exam: can bring 1 double-sided A4 w formulas created by you.

Lecture 9. Large Sample Tests.

|| used in standard setting

* we can control type 1 - size

Likelihood Ratio Test

All of these tests are MLE: have parametric family, know all but θ

$X_1, \ldots, X_n \sim$ i.i.d. $f(x|\theta)$ where θ is scalar

$H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$.

only at $\neq$

likelihood → for whole sample

$f(x|\theta)$ f i.e. so joint distn

Ratio betwn 0 & 1 numerator smaller than denominator

$$\lambda(x) = \frac{\mathcal{L}(\theta_0|x)}{\mathcal{L}(\hat{\theta}_{ML}|x)}$$

argmax $L(\theta)$

Theorem 1.

Under the same regularity conditions as the MLE theory and if $H_0 : \theta = \theta_0$

under null: H_0 we have chi-sq

calc. from data → $\lambda R =$

$$-2 \log \lambda(X) \Rightarrow \chi^2_1.$$

- one tail: bec (gaussian)^2;
- using χ^2

suggest not loose much.

Cut right tail

& if $\lambda R \leq \chi^2_1\,(0.95)$ accept

1 degree of freedom

likelihood fn locally quadratic

$\hat{\theta}_{ML}$ θ_0

see how close θ_{null} to $\hat{\theta}_{ML}$. want likely close.

Likelihood Ratio Test

Proof : $LR = -2 \log \dfrac{L(\theta_0)}{L(\hat\theta_{ML})} = -2\left(\ell(\theta_0) - \ell(\hat\theta_{ML})\right)$

Taylor approx at second order, at max $\hat\theta_{ML}$, will be derivative at θ_{ML} is 0 $\neq 0$.

$$= -2\Big[\underbrace{\frac{\partial \ell}{\partial \theta}(\hat\theta_{ML})}_{=0}(\theta_0 - \hat\theta_{ML}) + \frac{1}{2}\frac{\partial^2 \ell(\theta^*)}{\partial\theta\partial\theta}(\theta_0 - \hat\theta_{ML})^2.$$

$$= \underbrace{\frac{1}{n}\cdot\frac{\partial^2 \ell_n}{\partial\theta\partial\theta}(\theta^*)}\Big[\underbrace{\sqrt{n}(\theta_0 - \hat\theta_{ML})}\Big]$$ Normalize

we saw before that this converges to Fisher info

$\longrightarrow N(0, I_1^{-1})$

$$\Longrightarrow I_1 \cdot \left(N(0, I_1^{-1})\right)^2 = N(0,1)^2 = \chi^2_1$$

chi-sq w one deg. of freedom

Likelihood Ratio Test: Multi-dimensional

$X = (X_1 \dots X_n) \sim f(x|\theta)$ joint distbn.

- θ is k-dimensional parameter
- $H_0 : \{\theta \in \Theta : g_1(\theta) = 0, ..., g_p(\theta) = 0\}$
- $g_1(\theta) = 0$, ..., $g_p(\theta) = 0$ are restrictions on the model and $k \geq p$
- assume that restrictions are jointly linear independent

Under some regularity conditions (including smoothness of $g_1, ..., g_p$):

$$LR = 2\left(\underbrace{\max_{\theta\in\Theta}\ell(\theta|X)}_{=\ell(\hat{\theta}_{ML})} - \max_{\theta\in\Theta_0}\ell(\theta|X)\right) \Rightarrow \chi^2_p$$

opt. over restricted space set by null hypo

- $\arg\max_{\Theta_0} = \hat{\theta}_0$ (restricted estimator)

under H_0.

Ex) $X_1, \dots, X_n \sim \text{iid } N(\underbrace{\mu, \sigma^2}_{})$ $k=2$ bec two parameters to est.

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

$p = 1$: one restriction.

believe μ is true so set 0.

$$LR = 2\left[\max_{\mu,\sigma^2} \ell(\mu, \sigma^2) - \max_{\sigma^2} \ell(0, \sigma^2)\right] \overset{\text{under } H_0}{\Longrightarrow} \chi^2_1$$

unrestricted

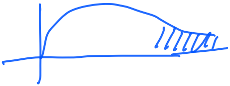
restricted

Then here,

$(\hat{u}_{ML}, \hat{\sigma}^2_{ML}) \quad (0, \hat{\sigma}^2_0)$

Likelihood Ratio Test: Multi-dimensional

- denote $\hat{\theta}_0 = \arg\max_{\theta \in \Theta_0} \ell(\theta|X)$ to be the restricted estimate
- $\ell(\theta|X)$ is log-likelihood
- $LR = 2(\ell(\hat{\theta}_{ML}|X) - \ell(\hat{\theta}_0|X))$
- LR test of level α rejects the null hypothesis if and only if $LR > \chi^2_p(1-\alpha)$.



cut 5%

Example

$X_1, \ldots, X_n \sim$ i.i.d. Poisson(λ)

$f(x|\lambda) = \lambda^x e^{-\lambda}/x!$ for $x = 0, 1, 2, \ldots$.

$H_0 : \lambda = \lambda_0 = 6$ against $H_a : \lambda \neq \lambda_0$.

Data: $\overline{X}_n = 5$, $n = 100$

$\sum x_i = n\bar{x}$
$= 500$

$l_1 = x_i \log \lambda - \lambda + \log (x_i)!$

$l_n = \log \lambda \cdot \sum x_i - n\lambda \cancel{+} \sum_i \log (x_i!) \to \max$

FOC : $\dfrac{\sum x_i}{\lambda} - n = 0$

$\hat{\lambda}_{MLE} = \bar{x} = 5$. $\lambda_0 = 6$

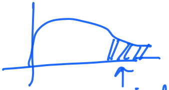
But hypo $H_0 : 6$.

$l_n(\hat{\lambda}_{MLE}) - l_n(\lambda_0) = \sum x_i (\log \hat{\lambda}_{MLE} - \log \lambda_0) - n(\hat{\lambda}_{MLE} - \lambda_0)$

$LR = 2[500 \log 5/6 + 100] \approx 14.6$

q $\chi^2_1 (0.95) = 3.98$. : reject huge no.

Wald test : sq. of T statistic in 1-dim.



- $X_1, \ldots, X_n \sim$ i.i.d. $f(x|\theta)$ where θ is scalar
- $H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$
- $\sqrt{n}(\hat{\theta}_{ML} - \theta) \Rightarrow N(0, I_1^{-1}(\theta))$
- under the null : $W = nI_1(\hat{\theta}_{ML})(\hat{\theta}_{ML} - \theta_0)^2 \Rightarrow \chi_1^2$ under H_0

$t = \sqrt{n} \cdot \sqrt{I_1} \, (\hat{\theta}_{ML} - \theta) \Rightarrow N(0,1)$

$W = (t)^2$

$W \approx LR$ (use same thing: now for $\hat{\theta}$ for Θ_0

requires $\hat{\theta}_{ML}$ & I_1 to calc

Wald test: Multi-dimensional

$X = (X_1, \ldots, X_n) \sim f(x|\theta)$

$g = \begin{bmatrix} g_1(\cdot) \\ \vdots \\ g_p(\cdot) \end{bmatrix}$

- ☐ θ is k-dimensional parameter
- ☐ $H_0 : \{\theta \in \Theta : \; g_1(\theta) = 0, \ldots, g_p(\theta) = 0\}$
- ☐ $\hat{\theta}$ is a consistent estimator of θ (may be ML)such that $\underbrace{\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, S)}$

By delta $\sqrt{n}(g(\hat{\theta}) - g(\theta_0)) \Rightarrow N(0, \Sigma)$

- ☐ $\hat{\Sigma}$ is consistent for $\left(\frac{\partial g(\hat{\theta})}{\partial \theta}\right)' S \frac{\partial g(\hat{\theta})}{\partial \theta}$

$$W = n(g(\hat{\theta}_{ML}) - 0)\hat{\Sigma}^{-1}(g(\hat{\theta}_{ML}) - 0) \Rightarrow \underline{\chi^2_p}$$

Example

$X_1, ..., X_n \sim$ i.i.d. Poisson(λ)
$f(x|\lambda) = \lambda^x e^{-\lambda}/x!$ for $x = 0, 1, 2, ...$
$H_0 : \lambda = \lambda_0 = 6$ against $H_a : \lambda \neq \lambda_0$.
Data: $\overline{X}_n = 5$, $n = 100$

$\hat{\lambda}_{MLE} = \bar{x} = 5$

$l_1 = x \log \lambda - \lambda - \log(x_i!)$

$\frac{\partial l_1}{\partial \lambda} = \frac{x_i}{\lambda} - 1$

1st info eq $E \frac{\partial l_1}{\partial \lambda} = 0$ $\quad E \frac{x_i}{\lambda} - 1 = 0$

$\underline{E x_i = \lambda}$ is exp. of distn

$\frac{\partial^2 l_1}{\partial \lambda \partial \lambda} = -\frac{x_i}{\lambda^2}$ $\quad I_1 = \frac{E x_i}{\lambda^2} = \frac{\lambda}{\lambda^2} = 1/\lambda$

$\text{Wald} = n(\hat{\lambda}_{ML} - \lambda_0)^2 \cdot \underset{\substack{\uparrow \\ \text{at } \hat{\lambda}_{ML}}}{I_1} = 100(5-6)^2 \cdot \frac{1}{5} = \underline{20} \gg 3.98$

reject as well.

Score (LM) Test

Lagrange multiplier

$H_0: \theta = \theta_0$

$X = (x_1, \ldots, x_n) \sim f(x|\theta)$

- the score is $S_n(\theta) = \frac{\partial \ell_n}{\partial \theta}(\theta|X) = \boxed{\sum_{i=1}^{n} \frac{\partial \log f(X_i|\theta)}{\partial \theta}}$
- the first information equality $\mathbb{E}[S(\theta_0)] = 0$
- idea to test $H_0 : \theta = \theta_0$ by testing info equality
- Under the null $\frac{1}{\sqrt{n}} S(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f(X_i|\theta)}{\partial \theta} \Rightarrow \underline{N(0, I_1(\theta_0))}$

$$LM = S(\theta_0)^2 / (n I_1(\theta_0)) \Rightarrow \chi_1^2.$$

- test rejects if $LM > \chi_1^2(1-\alpha)$

CLT it was a step in MLE gaussian

If θ_0 is true, $E\, S(\theta_0) = 0$ $S_n(\theta_0) \approx 0$??

$\frac{1}{\sqrt{n}} S(\theta_0) \cdot \frac{1}{\sqrt{I_1}} \sim N(0,1)$

Why Lagrange multiplier?

$$l(\theta) \to \max_{s.t.\ \theta \in \Theta_0}$$

$$H = l(\theta) - \lambda(\theta - \theta_0) \to \max_\theta$$

FOC: $\underbrace{\frac{\partial l}{\partial \theta}}_{S} = \lambda$ $\qquad$ λ = „Lagrange multiplier"

$$S_n(\theta_0) = S_n(\theta_0) - \underset{\substack{\| \\ 0}}{S_n(\hat{\theta}_{ML})} \overset{\text{Taylor Exp}}{=} \underbrace{\frac{\partial^2 l}{\partial\theta\partial\theta}(\theta^*)}_{n \cdot I_1}(\hat{\theta}_{ML} - \theta_0)$$

$$\frac{S_n(\theta_0)^2}{nI_1} = \frac{1}{nI_1}\left[nI_1(\hat{\theta}_{ML} - \theta_0)\right]^2 = nI_1(\hat{\theta}_{ML} - \theta_0)^2 = W$$

Asymptotically $\quad \underline{LM \approx LR \approx W}$

Score (LM) Test: Multi-dimensional

$X = (X_1, \ldots, X_n) \sim f(\cdot|\theta)$

- ☐ θ is k-dimensional parameter
- ☐ $H_0 : \{\theta \in \Theta : \ g_1(\theta) = 0, ..., g_p(\theta) = 0\}$
- ☐ denote $\hat{\theta}_0 = \arg\max_{\theta \in \Theta_0} \ell(\theta|X)$ to be the restricted estimate
- ☐ under the null $LM = \frac{1}{n} S(\hat{\theta}_0)' I_1(\hat{\theta}_0)^{-1} S(\hat{\theta}_0) \Rightarrow \chi_p^2$
- ☐ LM test of level α rejects the null hypothesis if and only if $LM > \chi_p^2(1-\alpha)$.

For LM need $\begin{cases} - \ \hat{\theta}_0 \text{ (restricted est.} \\ - \ \text{Info.} \end{cases}$

Example

$X_1, \ldots, X_n \sim$ i.i.d. Poisson(λ)
$f(x|\lambda) = \lambda^x e^{-\lambda}/x!$ for $x = 0, 1, 2, \ldots.$
$H_0 : \lambda = \lambda_0 = 6$ against $H_a : \lambda \neq \lambda_0$.
Data: $\overline{X}_n = 5$, $n = 100$

$l_n = \log \lambda \cdot \Sigma x_i - n\lambda \cdot \Sigma \log(x_i!)$

$s_n = \frac{\Sigma x_i}{\lambda} - n \qquad I_1(\lambda) = 1/\lambda$

$LM = \frac{[s_n(\lambda_0)]^2}{n I_1(\lambda_0)} = \frac{[\frac{500}{6} - 100]^2}{100 \cdot \frac{1}{6}} \approx 17 \quad 73.98$

reject!

Summary

θ : k dim. $H_0: g(\theta) = 0$ p dim / res.

under the null hypothesis:

$$LR = 2(\ell(\hat{\theta}_{ML}|X) - \ell(\hat{\theta}_0|X)) \Rightarrow \chi^2_p$$ ← need $\hat{\theta}_{ML}$ & $\hat{\theta}_0$

$$W = n(g(\hat{\theta}_{ML}) - 0)\hat{\Sigma}^{-1}(g(\hat{\theta}_{ML}) - 0) \Rightarrow \chi^2_p$$ ← $\hat{\theta}_{ML}$ & I_1

$$LM = \frac{1}{n}S(\hat{\theta}_0)I_1^{-1}(\hat{\theta}_0)S(\hat{\theta}_0) \Rightarrow \chi^2_p$$ ← $\hat{\theta}_0$ & I_1

where $\hat{\Sigma} = \left(\frac{\partial g}{\partial \theta}(\hat{\theta}_{ML})\right)' I_1^{-1}(\hat{\theta}_{ML}) \left(\frac{\partial g}{\partial \theta}(\hat{\theta}_{ML})\right)$

For all:

- MLE $\ell(\theta) \to \max_{\Theta} \rightsquigarrow \hat{\theta}_{ML}$
- $\hat{\theta}$ restricted : $\ell(\theta) \to \max_{H_0} \rightsquigarrow \hat{\theta}_0$
- $I_1(\theta) = Var(S_1) = -E\frac{\partial^2 \ell}{\partial \theta^2}$

$S_n = \frac{\partial \ell_n}{\partial \theta}$

→ In large samples & "good models" $W \approx LR \approx LM$

① ML. theory works. ML. fn has clear max & quad.

In model when max is not v. clear - not hold.

② Need restriction to be smooth. If non-linear

Wald.

most sensitive test.

$$\left\{\begin{array}{l} H_0: \mu = 1 \\ H_0: \frac{1}{\mu} = 1 \\ H_0: \mu^5 = 1 \end{array}\right\}$$

Wald st. produce 3 diff. $\sqrt{n}(\hat{\mu} - \mu_0) \Rightarrow N(0,V)$

$\leftarrow \frac{1}{\hat{\mu}} - \frac{1}{\mu_0}$

$\leftarrow \hat{\mu}^5 - \mu_0^5$

LR & LLM do not change, more stable