STAT 5430

Lec 33, F, Apr 18

(No new homework)

Interval Estimation I

Pivotal Quantities

↙ vector of parameters $\underset{\sim}{\theta}$

Definition: Let $X_1, \ldots, X_n$ be joint pdf/pmf $f(\underset{\sim}{x}|\underset{\sim}{\theta})$, $\underset{\sim}{\theta} \in \Theta \subset \mathbb{R}^p$. Then a random variable $Q(\underset{\sim}{X}, \underset{\sim}{\theta})$ is called a **pivot** or **pivotal quantity** if the distribution of $Q(\underset{\sim}{X}, \underset{\sim}{\theta})$ under $\underset{\sim}{\theta}$ *does not depend* on $\underset{\sim}{\theta}$.

Note: $Q(\underset{\sim}{X}, \underset{\sim}{\theta})$ is NOT a statistic (because can depend on $\underset{\sim}{\theta}$)

$$P_{\underset{\sim}{\theta}}(Q(\underset{\sim}{X}, \underset{\sim}{\theta}) \in A) = P(Q(\underset{\sim}{X}, \underset{\sim}{\theta}) \in A)$$

Some examples: (pivots, unlike statistics, can be functions of parameters $\underset{\sim}{\theta}$)

$\dfrac{X_1 - X_2}{X_3 - X_4}$ ancillary statistic

1. Let $X_1 \ldots X_n$ be iid $N(\mu, \sigma^2)$ random variables.

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\text{pivot}\left\{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)\right. \qquad \text{pivot}\left\{\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim T_{n-1} \text{ distribution}\right.$$

2. Let f_0 be a pdf on $\mathbb{R}$. Let $X_1 \ldots X_n$ be iid with random variables common pdf $f(x|\underset{\sim}{\theta})$ where

$$\underset{\sim}{\theta} = (\theta_1, \theta_2) \qquad f(x|\underset{\sim}{\theta}) = \frac{1}{\theta_2} f_0\left(\frac{x - \theta_1}{\theta_2}\right), \quad x \in \mathbb{R},$$

for $\underset{\sim}{\theta} = (\theta_1, \theta_2)$, $\theta_1 \in \mathbb{R}$ (location parameter) and $\theta_2 > 0$ (scale parameter).

Then, $Q(\underset{\sim}{X}, \underset{\sim}{\theta}) = \dfrac{\bar{X}_n - \theta_1}{\theta_2}$ ← pivot

Why? Note: $Y_i \equiv \dfrac{X_i - \theta_1}{\theta_2} \sim f_0(y)$ (i.e. $Y_1, \ldots, Y_n$ iid $f_0(y)$)

and $Q(\underset{\sim}{X}, \underset{\sim}{\theta}) = \frac{1}{n}\sum_{i=1}^{n} Y_i$ ← distribution doesn't depend on $\underset{\sim}{\theta}$

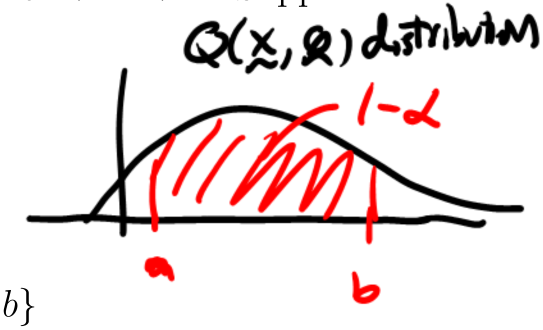
Interval Estimation I

Interval Estimation via Pivotal Quantities

Remarks:

.

1. Let $Q(\underset{\sim}{X}, \underset{\sim}{\theta})$ be a pivotal quantity ($\underset{\sim}{\theta} \in \Theta \subset \mathbb{R}^p$) and $0 < \alpha < 1$. Suppose $-\infty \leq a \leq b \leq \infty$ are such that



$$P(a \leq Q(\underset{\sim}{X}, \underset{\sim}{\theta}) \leq b) = P_{\underset{\sim}{\theta}}(a \leq Q(\underset{\sim}{X}, \underset{\sim}{\theta}) \leq b) = 1 - \alpha.$$

Then,

given data $\underset{\sim}{X}$ →

$$C_{\underset{\sim}{X}} = \{\underset{\sim}{\theta} : \underset{\sim}{\theta} \in \Theta,\ a \leq Q(\underset{\sim}{X}, \underset{\sim}{\theta}) \leq b\}$$

is a **confidence region** for $\underset{\sim}{\theta}$ with CC $(1 - \alpha)$

$$\text{That is, } \min_{\underset{\sim}{\theta} \in \Theta} P_{\underset{\sim}{\theta}}(\underset{\sim}{\theta} \in C_{\underset{\sim}{X}}) = \min_{\underset{\sim}{\theta} \in \Theta} \underbrace{P_{\underset{\sim}{\theta}}(a \leq Q(\underset{\sim}{X}, \underset{\sim}{\theta}) \leq b)}_{\underbrace{P(a \leq Q(\underset{\sim}{X}, \underset{\sim}{\theta}) \leq b)}_{1-\alpha}} = 1 - \alpha.$$

2. If $\Theta \subset \mathbb{R}$ and $Q(\underset{\sim}{X}, \underset{\sim}{\theta})$ is monotone in $\underset{\sim}{\theta} \in \mathbb{R}$, then the region $C_{\underset{\sim}{X}}$ will be an interval.

Interval Estimation I

Interval Estimation via Pivotal Quantities

shape is known

Example: Let $X_1 \dots X_n$ be iid Gamma(δ_0, θ) where $\theta > 0$ (δ_0 fixed/known). Using a pivotal quantity based on $\sum_{i=1}^n X_i$, find a CI for θ with C.C. $1-\alpha$.

scale (unknown)

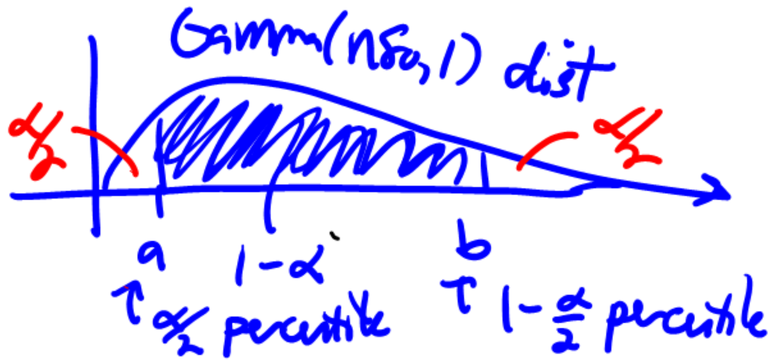
Solution: $\sum_{i=1}^{n} X_i \sim \text{Gamma}(n\delta_0, \theta)$

$$Q(\underset{\sim}{X}, \theta) = \sum_{i=1}^{n} X_i \Big/ \theta \sim \text{Gamma}(n\delta_0, 1)$$

pivot

dist. doesn't depend on θ

Find a & b such that



Confidence interval for $\theta > 0$ is $\{\theta : a \le Q(\underset{\sim}{X}, \theta) \le b\} = \{\theta : a \le \frac{\sum_{i=1}^{n} X_i}{\theta} \le b\}$

$$= \{\theta > 0 : \frac{\sum X_i}{b} \le \theta \le \frac{\sum X_i}{a}\}$$

$$= \left[\frac{\sum X_i}{b}, \frac{\sum X_i}{a}\right]$$

Interval Estimation I

Asymptotically Pivotal Quantities

Definition: A sequence of random variables $Q_n \equiv Q_n(X_1, \ldots, X_n, \underset{\sim}{\theta})$, $n \geq 1$, is called **asymptotically pivotal** or an **asymptotical pivot** if there exists a continuous random variable Q such that

$$Q_n \xrightarrow{d} Q \quad \text{as } n \to \infty$$

and Q has a distribution *not* depending on $\underset{\sim}{\theta}$.

e.g. $X_1, \ldots, X_n$ iid $EX_1 = \mu$ & $Var(X_1) = \sigma^2$

$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, Var(X_1) = \sigma^2)$ as $n \to \infty$ by CLT

Remark: $Q(X_1, \ldots, X_n, \mu, \sigma^2) \equiv \sqrt{n}\dfrac{(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \frac{1}{\sigma} N(0, \sigma^2) = N(0, \underbrace{\tfrac{1}{\sigma^2} \cdot \sigma^2}_{1})$

Suppose $Q_n(X_1, \ldots, X_n, \underset{\sim}{\theta})$ is asymptotically pivotal quantity and, for some $-\infty \leq a \leq b \leq \infty$, $P(a \leq Q \leq b) = 1 - \alpha$ holds for some $0 < \alpha < 1$. Then, for large n,

limit distribution
eg $Q \equiv N(0,1)$

$$C_{\underset{\sim}{X}} = \left\{ \underset{\sim}{\theta} : \underset{\sim}{\theta} \in \Theta, \ a \leq Q_n(X_1, \ldots, X_n, \underset{\sim}{\theta}) \leq b \right\}$$

is a **confidence region** for $\underset{\sim}{\theta}$ with **approximate** CC $(1 - \alpha)$

The main idea is that, for large n,

$$\begin{aligned}
\min_{\underset{\sim}{\theta} \in \Theta} P_{\underset{\sim}{\theta}} \left(\underset{\sim}{\theta} \in C_{\underset{\sim}{X}} \right) &= \min_{\underset{\sim}{\theta} \in \Theta} P_{\underset{\sim}{\theta}} \left(a \leq Q_n(X_1, \ldots, X_n, \underset{\sim}{\theta}) \leq b \right) \\
&\approx P(a \leq Q \leq b) \\
&= 1 - \alpha.
\end{aligned}$$

$$L(\theta) = \prod_{i=1}^{n} \binom{m}{x_i} \theta^{x_i} (1-\theta)^{m-x_i}$$

Interval Estimation I

Asymptotically Pivotal Quantities: Illustration

("large n + CLT + variance estimation = asymptotic pivot")

Example 1: Let $X_1, \ldots, X_n$ be iid Binomial(m, θ), $0 < \theta < 1$, with fixed m. Find an asymptotic pivot using the MLE $\hat{\theta}_n$ and find a corresponding CI for θ with approximate C.C. $1 - \alpha$.

Solution: $\hat{\theta}_n = \frac{\bar{X}_n}{m} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ where $Y_i = \frac{X_i}{m}$ $\left(E_\theta Y_1 = \theta,\ Var_\theta(Y_1) = \frac{(1-\theta)\theta}{m}\right)$

By CLT,

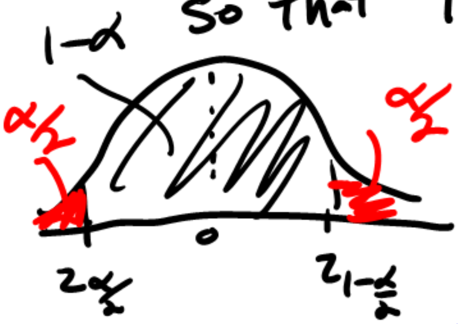
$$\sqrt{n}\,(\underbrace{\bar{Y}_n}_{\hat{\theta}_n} - \underbrace{\theta}_{E_\theta Y_1}) \xrightarrow{d} N\left(0, Var_\theta(Y_1) = \frac{(1-\theta)\theta}{m}\right) \text{ as } n \to \infty.$$

Letting $g(\underbrace{\hat{\theta}_n}_{\bar{Y}_n}) = \sqrt{\frac{(1-\hat{\theta}_n)\hat{\theta}_n}{m}} \xrightarrow{P} g(\underbrace{\theta}_{E_\theta Y_1})$ as $n \to \infty$ by WLLN ($\bar{Y}_n \xrightarrow{P} E_\theta Y_1$) & continuous mapping theorem.

Hence,

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{g(\hat{\theta}_n)} \equiv Q_n(X_1, \ldots, X_n, \theta) \xrightarrow{d} Q \equiv \frac{1}{\underbrace{\sqrt{\frac{\theta(1-\theta)}{m}}}_{g(\theta)}} \cdot N\left(0, \frac{\theta(1-\theta)}{m}\right) = N\left(0, \frac{1}{\frac{\theta(1-\theta)}{m}} \frac{\theta(1-\theta)}{m}\right) = N(0,1)$$

Now use $Q \equiv Z \sim N(0,1)$ & find a & b so that $P(\underbrace{a}_{z_{\alpha/2}} \le Z \le \underbrace{b}_{z_{1-\frac{\alpha}{2}}}) = 1 - \alpha$.



$$C_{\underline{X}} = \left\{0 < \theta < 1 : z_{\alpha/2} \le \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{g(\hat{\theta}_n)} \le z_{1-\alpha/2}\right\}$$

$$= \left[\hat{\theta}_n - z_{1-\alpha/2}\frac{g(\hat{\theta}_n)}{\sqrt{n}},\ \hat{\theta}_n + z_{\alpha/2}\frac{g(\hat{\theta}_n)}{\sqrt{n}}\right]$$

is a CI for θ with approximate C.C. $1-\alpha$.

Interval Estimation I

Asymptotically Pivotal Quantities: Illustration

("large n + LRS + χ^2-approximation = asymptotic pivot")

Example 2: Let $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ be independent random variables where $X_1, \ldots, X_n$ are iid Exponential(θ), $\theta > 0$, and $Y_1, \ldots, Y_m$ are iid Exponential(λ), $\lambda > 0$. Find a large-sample confidence region for (θ, λ), with approximate C.C. $1 - \alpha$, based on a likelihood ratio statistic. (LRS)