$LR \longrightarrow \chi^2(1)$

Wald $\longrightarrow \chi^2(1)$ as well.

e.g. $X_i \overset{iid}{\sim} N(\theta, \sigma^2)$

$$W = \hat{I}(\hat{\theta}|X)^{-1}(\hat{\theta}-\theta_0)^2$$

For normal case: $I(\theta|X) = \frac{n}{\sigma^2}$

$$\Rightarrow W = \frac{n(\bar{X}-\theta_0)^2}{\sigma^2} = \left(\frac{(\bar{X}-\theta_0)}{\sigma/\sqrt{n}}\right)^2$$

Lagrange Multiplier (LM) or Score Test

$$H = \ln L(\theta|X)\big|_{\theta=\theta_0} + \underbrace{h}_{\text{Lagrange multiplier}}(\theta-\theta_0)$$

h - cost of imposing the restriction

h large $\Rightarrow$ reject the null

$\frac{\partial H}{\partial h}$: $\theta = \theta_0$

$$\frac{\partial H}{\partial \theta_0}: \frac{\partial \ln L(\theta_0|X)}{\partial \theta_0} - h = \hat{S}(\theta_0|X) - h = 0$$

$$\Rightarrow h = \hat{S}(\theta_0|X)$$

$$\Rightarrow \frac{h}{\sqrt{n}} = \frac{\hat{S}(\theta_0|X)}{\sqrt{n}}$$

expand around $\hat{\theta}_{MLE}$

$$= n^{-1/2}\frac{\partial^2}{\partial\theta^2}\log L(\theta|X)\Big|_{\tilde{\theta}}(\theta_0-\hat{\theta}_{MLE}) + n^{-1/2}\underbrace{\hat{S}(\hat{\theta}_{MLE}|X)}_{=0}$$

$\tilde{\theta} \in conv(\theta_0, \hat{\theta}_{MLE})$

Thus,

$$n^{-1/2}\,\hat{s}(\theta_0|\underline{X}) = n^{-1/2}\frac{\partial^2}{\partial\theta^2}\log L(\theta|\underline{X})\Big|_{\tilde{\theta}}(\theta_0-\hat{\theta})$$

$$= -\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\log L(\theta|\underline{X})\Big|_{\tilde{\theta}}\;\underbrace{\sqrt{n}\,(\hat{\theta}_{MLE}-\theta_0)}$$

$$\xrightarrow{p} I(\theta_0|\underline{X}_1) \qquad \xrightarrow{d} N(0, v(\theta))$$

where $v(\theta) = I(\theta|\underline{X}_1)^{-1}$

$$\Rightarrow \frac{\hat{s}(\theta_0|\underline{X})}{\sqrt{n}} \xrightarrow{d} N(0, I(\theta_0|\underline{X}_1))$$

$$\Rightarrow \hat{I}(\theta_0|\underline{X}_1)^{-1/2}\frac{\hat{s}(\theta_0|\underline{X})}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

$$\Rightarrow LM = \hat{s}(\theta_0|\underline{X})^2\frac{1}{n\hat{I}(\theta_0|\underline{X}_1)} = \frac{\hat{s}(\theta_0|\underline{X})^2}{\hat{I}(\theta_0|\underline{X})} \xrightarrow{d} \chi^2(1)$$

---

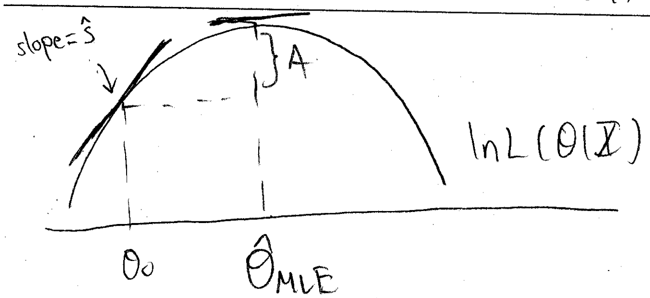
eg. $X_1,\dots,X_n \overset{iid}{\sim} N(\theta,\sigma^2)$

Scaled score $\quad n^{-1/2}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\theta_0)\right)$

$$= n^{-1/2}\left(\frac{n}{\sigma^2}(\bar{X}-\theta_0)\right)$$

$$\hat{I}(\theta|\underline{X}) = \frac{n}{\sigma^2}$$

$$LM = \frac{\hat{s}(\theta_0|\underline{X})^2}{\hat{I}(\theta_0|\underline{X})} = \frac{\left(\frac{n}{\sigma^2}(\bar{X}-\theta_0)\right)^2}{\frac{n}{\sigma^2}} = \left(\frac{\sqrt{n}\,(\bar{X}-\theta_0)}{\sigma}\right)^2 \sim \chi^2(1)$$



LM takes $\hat{s}$ and the curvature $\hat{I}$ to approximate A.

Under normality, $W > LR > LM$.

Decision Theory:

Θ parameter space (states of nature)

$X \overset{pdf}{\sim} f(x|\theta)$

actions: $a \in A$

Loss fcn $L(\theta, a): \Theta \times A \to \mathbb{R}$

eg. $A = \mathbb{R}$

$L(\theta, a) = (\theta - a)^2$

eg (hypothesis testing)

$A = \{0, 1\}$

choose H_0 (0), choose H_1 (1)

	$a = 0$	$a = 1$
H_0	0	1
H_1	d	0

Loss function

Simplify: $\Theta = \{\theta_1, \dots, \theta_m\}$

$A = \{a_1, \dots, a_k\}$

mixed action (convexifies the action set)

- Let $q_1, \dots, q_k$ be s.t. $\sum_{i=1}^{k} q_i = 1$, $q_j \geq 0$. Then $\{\sum_{i=1}^{n} a_i q_i\}$ is a convex set.

$\delta(\underline{X})$: decision rule

Risk: $R(\theta,\delta) = E[L(\theta,\delta(\underline{x}))]$

$= E\left[\sum_{i=1}^{k} \delta_i(\underline{X}) L(\theta, a_i)\right]$

Example 1: $\delta(\underline{X}) = \overline{X}$

Then $R(\theta,\overline{X}) = E[(\overline{X}-\theta)^2] = \frac{Var(X)}{n}$

You want to choose δ to minimize risk.

Defn: a decision rule δ is inadmissible if there exists another decision rule δ' such that

$R(\theta,\delta) \geq R(\theta,\delta') \quad \forall \theta \in \Theta$

and $R(\theta,\delta) > R(\theta,\delta')$ some $\theta \in \Theta$.

If a decision rule is not inadmissible, we say it is admissible.

$H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$

Neyman-Pearson: $\frac{f(x|\theta_0)}{f(x|\theta_1)} < k$

This is admissible decision rule.

Bayesian decision rules.

Prior: $\pi(\theta)$

- in discrete case: $\{\pi_1, \dots, \pi_m\}$

Defn: $r(\pi, \delta) = \int R(\theta, \delta)\pi(\theta)d\theta$ is referred to as Bayes' risk.

Defn: a Bayes decision rule is a rule δ^* satisfying

$$r(\pi, \delta^*) = \inf_{\delta \in D} r(\pi, \delta).$$

$$r(\pi, \delta) = \sum_{i=1}^{m} \pi_i R(\theta_i, \delta)$$

$$= \sum_{i=1}^{m} \pi_i \int L(\theta_i, \delta(x)) f_X(x|\theta)dx$$

$$= \int \underbrace{\left[\sum_{i=1}^{m} L(\theta_i, \delta(x)) f_X(x|\theta)\pi_i\right]}_{\text{inside here, } x \text{ is fixed}} dx$$

Bayes decision rule solves problem given $X = x$.

$$\frac{f_X(x|\theta)\pi_i}{m(x)} = \text{posterior}$$

- find δ that minimizes risk given a posterior

e.g. 1
- $X_i \overset{iid}{\sim} N(\theta, n) \Rightarrow \bar{X} \sim N(\theta, 1)$
- $\pi(\theta) \overset{d}{=} N(q, k^{-1})$
- $\delta^*(x) = \dfrac{\bar{X} + kq}{1+k} \to \bar{X}$ as $k \to 0$ (uniform prior)

e.g.2 prior: $\underbrace{\pi}_{\text{prior that } \theta=\theta_0}$, $1-\pi$

$H_0: \theta = \theta_0$
$H_1: \theta = \theta_1$

$$\pi(\theta_i|x) = \begin{cases} \dfrac{f(x|\theta_0)\pi}{m(x)} & \theta=\theta_0 \\[2ex] \dfrac{f(x|\theta_1)(1-\pi)}{m(x)} & \theta=\theta_1 \end{cases}$$

Risk: $r(\pi, 0) = 0 \cdot \dfrac{f(x|\theta_0)\pi}{m(x)} + d\,\dfrac{f(x|\theta_1)(1-\pi)}{m(x)}$

$r(\pi, 1) = 1 \cdot \dfrac{f(x|\theta_0)\pi}{m(x)} + 0\,\dfrac{f(x|\theta_1)(1-\pi)}{m(x)}$

- will choose $a=1$ if $\dfrac{\pi f(x|\theta_0)}{(1-\pi) f(x|\theta_1)} < d$

iff $\dfrac{f(x|\theta_0)}{f(x|\theta_1)} = \lambda(x) < \underbrace{d\,\dfrac{(1-\pi)}{\pi}}_{k}$

- k is chosen here based on the prior and cost of being wrong.
  - cf classical way of choosing k s.t. $\alpha = \bar{\alpha}$.

Thm: (admissibility of Bayes rules) If a Bayes rule $\delta^*(x)$ exists wrt prior $\pi(\theta)$ with positive weight everywhere, then $\delta^*(x)$ is admissible.

Thm (Complete class theorem): If $\delta(x)$ is admissible, then $\delta(x)$ is a Bayes decision rule wrt some $\pi(\theta)$. (There are assumptions necessary to generate this.)

If π is a flat prior, then $R(\theta, \bar{X}) = \frac{\sigma^2}{n}$ might be a sensible choice. (Graham doesn't agree with this.)

Everything we've done so far has been parametric.

Extra OH on Monday 3:30-5:00