Lecture 13

Confidence intervals

Relate true value with value of estimate.

$f(D)$ true value

$X = cf(D')$ estimate

Want to make a statement that tells where $f(D)$ is w.r.t. X

An interval $[A, B]$ is called $(1-\delta)$ conf interval if $f(D) \in (A, B)$ with probability at least $1-\delta$

δ - small $\Rightarrow$ good confidence.

$\delta = 0.05 \Rightarrow 95\%$ conf interval.

$X \rightarrow [X_1, X_2] \rightarrow \delta$

$$P[f(D) \in [X_1, X_2]] \geq 1-\delta$$

$$= P[f(D) \notin [X_1, X_2]] < \delta$$

$$- P[f(D) < X_1 \lor f(D) > X_2] < \delta$$

$$P[f(D) > X_1, f(D) < X_2] > 1-\delta$$

Useful Inequalities

1° Markov Inequality. For given r.v. X

$$\forall \lambda > 0, \; P\left[|X| \geq \lambda \right] \leq \frac{E[|X|]}{\lambda}$$

$X \Rightarrow E(X) = 1$ Take $X > 0$

$\lambda \longrightarrow 100$

$$P\left[|X| \geq 100 \right] < \frac{1}{100}$$

2° Chebyshev Inequality:

For X s.t. $Var(X) = \sigma^2 < \infty$

$$P\left[|X - E(X)| \geq \lambda \right] \leq \frac{Var(X)}{\lambda^2}$$

pick $\lambda = \gamma \cdot \sqrt{Var(X)}$
$= \gamma \cdot Std(X)$

$$P\left[|X - E(X)| \geq \gamma \, Std(X) \right] \leq \frac{Var(X)}{\gamma^2 Var(X)} = \frac{1}{\gamma^2}$$

if we pick $\gamma = 10$

$$P\left[X \text{ is off by } 100\text{ std from } f(\theta) \right] \leq \frac{1}{100}$$

Recipe for conf. interval:

- make X an unbiased estimate $f(\theta) = E(X)$
- compute $\sigma = \sqrt{Var(X)}$

$(X - \gamma\sigma, X + \gamma\sigma)$ with conf $1 - \frac{1}{\gamma^2}$

How can we make CI's better?

- determine and use the distrib of X
  - cumulative distrib. fct of X

$F(x) = P[X \leq x]$

$F^{-1}(x)$

  - symmetric conf. int.

$E(X) = f(\theta)$

Find x_l s.t. $F(x_l) = \frac{\delta}{2}$

x_h s.t. $1 - F(x_h) = \frac{\delta}{2}$

Central limit theorem

For any $X_1, X_2, \ldots X_n$ iid

s.t. $\mathrm{var}(X_i) < \infty$

$= \sigma^2$

$\mu = E[X]$

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \longrightarrow \frac{\frac{\sum_{i=1}^{n} X_i}{n} - \mu}{\sigma} \cdot \sqrt{n}$$

$$\longrightarrow N(0,1)$$

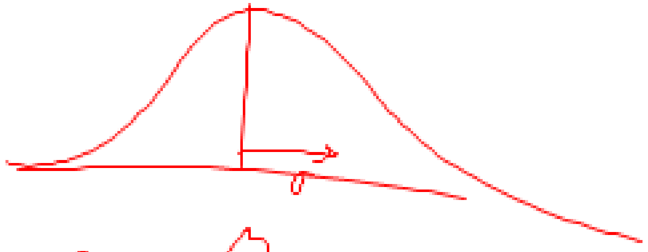
$$\frac{\sum_{i=1}^{n} X_i}{n} \longrightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$$

$n > 100$

Hays & Hellenstein

clt for sampling with replacement.

Normal



Gamma

χ^2-dist

normal