Utilitarian Use of Models

Theory

$$f(x;\theta) \quad \Rightarrow \quad \mathbb{P}(X \in A) = \int_{x\in A} f(x;\theta)dx = ???$$

Practice

$$\widehat{\mathbb{P}}(X \in A) \equiv \int_{x\in A} f(x;\widehat{\theta})dx \qquad \text{if frequentist}$$

$$\begin{aligned}\mathbb{P}(X \in A|X_1,\ldots,X_n) &= \int \mathbb{P}(X \in A|\theta)\pi(\theta|x_1,\ldots,x_n)d\theta \\ &= \int \left[\int_{x\in A} f(x|\theta)dx\right]\pi(\theta|x_1,\ldots,x_n)d\theta \qquad \text{if Bayesian}\end{aligned}$$

$\boxed{\text{Need to say something about } \theta \text{ but do not care about it directly.}}$

© 2017-23 by M. Zhu, PhD

Uncertainty Quantification

If θ means something, may want a probability statement like

$$\mathbb{P}(1.913 < \theta < 2.004) = 95\%.$$

Bayesian Straight-forward (since θ is random)!

$$\mathbb{P}(c_\alpha^{\text{lo}} < \theta < c_\alpha^{\text{hi}} | X_1, \ldots, X_n) = 1 - \alpha$$

e.g., find them by solving

$$\int_{-\infty}^{c_\alpha^{\text{lo}}} \pi(\theta | x_1, \ldots, x_n) d\theta = \int_{c_\alpha^{\text{hi}}}^{\infty} \pi(\theta | x_1, \ldots, x_n) d\theta = \frac{\alpha}{2}.$$

Frequentist Puzzling (since θ is non-random)!

© 2017-23 by M. Zhu, PhD

A Simple Example

[Example 7.2 (p. 118)]

$$\boxed{X_1, X_2, ..., X_n \overset{iid}{\sim} \mathrm{N}(\mu, \sigma^2),\ \sigma^2\ \underline{\text{known}}}$$

Starting from the sampling distribution (of the estimator $\widehat{\mu} = \bar{X}$), $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, we can deduce

$$\bar{X} - \mu \sim \mathrm{N}(0, \sigma^2/n) \Rightarrow \mathbb{P}\left(\frac{-1.96\sigma}{\sqrt{n}} < \bar{X} - \mu < \frac{1.96\sigma}{\sqrt{n}}\right) = 95\%$$

$$\Rightarrow \mathbb{P}\left(\bar{X} - \frac{1.96\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right) = 95\%.$$

Lesson The last (rearranged) probability statement (about $\bar{X}$, <u>NOT</u> about μ) quantifies our uncertainty about the unknown parameter μ, leading naturally to a notion of "margin of error".

© 2017-23 by M. Zhu, PhD

Ideal Scenario

Know the distribution of $\widehat{\theta} - \theta \Rightarrow$ can then say (for <u>scalar</u> θ)

$$\mathbb{P}\left(c_{\alpha}^{\text{lo}} < \widehat{\theta} - \theta < c_{\alpha}^{\text{hi}}\right) = 1 - \alpha$$

$$\Rightarrow \mathbb{P}\left(\widehat{\theta} - c_{\alpha}^{\text{hi}} < \theta < \widehat{\theta} - c_{\alpha}^{\text{lo}}\right) = 1 - \alpha\%.$$

The set,

$$\left\{\theta : c_{\alpha}^{\text{lo}} < \widehat{\theta} - \theta < c_{\alpha}^{\text{hi}}\right\} \quad = \quad \left(\widehat{\theta} - c_{\alpha}^{\text{hi}},\ \widehat{\theta} - c_{\alpha}^{\text{lo}}\right),$$

is called a $(1 - \alpha) \times 100\%$-level confidence set for θ.

Difficulty Distribution of $\widehat{\theta} - \theta$ not exactly easy to derive!

© 2017-23 by M. Zhu, PhD

A Not-So-Simple Example

[Example 7.3 (pp. 120–122)]

$$\boxed{X_1, X_2, ..., X_n \overset{iid}{\sim} \mathrm{N}(\mu, \sigma^2),\ \sigma^2\ \underline{\text{unknown}}}$$

Previous logic

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1) \Rightarrow \cdots \Rightarrow \left(\bar{X} - \frac{1.96\sigma}{\sqrt{n}}, \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right)$$

still applies, but answer not usable.

Solution Was able to derive the distribution of

$$T \equiv \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)} \quad \text{where} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

a celebrated early achievement of statistics!

© 2017-23 by M. Zhu, PhD

Student's t-Distribution

Nontrivial S^2 also a random variable $\Rightarrow T$ a much more complicated (non-linear) function of $X_1, \ldots, X_n$ than just $\bar{X}$.

A "Big Deal" Can now do

$$\mathbb{P}(c_\alpha^{\text{lo}} < T < c_\alpha^{\text{hi}}) = 1 - \alpha \Rightarrow \mathbb{P}\left(c_\alpha^{\text{lo}} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < c_\alpha^{\text{hi}}\right) = 1 - \alpha$$

$$\Rightarrow \mathbb{P}\left(\bar{X} - c_\alpha^{\text{hi}} \frac{S}{\sqrt{n}} < \mu < \bar{X} - c_\alpha^{\text{lo}} \frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Facts (i) $c_\alpha^{\text{lo}} = -c_\alpha^{\text{hi}}$; (ii) for <u>small</u> n=2, 3, ..., $c_{0.05}^{\text{hi}} \approx 12.7, 3.2$, ... can be "large"; (iii) for <u>moderate</u> n=20, 30, ..., $c_{0.05}^{\text{hi}} \approx 2 > 1.96$.

© 2017-23 by M. Zhu, PhD

Pivotal Quantity

In both cases, relied on a certain quantity, say, $h(\theta; \boldsymbol{D})$,

(i) which involved only the parameter θ of interest (but no other unknown parameter) and data $\boldsymbol{D} \equiv \{X_1, \ldots, X_n\}$, and

(ii) whose distribution was completely known.

For example,

$$h(\mu; \boldsymbol{D}) = \bar{X} - \mu \sim \mathrm{N}(0, \sigma^2/n), \quad \text{[when } \sigma^2 \text{ known]}$$

$$h(\mu; \boldsymbol{D}) = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

© 2017-23 by M. Zhu, PhD

Pivotal Quantity

Property (ii) allowes us to identify a set C_α—e.g., $(c_\alpha^{\text{lo}}, c_\alpha^{\text{hi}})$—such that

$$\mathbb{P}[h(\theta; \boldsymbol{D}) \in C_\alpha] = 1 - \alpha.$$

Property (i) allowes us to "rearrange" or "solve" the relationship

$$h(\theta; \boldsymbol{D}) \in C_\alpha \Rightarrow \theta \in h^{-1}(C_\alpha; \boldsymbol{D}) \equiv \{\theta : h(\theta; \boldsymbol{D}) \in C_\alpha\}$$

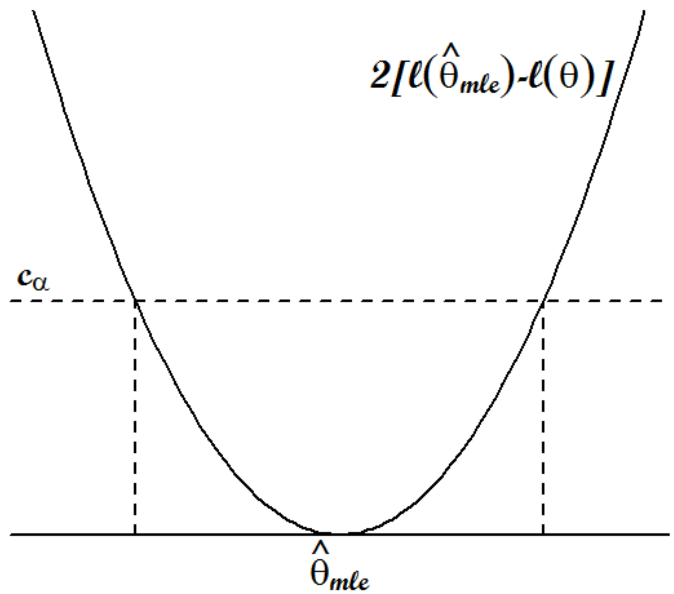
for θ.

Remark Property (ii) not easy!

© 2017-23 by M. Zhu, PhD

Likelihood Ratio

$$h(\theta;\boldsymbol{D}) = 2\log\left[\frac{L(\widehat{\theta}_{mle};\boldsymbol{D})}{L(\theta;\boldsymbol{D})}\right] = 2[\ell(\widehat{\theta}_{mle};\overset{\text{will omit}}{\overset{\downarrow}{\boldsymbol{D}}}) - \ell(\theta;\overset{\text{sometimes}}{\overset{\downarrow}{\boldsymbol{D}}})]$$



© 2017-23 by M. Zhu, PhD

The Simple Example

[Example 7.4 (pp. 124–125)]

$$\boxed{X_1, X_2, ..., X_n \overset{iid}{\sim} \mathrm{N}(\mu, \sigma^2),\ \sigma^2\ \underline{\text{known}}}$$

$$\frac{L(\widehat{\mu}_{mle})}{L(\mu)} = \frac{L(\bar{X})}{L(\mu)} = \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(X_i - \bar{X})^2}{2\sigma^2}\right]}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(X_i - \mu)^2}{2\sigma^2}\right]} = ...$$

$$... = \exp\left[\frac{n\bar{X}^2 - 2\mu(n\bar{X}) + n\mu^2}{2\sigma^2}\right]$$

$$2\log\left[\frac{L(\widehat{\mu}_{mle})}{L(\mu)}\right] = \frac{n(\bar{X} - \mu)^2}{\sigma^2} = \left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right]^2 \sim \chi^2_{(1)}$$

© 2017-23 by M. Zhu, PhD

Likelihood Ratio (Cont'd)

In general, may have "other parameters" (those we are NOT making inference about, aka nuisance parameters, say, ψ), so extend likelihood ratio idea to

$$\Lambda(\theta; \boldsymbol{D}) \equiv \frac{\sup\limits_{\theta,\psi} L(\theta,\psi)}{\sup\limits_{\psi} L(\theta,\psi)} = \frac{L(\widehat{\theta}_{mle}, \widehat{\psi}_{mle})}{L(\theta, \widehat{\psi}_{mle|\theta})},$$

where $\widehat{\psi}_{mle|\theta}$ is the constrained (or partial) MLE of ψ while keeping θ fixed.

Subtlety Often, $\widehat{\psi}_{mle|\theta} \neq \widehat{\psi}_{mle}$.

© 2017-23 by M. Zhu, PhD

The Not-So-Simple Example

[Example 7.5 (pp. 125–126)]

$$\boxed{X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathrm{N}(\mu, \sigma^2),\ \sigma^2\ \underline{\text{unknown}}}$$

$$\widehat{\mu}_{mle} = \bar{X}, \quad \underbrace{\widehat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}_{\widehat{\sigma}^2}, \quad \underbrace{\widehat{\sigma}^2_{mle|\mu} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2}_{\widehat{\sigma}^2_{\mu}}$$

$$\Lambda(\mu; \boldsymbol{D}) = \frac{L(\widehat{\mu}_{mle}, \widehat{\sigma}^2_{mle})}{L(\mu, \widehat{\sigma}^2_{mle|\mu})} = \frac{L(\bar{X}, \widehat{\sigma}^2)}{L(\mu, \widehat{\sigma}^2_{\mu})}$$

$$= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\widehat{\sigma}^2}} \exp\left[-\frac{(X_i - \bar{X})^2}{2\widehat{\sigma}^2}\right]}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\widehat{\sigma}^2_{\mu}}} \exp\left[-\frac{(X_i - \mu)^2}{2\widehat{\sigma}^2_{\mu}}\right]} = \ldots$$

© 2017-23 by M. Zhu, PhD

The Not-So-Simple Example

[Example 7.5 (pp. 125–126)]

$$\cdots = \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\widehat{\sigma}^2}} \exp\left[-\frac{(X_i - \bar{X})^2}{2\widehat{\sigma}^2}\right]}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\widehat{\sigma}_\mu^2}} \exp\left[-\frac{(X_i - \mu)^2}{2\widehat{\sigma}_\mu^2}\right]}$$

$$= \left[\frac{\widehat{\sigma}_\mu^2}{\widehat{\sigma}^2}\right]^{n/2} \times \frac{\exp\left[-\frac{\sum(X_i - \bar{X})^2}{2\widehat{\sigma}^2}\right]}{\exp\left[-\frac{\sum(X_i - \mu)^2}{2\widehat{\sigma}_\mu^2}\right]}$$

$$2\log\Lambda(\mu; \boldsymbol{D}) = n\log\left[\frac{\sum(X_i - \mu)^2}{\sum(X_i - \bar{X})^2}\right] \sim ???$$

© 2017-23 by M. Zhu, PhD

Asymptotic Result

Theorem Let

$$n = \text{sample size}, \quad \Lambda_n(\theta; \boldsymbol{D}) = \frac{L(\widehat{\theta}_{mle}, \widehat{\psi}_{mle})}{L(\theta, \widehat{\psi}_{mle|\theta})},$$

θ is the true but unknown parameter of interest, and all MLEs are based on n samples. Then, under <u>some regularity conditions</u>, the distribution of

$$2 \log \Lambda_n(\theta; \boldsymbol{D})$$

converges to that of $\chi^2_{(df)}$ as $n \to \infty$, where $df = \dim(\theta)$.

© 2017-23 by M. Zhu, PhD

Comparison

[Example 7.6 + Exercise 7.2 (p. 128)]

$$\{\mu:\ 2\log\Lambda(\mu;\boldsymbol{D}) < c_\alpha\} \ \Leftrightarrow\ \left\{\mu:\ n\log\left[\frac{\sum(X_i-\mu)^2}{\sum(X_i-\bar{X})^2}\right] < 3.84\right\}$$

versus

$$\{\mu:\ c_\alpha^{\text{lo}} < T(\mu;\boldsymbol{D}) < c_\alpha^{\text{hi}}\} \ \Leftrightarrow\ \left\{\mu:\ -2 < \frac{\bar{X}-\mu}{S/\sqrt{n}} < 2\right\}$$

<u>Remark</u>

Appearances are deceiving! Try Exercise 7.2 (p. 128).

© 2017-23 by M. Zhu, PhD