Lecture 8. Testing concepts.

Testing

- Hypothesis= statement about population, that can be correct or false
- Simple hypothesis fully specifies distribution of data
- Composite hypothesis= not simple
- Testing hypothesis= deciding based on sample X to accept or reject a null hypothesis
- Critical region C set of values x for which accept

Eg: Is it a fair coin

$X_1, \dots, X_n \sim$ iid Bernoulli(p)

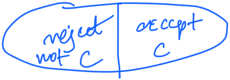
$H_0: p = 1/2$ // simple

composite hypothesis

$$\begin{cases} H_1: p \neq 1/2 \text{ or} \\ H_1: p > 1/2 \end{cases}$$

$X = (X_1, \dots, X_n)$

accept or reject H_0



Example

$X_1, \ldots, X_n \sim$ iid $N(\mu, 1)$

$H_0: \mu \leq$ or $H_1: \mu >$ → we want to test this

$C =$ accept $= \{\bar{X} < 8\}$: accept large values & reject low values.

False positive-false negative

- ☐ Type 1 mistake= reject when the null is true
- ☐ Type 2 mistake = accept the null when it is false
- ☐ Size= probability of type 1 mistake
- ☐ Power= 1-probability of type 2 mistake

accept C	reject not C

more making tradeoff T1 & T2 accepts.

Ho: status quo.
only move when we have compelling evidence, contradictory (rather) move away.
reject Ho: found something

Sample

decision is random bec sample random

population

power = 1 - p(T2)

	Ho is true	H1 is true
accept	correct	type 2
reject	type 1	correct

prob = size

Example: size-power trade-off (worst happens at boundary)

$X_1, \dots, X_n \sim \text{iid } N(\mu, 1)$

$H_0 : \mu \le \mu_0 \qquad H_1 : \mu > \mu_0$

$\bar{X} \sim N(\mu, 1/n)$

accept when $\bar{X} < \delta$

$$\text{size} = P_{H_0}\{\text{reject}\}_{\mu^* \in H_0} = P_{\mu_0}\{\bar{X} > \delta\} = P\{\underbrace{(\bar{X} - \mu^*)\sqrt{n}}_{N(0,1)} \ge (\delta - \mu^*)\sqrt{n}\}$$

$$= 1 - \underset{\text{cdf}}{\Phi}((\delta - \mu^*)\sqrt{n}) : \text{fn} \uparrow \text{in } \mu^*$$

power & size same fn. so cannot max power min size

$\text{size}(\mu_*) \uparrow$ with μ_* and $\downarrow \delta$

$$\text{size} = \max_{\mu^* \in H_0} \text{size}(\mu_0)$$

$$\text{power of } (\underset{H_1}{\mu^*}) = P_{\mu_*}\{\text{reject}\} = 1 - \Phi((\delta - \mu^*)\sqrt{n})$$

5% is the most often size.

Example: test of given size

$\bar{X} \sim N(\mu_0, \frac{1}{n})$

Need to find out: which δ?

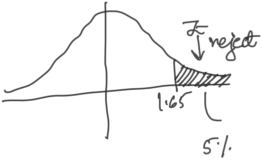
$\{\bar{X} > \delta\}$ = reject

$P_{\mu_0}\{\bar{X} > \delta\} \le 5\%$

Z stat: variance is known.

$$Z = \sqrt{n}(\bar{X} - \mu_0) \Big|_{\text{under } H_0(\mu_0)} \sim N(0,1)$$

centralised, normalised



test $\{$reject when $Z = \sqrt{n}(\bar{X} - \mu_0) > 1.65\}$

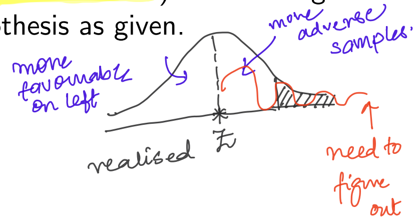
can write: $\bar{X} > \mu_0 + \frac{1.65}{\sqrt{n}}$.

P-value

The *p-value* is the probability (calculated under the null) of obtaining a sample at least as adverse to the null hypothesis as given.

$H_0 : \mu \le \mu_0 \quad H_1 : \mu > \mu_0$

reject $Z = \sqrt{n}(\bar{x} - \mu_0) > 1.65$



pvalue = $1 - \Phi(Z)$

pvalue < 5% — you reject (how often under the null such samples happen)

is uniform

Example

$X_1, \dots, X_n \sim \text{iid} \quad N(\mu, \sigma^2)$

$H_0: \sigma^2 = \sigma_0^2 \quad \text{vs} \quad H_1: \sigma^2 < \sigma_0^2$

$S^2 = \frac{1}{n-1} \sum_i (x_i - \bar{X})^2$

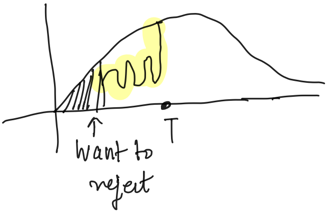
if $\sigma^2 = \sigma_0^2$; $T = \frac{(n-1) S^2}{\sigma_0^2} \sim \chi^2_{n-1}$

reject $T < Q_{\chi^2_{n-1}}(0.05)$

Quantile

p-value: $F_{\chi^2_{n-1}}(T)$

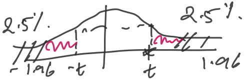
(cdf)



Pivotal Statistics

- ☐ Statistic is pivotal if distribution under the null is known (does not depend on unknown parameters)
- ☐ Asymptotically pivotal= distribution converges to known as the sample size increases

pvalue



Eg) $X_1, \dots, X_n \sim$ iid $\quad EX_i = \mu$

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

|| both null & alt is composite. don't know how data distributed

Want to accept when $|\bar{X} - \mu_0|$ is small

From CLT $\sqrt{n}(\bar{X} - \mu_0) \Big|_{\text{under } H_0} \Longrightarrow N(0, \sigma^2)$

$\uparrow$ Var(Xi)

$s^2 \xrightarrow{P} \sigma^2.$

don't know σ^2 but can approx. using s^2

$$t = \frac{\bar{X} - \mu_0}{\sqrt{s^2/n}} = \frac{\bar{X} - \mu_0}{\underbrace{\sqrt{s^2}}_{\to_P \sigma^2}} \sqrt{n} \underbrace{\xrightarrow[\text{under } H_0]{}}_{\text{By CMP, Slutsky}} N(0,1)$$

accept if $|t| < 1.96$

pvalue $= 2\Phi(-|t|)$

Eg2) $X_1, \dots, X_m \sim$ iid F_X $\quad$ $Y_1, \dots, Y_n \sim$ iid F_Y

(μ_x, σ_x^2) $\quad$ parameters not known. $\quad$ (μ_y, σ_y^2)

$H_0: \mu_x = \mu_y$ $\quad$ $H_1: \mu_x \neq \mu_y$

$\bar{X}_m - \bar{Y}_n$ (if v.v. small can claim / support H_0) but don't know distbn

$\mathrm{Var}(\bar{X}_m - \bar{Y}_n) = \mathrm{Var}(\bar{X}_m) + \mathrm{Var}(\bar{Y}_n)$ — indp. so sum of var's.

$$= \frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}$$

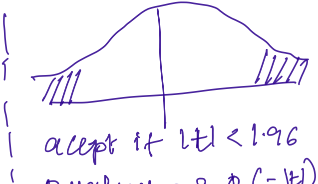
good est. be: $\dfrac{s_x^2}{m} + \dfrac{s_y^2}{n}$

$$t = \frac{\bar{X}_m - \bar{Y}_n}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}} \overset{\text{under } \mu_x = \mu_y}{\Longrightarrow} N(0,1)$$

std. errors

apply all these get to result

- CLT $\bar{X}$
- CLT $\bar{Y}$
- consis s_x^2
- consis s_y^2
- CMT + slutsky.



accept if $|t| < 1.96$

p-values $= 2\,\Phi(-|t|)$

(pop) std. dev $= \sqrt{\mathrm{Var}(\cdot)}$

s.err is an est. of std. dev.

Bootstrap your SE : by calc. variance

Example. Test of the variances. $X = (X_1, ..., X_n) \sim$ i.i.d.
$H_0 : Var(X_i) = \sigma_0^2$
$H_a : Var(X_i) \neq \sigma_0^2$
statistic $z = \sqrt{n}(s^2 - \sigma_0^2)$ and reject when it is too large or too small

$\sqrt{n}(s^2 - \sigma_0^2)$ is gaussian (saw in lecture)
$\Rightarrow N(0, v)$?? proved & now simulate variance by creating own world.

draw $(X_{1b}^*, \dots, X_{n,B}^*)$ from $\hat{F}_n$ (from orig sample values $\{x_1, \dots, x_n\}$ randomly, equally with replacement)

for each sample, calc. $(s_B^*)^2$

$v^* = \frac{1}{1-B} \sum_b (z_b^* - \bar{z}^*)^2$

$v^* =$ sample var $(\underbrace{\sqrt{n}(s_B^{*2} - \hat{\sigma}^2)}_{z_B^*})$

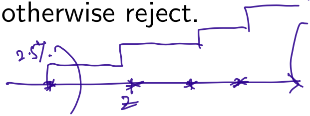
$v^* \longrightarrow v$

$t = \frac{\sqrt{n}(s^2 - \sigma_0^2)}{\sqrt{v^*}} \xrightarrow[\text{under } H_0]{} N(0,1)$ $\xleftarrow{n \to \infty,\ B \to \infty}$

Second way to

Bootstrap

- For $b = 1, ..., B$ repeat:
  - Draw i.i.d. sample $X_b^* = (X_{1b}^*, ..., X_{nb}^*)$ from a set of initial observations $\{X_1, ..., X_n\}$ with replacement;
  - Calculate s_b^2 to be a sample variance of X_b^*;
  - Calculate $z_b^* = \sqrt{n}(s_b^{*2} - s^2)$; — from data.
- Order z_b in ascending order: $z^*_{(1)} \leq ... \leq z^*_{(B)}$;
- For test of size α, if $z^*_{([\frac{\alpha}{2}B])} < z < z^*_{([(1-\frac{\alpha}{2})B])}$ accept the null, otherwise reject.



2.5%

no. of draws & wrt 2 ent.

( )
order statistics.