Lecture 10. Confidence Sets and intro to Bayes

Confidence set

Assume sample: $X \sim F_\theta$

$\to C(X)$ - set in Θ

Definition 1.

Coverage probability of the set $C(X) \subset \Theta$ is the probability that confidence set $C(X)$ contains θ,

(not random, parameter)

use 95% for coverage set

$$\text{Coverage}(\theta) = \mathbb{P}_\theta\{\theta \in C(X)\}$$

Definition 2.

We say that confidence set $C(X)$ has confidence level $1 - \alpha$ if

$$\inf_{\theta \in \Theta} \mathbb{P}_\theta\{\theta \in C(X)\} \geq 1 - \alpha$$

Test Inversion

- For each $\theta_0 \in \Theta$, consider $H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$.
- Suppose a test is of size α:

$$\mathbb{P}_{\theta_0}\{\text{the test rejects } \theta = \theta_0\} \leq \alpha$$

- The confidence set θ_0 that did not reject, collect them
  $C(X) = \{\theta_0 \in \Theta :$ the null hypothesis that $\theta = \theta_0$ is not rejected$\}$ is of confidence level $1 - \alpha$.
- This procedure is known as test inversion.

$P_{\theta_0}\{\theta_0 \text{ is rejected}\} \leq \alpha$

$P_{\theta_0}\{\theta_0 \text{ is accepted}\} \geq 1-\alpha$

How test each θ? — grid testing/inversion.
Inversion of t-statistic most common.

Example

$X_1, ..., X_n \sim$ i.i.d. from a distribution with two finite moments
Construct a confidence set for $\mu = EX_i$ of (asymptotic) level $1 - \alpha$.

$H_0 : \mu = \mu_0 \quad$ v/s $\quad H_1 : \mu \neq \mu_0$

$\hat{\mu} = \bar{x}$ saw before $\hat{\mu}$ is good est. of $\bar{x}$

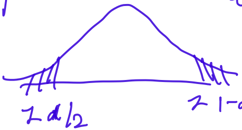
$$\frac{\sqrt{n}(\bar{x} - \mu_0)}{\cancel{\sigma}\ \sqrt{s^2}} \xrightarrow{\text{under } H_0} \mathcal{N}(0,1)$$

to normalize

use sample instead, don't know σ

Calc. from campu:
$\hat{\mu} = \bar{x}$ & s^2

t-std. normal



accept if $z_{\alpha/2} \leq \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} \leq z_{1-\alpha/2}$

What μ_0 will be accepted by this procedure? Which one to accept?

$$\frac{s}{\sqrt{n}} \cdot z_{\alpha/2} \leq \bar{x} - \mu_0 \leq \frac{s}{\sqrt{n}} z_{1-\alpha/2}$$

$$\bar{x} - \frac{s}{\sqrt{n}} z_{1-\alpha/2} \leq \mu_0 \leq \bar{x} - s/\sqrt{n}\, z_{\alpha/2}$$

$$\left[\bar{x} - \frac{s}{\sqrt{n}} z_{1-\alpha/2},\ \bar{x} - \frac{s}{\sqrt{n}} z_{\alpha/2}\right]$$

std. error

≈ 2

Asymptotic Wald Intervals

- $X_1, \ldots, X_n \sim i.i.d. f(x|\theta)$ with $\theta \in \Theta$.
- Confidence interval for $\tau = h(\theta)$?
- Under some regularity $\sqrt{n}(\hat{\theta}_{ML} - \theta) \Rightarrow N(0, I^{-1}(\theta))$
- Delta-method: $\sqrt{n}(h(\hat{\theta}_{ML}) - h(\theta)) \Rightarrow N(0, \Sigma)$, $\Sigma = (h'(\theta))^2 I^{-1}(\theta)$
- $\hat{\Sigma} = (h'(\hat{\theta}_{ML})^2(-\partial^2 l_n(\hat{\theta}_{ML})/\partial\theta^2)^{-1}$ estimates Σ consistently
- Invert Wald test we get

$$\left[\underset{\hat{\tau}}{h(\hat{\theta}_{ML})} - z_{1-\alpha/2}\sqrt{\frac{\hat{\Sigma}}{n}}, h(\hat{\theta}_{ML}) - z_{\alpha/2}\sqrt{\frac{\hat{\Sigma}}{n}}\right]$$

$\hat{\tau} = h(\hat{\theta}_{ML})$ you need to prove:

est : consistent & asymp. g

$\sqrt{n}(\hat{\tau} - \tau) \Rightarrow N(0, \Sigma)$

- est. for Σ, $\hat{\Sigma}$ ← consistent

$\left[\hat{\tau} \pm 1.96\sqrt{\frac{\hat{\Sigma}}{n}}\right]$

Example

$X_1, ..., X_n \sim i.i.d.\text{Bernoulli}(p)$
want to construct a confidence set for $h(p) = p/(1-p)$ $= \tau$

$\hat{p} = \bar{x}$ $\quad \sqrt{n}\,(\hat{p} - p) \Rightarrow N(0, p(1-p))$

$\tau = \frac{p}{1-p}$ $\quad \hat{\tau} = \frac{\hat{p}}{1-\hat{p}} = h(\hat{p})$

$h = \frac{p}{1-p} = -1 + \frac{1}{1-p}$

$h' = \frac{-1}{(1-p)^2}$

Delta . method :

$\sqrt{n}\,(\hat{\tau} - \tau) \Rightarrow N\left(0, p(1-p) \cdot \left[\frac{1}{(1-p)^2}\right]^2\right) = N\left(0, \frac{p}{(1-p)^3}\right)$

$\hat{\Sigma} = \frac{\hat{p}}{(1-\hat{p})^3} \xrightarrow[\text{CMT}]{} \Sigma$ $\quad \Sigma$

$\left[\hat{\tau} - z_{1-\alpha/2} \sqrt{\frac{p}{(1-p)^3}} \cdot \frac{1}{\sqrt{n}} \; ; \; \hat{\tau} - z_{\alpha/2} \sqrt{\frac{\hat{p}}{(1-\hat{p})^3}} \cdot \frac{1}{\sqrt{n}} \right]$

Bootstrap confidence sets

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

Trying to avoid calc Σ but know its gaussian

- ☐ confidence set for a parameter θ based on a consistent estimator $\hat{\theta} = \delta(X)$
- ☐ test $H_0 : \theta = \theta_0$ by bootstrapping $T(\theta_0) = \hat{\theta} - \theta_0$
- ☐ draw a bootstrapped sample X^* from $\hat{F}$ and calculate quantiles of $T^* = \delta(X^*) - \hat{\theta}$

* est always for sample of size n.

repeat B

$X_1^*, \ldots, X_n^* \sim \text{iid } \hat{F}_n$

$\hat{\theta}^* = \delta(X^*)$

$T^* = \hat{\theta}_B^* - \hat{\theta} \overset{d}{\approx} T = \hat{\theta} - \theta_0$

truth for botstrp world.

g have multiple draws: $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$

T_1^* … T_B^*

order

$T_{(1)}^* \le \ldots \le T_{(B)}^*$

accept θ_0 if $T^*_{(\lceil \alpha/2 B \rceil)} \le \hat{\theta} - \theta_0 \le T^*_{(\lceil (1-\alpha/2)B \rceil)}$

Bootstrap confidence sets $\hat{\theta} - T^*([(1-\alpha/2)B]) \le \theta_b \le \hat{\theta} - T^*_{([\alpha/2])}$

- For $b = 1, ..., B$ repeat the following:
  - Draw a random sample X_b^* from distribution $\hat{F}$;
  - Calculate $T_b^* = \delta(X_b^*) - \hat{\theta}$;
- Order from smallest to largest: $T^*_{(1)} \le ... \le T^*_{(B)}$.
- Test of $H_0 : \theta = \theta_0$ accepts if $T^*_{([\frac{\alpha}{2}B])} \le \hat{\theta} - \theta_0 \le T^*_{([(1-\frac{\alpha}{2})B])}$.
- Confidence set is $\hat{\theta} - T^*_{([(1-\frac{\alpha}{2})B])} \le \theta_0 \le \hat{\theta} - T^*_{([\frac{\alpha}{2}B])}$.

Second way: $\sqrt{n}(\hat{\theta} - \theta_0) \approx N(0, \Sigma)$

$$\hat{\Sigma}^* = \frac{1}{B} \sum_B [\sqrt{n}(\hat{\theta}^*_B - \hat{\theta})]^2$$

$$\Sigma = \hat{\Sigma}^*$$

$$\left[\hat{\theta} \pm \frac{\sqrt{\hat{\Sigma}^*}}{\sqrt{n}} 1.96\right]$$

Frequentists Paradigm

- θ is non-random, fixed and unknown
- Data X is random from distribution dependent on θ
- Inferences are random because data is random
- Statistical guarantees are about sampling uncertainties

$\hat{\theta}$ is unbiased $E\hat{\theta} = \theta$ // on average be right

- uncertainty / randomness comes from data / sampling

Bayesian Paradigm

- θ is random with prior distribution $\pi(\cdot)$
- Data X is random from $f(\cdot|\theta)$ Data random befor realised.
- After you observe data you treat is as given (conditional on data) update your belief of prior
- Inferences describe evolution of randomness about θ

θ is random $\pi(\cdot)$

$X \sim f(x|\theta)$

you see data

posterior $\pi(\theta | \underbrace{X = x}_{\text{data that gets realised}})$

Bayesian Paradigm

- Inputs
  - prior distribution $\pi(\cdot)$ - for θ
  - distribution of data $f(\cdot|\theta)$ - x

how likely data distribute this way

- joint pdf of θ and X is $f(x,\theta) = \pi(\theta)f(x|\theta)$ - Joint disbn x & θ
- predictive distribution is $m(x) = \int \pi(\tilde{\theta})f(x|\tilde{\theta})d\tilde{\theta}$ marginal for X
- the posterior distribution for θ is

$$\pi(\theta|X=x) = \frac{f(x,\theta)}{m(x)} = \frac{\pi(\theta)f(x|\theta)}{\int \pi(\tilde{\theta})f(x|\tilde{\theta})d\tilde{\theta}}$$

formula for Bayesian updating.

challenging to calculate

Algorithms allow you to calc. Posterior : eg mcmc

$X = (X_1, ...X_n) \sim i.i.d.\ \ N(\mu, \sigma^2)$ ← $f(\cdot|\theta)$
Suppose σ^2 is known
prior distribution for μ is $N(\mu_0, \tau^2)$ ← $\pi(\cdot)$

$X = (x_1, \dots, x_n)$ $x_i|\theta \overset{iid}{\sim} N(\mu, \sigma^2)$

$\theta \sim N(\mu_0, \tau^2)$

$\begin{Bmatrix} x_1 \\ \vdots \\ x_n \\ \theta \end{Bmatrix} \sim N$

(2) Accuracy of the prior

For posterior: $\theta|x$

$\mu|x \sim N(\tilde{\mu}, \tilde{\sigma}^2)$

accuracy of posterior distn be smaller than frequist approach.

$$\boxed{\frac{1}{\tilde{\sigma}^2} = \frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

acc. of fisher info — acc. of prior

Bayesian
* weighted avg

from data — prior belief of μ

$$\boxed{\tilde{\mu} = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}} \bar{x}_n + \frac{\frac{1}{\tau_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \mu_0}$$

How much weight on each? : (1) Fisher info from sample

If have lot of data, more weight on (1) else on (2)

→ as n↑, prior will get washed out

always bigger

Example

$X_1, ..., X_n \sim$ i.i.d. from Bernoulli (p)
Uniform prior: $\pi(p) = 1$ if $p \in [0, 1]$ and 0 otherwise