \end{bmatrix}$$

- It is easy to check that  $M\Sigma$  becomes

$$\begin{bmatrix} I & 0 \\ -\mathcal{I}_{\gamma\phi}\mathcal{I}_{\phi\phi} & 0 \end{bmatrix}$$

where  $I$  is a  $p \times p$  identity matrix.

- It follows that  $\Sigma M \Sigma M \Sigma = \Sigma M \Sigma$  and  $\text{trace}(M\Sigma) = p$ .

