STATISTICAL INFERENCE

LECTURE 6

Recap }

Normal distro $X \sim N(\mu, \sigma^2)$

- pdf: $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

- If $X_i \sim N(\mu_i, \sigma_i^2)$ are independent,
$$Y = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$
} proof: Use MGF tech.

- If they are i.i.d, then $\mu_i = \mu_j$, $\sigma_i = \sigma_j$ $\forall i, j$
$X_i \sim \underbrace{N(\mu, \sigma^2)}$ $\forall i \in I_n$
indep & same / identical distro

- If they are not independent, Covariance will come into play
eg: Bivariate Normal distro.

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

$X \sim N(\mu_X, \sigma_X^2)$, $Cov(X,Y) = \rho$
$Y \sim N(\mu_Y, \sigma_Y^2)$

- If $X \sim N(\mu, \sigma^2)$ then $Z := \underbrace{\frac{X-\mu}{\sigma}}_{\text{Standardization}} \sim N(0,1)$ ← Standard Normal distro

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

++ CDF is tabulated for z
$X = \mu + \sigma z$.

- Let $X_1, X_2, \ldots X_n$ be a sequence of iid $N(\mu, \sigma^2)$
then $\bar{X} = \sum_{i=1}^{n} X_i / n$ } still an r.v.

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \equiv \sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1)$$

Approx statements

- Weak Law of Large Numbers (WLLN)

  If $X_1, X_2, \ldots$ are iid such that $E(X_i) = \mu$,

  As # samples increases, the sample mean tends to the population mean. in probability.

  $$\lim_{n\to\infty} \bar{X} \xrightarrow[\text{prob.}]{\text{in}} E(X) \text{ or } \mu.$$

  weak convergence.

- Central Limit Theorem (CLT)

  note: last line of prev page.
  $\sqrt{n}(\bar{X}-\mu)/\sigma \sim N(0,1)$ for $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$

  CLT generalizes this.

  If $X_1, X_2, \ldots$ are iid such that $E(X_i) = \mu$,

  The normalized sample mean for ANY "nice" distro. converges to std. normal distro. in distribution. as # samples increases.

  Nice := finite variance ($\sigma^2 < \infty$)

  stronger claim.

* Many other forms & formal statements are online. in terms of $\sum X_i$ for instance.

- Probability ⇝ theoretical base.

  ⇝ assumes many things like i.i.d, $N(\mu, \sigma^2)$, ...

-- In practice this is rare. All you have are samples $x_1, x_2, \ldots, x_n$ (finite, may/not be iid, distro unknown, even if distro is known, parameters unknown ...)

-- Closed form of CDF/pdf of continuous rvs are not specified/known. We have access to an Empirical (=experiment based) frequency distro.

Statistical inference

Toolkit to estimate things about population ⇝ "parameter"

from things about sample ⇝ "statistic"

Inference → Non parametric (difficult, but flexible, less assumptions)

→ Parametric → point estimation. } like this.

→ interval estimation.

estimation & hypothesis testing

Parametric: we know that the population is characterized by a distro, with some parameters

LECTURE 7

POINT ESTIMATION

Given: $X_1, X_2, \ldots X_n \overset{iid}{\sim}$ some parametric distro.

eg: Say we are given outcomes of tossing a coin.

(A) fit a suitable model.

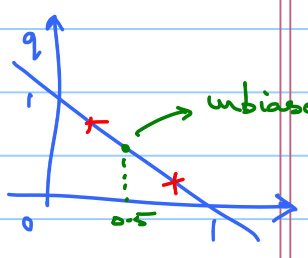
no info about the coin being biased / unbiased.

Bernoulli (p). $\underbrace{0 < p < 1}_{\text{unknown.}}$

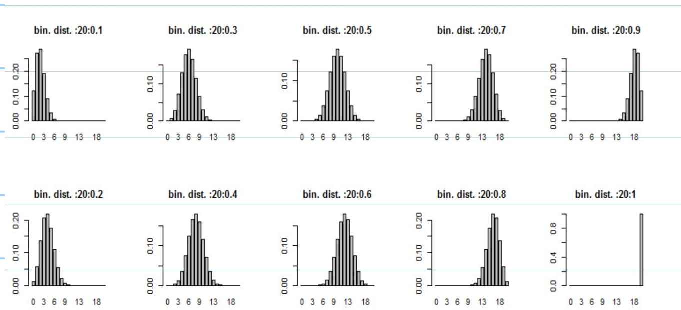
(B) find a likely value of p from sample & make conclusions about the cause (coin in our case)

Harvey Dent's Biased Two headed coin - Batman wiki

- Say the coin gives output $\{0,1,1,1,1,1,0,1,1,1,\ldots\}$
  intuition: a biased coin, gives mostly heads.
- Say the coin gives output $\{0,1,1,0,0,1,0,1,0,1,\ldots\}$
  intuition: may be an unbiased coin



unbiased coin



Given the data, what value of p is most likely to have caused that data?

If $P(\text{head}) = p$ & $P(\text{tail}) = q$ from data,

all we know is that $p + q = 1$.

What is the most optimal solution?

(C) One method MLE

MLE : Maximum Likelihood Estimation.

given $X_1, X_2, \dots X_n \overset{iid}{\sim} f_\theta(x)$,

optimum parameter $\Big\}$ $\theta_* = \underset{\theta \in \Theta}{\operatorname{argmax}} \underbrace{f(x_1, x_2, \dots, x_n; \theta)}$ joint pdf of X_i's under the assumption pdf is function of θ.

possible values of θ

parameter space.

Choose the θ that maximizes the jpdf of X_i's $\equiv$ the chance of observing the sequence. $(x_1, x_2, \dots x_n)$

Due to the iid assumption $\Big\}$, $f_\theta(x_1, x_2, \dots, x_n) = f_\theta(x_1) \cdot f_\theta(x_2) \cdots f_\theta(x_n)$

$$\theta_* = \underset{\theta}{\operatorname{argmax}} \left(\prod_{i=1}^{n} f_\theta(x_i) \right)$$ called the Likelihood function $L(\theta \mid x)$

eqvt to $\Big\}$

$$\theta_* = \underset{\theta}{\operatorname{argmax}} \log \left(L(\theta \mid x) \right)$$

(due to monotonicity of log function)

$$\theta_* = \underset{\theta}{\operatorname{argmax}} \left(\sum_{i=1}^{n} \log f_\theta(x_i) \right)$$ Log likelihood function.

eg: For the coin toss example,

$$\theta \equiv p, \quad \Theta = [0, 1]$$

$$f_\theta(x_i) = p^{x_i} (1-p)^{1-x_i}$$

- $$L(\theta \mid x) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$
- $$\mathcal{L} = \log L(\theta \mid x) = \sum_{i=1}^{n} x_i \log p + (1-x_i) \log (1-p)$$
- $$p_* = \underset{p \in [0,1]}{\operatorname{argmax}} \left((\log p) \sum x_i + \left(n - \sum x_i\right) \log (1-p) \right)$$

@ p_*, $\dfrac{d\mathcal{L}}{dp} = 0$ (maxima)

$$\Rightarrow \left. \frac{d\mathcal{L}}{dp} \right|_{p_*} = \left(\sum x_i\right) \frac{1}{p_*} + \left(n - \sum x_i\right) \left(\frac{-1}{1-p_*} \right)$$

$$p_* = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$p_{*_{MLE}} = \bar{x}$$

$\rightsquigarrow$ p_{MLE} = proportion of heads in the empirical data.

Examples from problem set 5.

1.

a) $\{X_i\}_{i\in I_n} \overset{iid}{\sim}$ Poisson (λ).

$$f(x_i) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}, \quad x_i = 0,1,2,3,\ldots$$

$$L = \prod_i f(x_i|\lambda) = \prod_i \frac{e^{-\lambda}\cdot\lambda^{x_i}}{(x_i)!} = \frac{e^{-n\lambda}\cdot\lambda^{\sum_i x_i}}{\prod_i (x_i)!}$$

$$\mathcal{L} := \log L = -n\lambda + \sum_i x_i \log\lambda - \sum \ln(x_i!)$$

$$\left.\frac{\partial\mathcal{L}}{\partial\lambda}\right|_{\hat\lambda} = 0 \Rightarrow \quad -n + \left(\sum_i x_i\right)\frac{1}{\hat\lambda} = 0$$

$$\Rightarrow \quad \boxed{\hat\lambda = \sum_i x_i / n}$$

b) $\{X_i\}_{i\in I_n} \overset{iid}{\sim}$ Geo (p) // # of failures b4 I success

$$f(X = x_i) = (1-p)^{x_i - 1} p, \quad x_i = 1,2,\ldots$$

I success in x_i^{th} trial.

$$L = \prod_i f(x_i) = \prod_{i=1}^{n} (1-p)^{x_i - 1} p$$

$$\mathcal{L} = \ln L = \ln\left[(1-p)^{\sum x_i - n} p^n\right] = n\ln p + \left(\sum x_i - n\right)\ln(1-p)$$

$$\left.\frac{\partial\mathcal{L}}{\partial p}\right|_{\hat p} = 0 \Rightarrow \quad \frac{n}{\hat p} + \left(\sum_i x_i - n\right)\frac{1}{1-\hat p}(-1) = 0$$

$$\frac{1}{\hat p} - 1 = \frac{\sum x_i - n}{n} = \bar{X} - 1$$

$$\boxed{\hat p = \frac{1}{\bar X}}$$

// Observe: MLE is not always $\bar X$.

5. $\{X_i\}_{i=1}^{n} \overset{iid}{\sim} N(\mu, \sigma^2)$

to find:

a) • MLE (σ^2)

b) • Sample variance of σ^2.

⇝ a) likelihood, $L = (2\pi\sigma^2)^{-n/2} \exp\left[-\sum_i \frac{(X_i-\mu)^2}{2\sigma^2}\right]$

$$\left.\begin{array}{c}\text{log likelihood,} \\ \mathcal{L}\end{array}\right\} = \left(\frac{-n}{2}\right)\log(2\pi\sigma^2) - \sum_i \frac{(X_i-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = \left(\frac{-n}{2}\right)\left(\frac{1}{\sigma^2}\right) - \sum_i \frac{(X_i-\mu)^2}{2}\left(\frac{-1}{\sigma^4}\right)$$

@ MLE $\hat{\sigma}^2$, $\frac{\partial \mathcal{L}}{\partial \sigma^2} = 0$

$$\Rightarrow \frac{-n}{2\sigma^2} = -\frac{\sum (X_i-\mu)^2}{2\sigma^4}$$

$$\Rightarrow \hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{n}$$

- if μ is known, $\hat{\sigma}^2 = \sigma^2$
- if μ is not known, use MLE of μ.

$$\hat{\mu} = \bar{X}$$

$$\hat{\sigma}^2 = \sum_i \frac{(X_i - \bar{X})^2}{n}$$

$$= \left(\frac{n-1}{n}\right) \sum_i \frac{(X_i-\bar{X})^2}{(n-1)} \longrightarrow \text{sample variance.}$$

$$\boxed{\hat{\sigma}^2 = \left(\frac{n-1}{n}\right) s^2}$$

// Find MLE of m & p, Binomial (m, p), → in video.
MLE of μ for $N(\mu, \sigma^2)$

- MLE is one method of estimating the parameters.

- Other functions of samples can also be used.

$$\Theta_* = g(X_1, X_2, X_3, \dots, X_n)$$

g → called the <u>estimator</u> • the function.

Θ_* → Called the <u>estimate</u>. • the value.

Uniform distros } • $\{X_i\}_{i=1}^n \overset{iid}{\sim} U(0, \theta)$

estimators : $\hat{\theta} = \max \{X_i\}_{i=1}^n$

$\hat{\theta} = 2\bar{X}$; $\hat{\theta} = 12\sqrt{S_x^2}$

S_x^2 : sample variance.

⟹ Method of Maximizing Moments (MME, often used).

⚠ Other methods: // Not in this course

Method of
a) Least Squares
b) Minimum variance
c) Minimum Chi-square
d) Inverse probabilities

⋮

* Properties of estimators.

• Given two estimators, which one is better?

a) Bias : $|E(\hat{\theta}) - \theta| \equiv$ How far θ is from expected value of $\hat{\theta}$

If we repeat the experiment many times, does $\hat{\theta} \to \theta$? on an average?

b) Variance : $E\left([\hat{\theta} - E(\hat{\theta})]^2\right)$

Mean Squared Error } $MSE = E\left([\hat{\theta} - \theta]^2\right)$

$= \text{bias}^2 + \text{Variance}.$

Other Characteristics } • Consistency • Sufficiency. // Not in this course

- $\{X_i\}_{i \in I_n} \sim U(0, \theta)$

a) $\hat{\theta}_2 = 2\bar{X}$ $\quad E(\hat{\theta}) = 2E(\bar{X}) = 2 \cdot \frac{\theta}{2} = \theta$

<u>Unbiased</u>

$$MSE(\hat{\theta}_2) = Var(\hat{\theta}_2) = 4V(\bar{X})$$
$$= 4 \cdot \left(\frac{1}{n^2} \cdot n V(X)\right) = \frac{\theta^2}{3n}$$

- $\hat{\theta}_1 = \max\{X_i\}$ $\quad f_{\hat{\theta}_1}(x) = \left\{ \frac{n}{\hat{\theta}_1} x^{n-1} \right.$

$E(\hat{\theta}_1) = \frac{n}{n+1}\theta$

<u>biased</u>

& underestimates θ.

$$MSE(\hat{\theta}_1) = \frac{2\theta^2}{(n+1)(n+2)}$$

$\rightsquigarrow$ Biased, but $MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$

⚠ Example 7.7 in Sheldon Ross.

<u>LECTURE 8</u>

Some properties of MLE:

++ Usually gives unbiased & low MSE estimates.

-- If sample is small, bias maybe large.

++ Well understood method.

-- Strong assumptions about the observations.

++ MLE of iid samples is consistent estimator.

-- Computationally intensive.

-- <u>Likelihood function</u> & <u>optimization problems</u>.

Very complex in some cases.

- existence of solution via calculus.
- Multiple solutions.
- minimization w.r.t one param may increase cost w.r.t others.

* INTERVAL ESTIMATION.

→ A range in which we have a high confidence that the parameter might lie is more useful & mathematically treatable compared to a point estimate.

eg: • It's not very sensible to call a coin unbiased if $p = 0.48$ or $p = 0.54$. (Sampling errors maybe)
- For given # samples, & confidence level, we'd like an interval (L, U) such that $P[\theta \in (L, U)]$ is upto the specified confidence.

We usually calculate this interval from the point estimate $\hat{\theta}$ &/or its properties.

- Read more here : https://en.wikipedia.org/wiki/Confidence_interval
- https://en.wikipedia.org/wiki/Type_I_and_type_II_errors
- The confidence is specified by α. (Common: 0.05, 0.01, 0.1, 0.005 — 5%, 1%, 10%, 0.5%)

Consider Covid-19 testing.

Contingency table →

Test	Truth: Has disease	Truth: No disease	
Detects disease	Good test (True positive)	Bad test (False positive)	Type I error
Doesn't detect	Very Bad test (False negative)	Good test (True negative)	
	Type II error		

Type I error probability } α ⟵ Reducing one increase other. ⟶ Type II error probability } β

- This is the confidence that's usually given.
- More serious than α in consequence. (In some situations)

* Confidence of $(1-\alpha)$ or $100(1-\alpha)\%$ means that if we repeat the test sufficiently large # times, we can expect θ to lie in the interval (L, U) $100(1-\alpha)\%$ # times.

* A tight interval (small $U - L$) is sensible & desired. (eg. 3σ limits around μ has almost all samples - it's not useful).

- ($\hat{\theta}-\varepsilon$ | $\hat{\theta}$ | $\hat{\theta}+\varepsilon$) } 2ε interval around $\hat{\theta}$ → the point estimate

  $\hat{\theta}$ ← θ_{MLE}, θ_{MME}, ...

- Need not be symmetric interval (eg in skewed distros, eg. χ^2, one sided intervals, etc).

* Mainly $P\left[\theta \in (\hat{\theta}-\varepsilon, \hat{\theta}+\varepsilon)\right] \geq 1-\alpha$

$= P\left[|\theta - \hat{\theta}| < \varepsilon\right] = 1-\alpha$

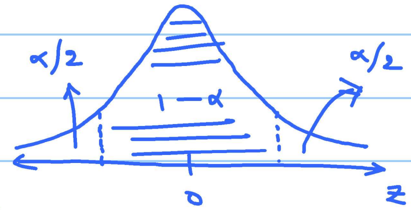
* Say $\{X_i\}_{i=1}^{n} \overset{iid}{\sim} N(\mu, \sigma^2)$

a) Estimating Confidence Interval for μ.

a.1) σ^2 = pop. variance is known.

A good candidate for μ is $\bar{X}$.

We want to find ε s.t } $P\left[|\bar{X}-\mu| < \varepsilon\right] = 1-\alpha$

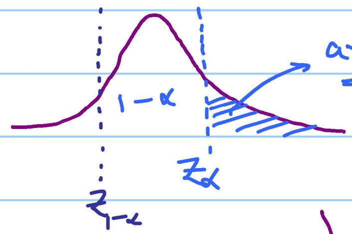


$\Rightarrow P\left[\left|\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < \dfrac{\varepsilon\sqrt{n}}{\sigma}\right] = 1-\alpha$

$\Rightarrow P\left[-\dfrac{\varepsilon\sqrt{n}}{\sigma} < z < \dfrac{\varepsilon\sqrt{n}}{\sigma}\right] = 1-\alpha$

$\equiv F\left(\dfrac{\varepsilon\sqrt{n}}{\sigma}\right) - F\left(-\dfrac{\varepsilon\sqrt{n}}{\sigma}\right) = 1-\alpha$

Def: $z_\alpha :=$ z such that $P\left[z > z_\alpha\right] = \alpha$



- By symmetry, $P\left[z > z_{1-\alpha}\right] = 1-\alpha$

  or $P\left[z < z_{1-\alpha}\right] = \alpha$

Common z values

$\Rightarrow z_{\alpha/2} = \dfrac{\varepsilon\sqrt{n}}{\sigma} \leadsto \varepsilon = z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

$\Rightarrow$ The $100(1-\alpha)\%$ Confidence interval for μ is

$$\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

→ These $z_{\alpha/2}$ or z_α (t_α, χ^2_α, ...) are tabulated, called critical values @ some α or table values.

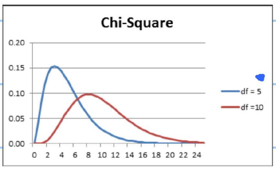
Confidence level	z critical value
80%	1.28
90%	1.645
95%	1.96
98%	2.33
99%	2.58
99.8%	3.09
99.9%	3.29

LECTURE 9

a.2) $100(1-\alpha)\%$ CI for μ of $N(\mu, \sigma^2)$,
σ^2 is unknown.

$\leadsto$ good candidate for σ^2, its estimate sample variance. $\hat{\sigma} = S$

Recap:



- $\chi^2_n := \sum_{i=1}^{n} Z_i^2$ Chi-squared $\equiv$ sum of squares of n indep. std. normal variates (dof : n)

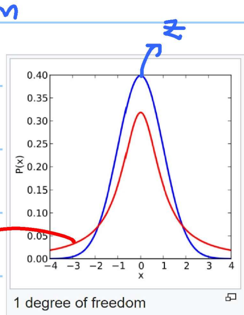
$$\frac{(n-1)S_x^2}{\sigma^2} \sim \chi^2_{(n-1)} \quad \text{if } \{X_i\}_{i=1}^{n} \overset{iid}{\sim} N(\mu, \sigma^2)$$

- $$\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \quad \left\{ \underline{\underline{\text{Not}}} \text{ normal / std normal} \right\}$$

$$= \left(\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}\right) \Bigg/ \left(\sqrt{\frac{(n-1)S^2}{\sigma^2} \times \frac{1}{n-1}}\right)$$

$$= \text{std normal variate} \Bigg/ \sqrt{\frac{\chi^2_{n-1} \text{ variate}}{\text{deg of freedom}}} \qquad \underline{\underline{n-1}}$$

Student's-t random variable.
https://en.wikipedia.org/wiki/Student%27s_t-distribution



1 degree of freedom

* $t_n = \dfrac{z^2}{\sqrt{\chi^2_n}}$
- t is heavy tailed
- but $t \to z$ as $n \to \infty$

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1} \quad \Big\} \text{ for small samples, use } t \text{ distbn.}$$

these are tabulated for different n.

$$\Rightarrow \quad P\left[-t_{(n-1),\alpha/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{(n-1),\alpha/2}\right] = 1-\alpha$$

$\Rightarrow$ $100(1-\alpha)\%$ CI for μ when σ^2 is unknown is

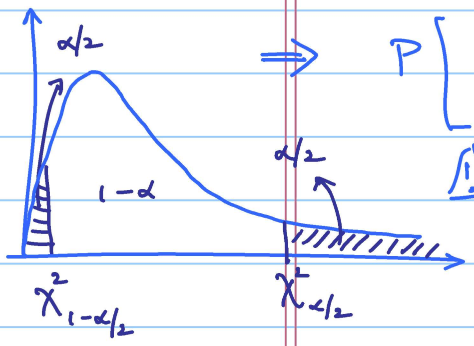
$$\left(\bar{X} - t_{(n-1),\alpha/2}\frac{S}{\sqrt{n}},\ \bar{X} + t_{(n-1),\alpha/2}\frac{S}{\sqrt{n}}\right)$$

Mnemonic ⚠ Replace everything in a.1) formula by $z \to t_{n-1}$ & $\sigma \to S$

⚠ Last row of t table is z critical values. ⚠ Not n

b) $100(1-\alpha)\%$ CI for σ^2 in $N(\mu, \sigma^2)$

Start: $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$



$$\Rightarrow P\left[\chi^2_{(n-1),1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{(n-1),\alpha/2}\right] = 1-\alpha$$

⚠ χ^2 support is $[0,\infty)$, so no negative values.

⚠ $(n-1)$ deg of freedom, not n in table.

$$\Rightarrow P\left[\frac{(n-1)S^2}{\chi^2_{(n-1),\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(n-1),1-\alpha/2}}\right] = 1-\alpha$$

$\Rightarrow$ $100(1-\alpha)\%$ CI for σ^2 of $N(\mu,\sigma^2)$ is

$$\left(\frac{(n-1)S^2}{\chi^2_{(n-1),\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{(n-1),1-\alpha/2}}\right)$$

LECTURE 10

HYPOTHESIS TESTING.

Aim
- To test if the sample supports/rejects a claim/hypothesis.
  data ~ claim or not.

Structure: (in gen)
- Null Hypothesis } $H_0 : f(x,\theta) = g_1(x_1, x_2, \ldots, x_n)$
  eg: $H_0 : \mu = 50\,g$
- Alternate Hypothesis (or H_1) } $H_a : f(x,\theta) = g_2(x_1, x_2, \ldots, x_n)$
  eg: $H_1 : \mu = 60g$
  $H_1 : \mu > 50g$ or $\mu \neq 50g$ ...
- Test statistic – a function of the samples
  T eg: $\sum x_i$ or $\bar{x}$ or S_x ...
  → This reduces the samples to a single numerical value
- Level of Significance } $\alpha \equiv$ Maximum permissible Type I error
  eg: 5% or 1%.

Miss detection $\alpha = P[\text{reject } H_0 \mid H_0 \text{ is true}]$

False alarm $\beta = P[\text{do not reject } H_0 \mid H_0 \text{ is not true}]$

* We talk of rejecting (or not) H_0. We don't say "accepting" H_0.

- Critical/rejection region } the possible values of T for which we reject H_0.
  - $\hookrightarrow$ Its complement is called acceptance region.

Eg 1: Test for normal mean μ

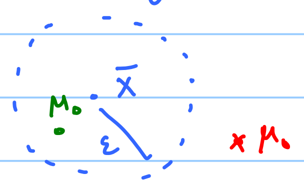
given: $\{X_i\}_{i=1}^{n} \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 known.

a) fix Hypotheses
- $H_0 : \mu = \mu_0$
- $H_a : \mu \neq \mu_0$

b) fix test statistic
- $T: \bar{X} = \frac{\sum X_i}{n}$ } We know its distribution ++ / It is MLE ++, WLLN, CLT, ...

c) fix α
- LOS: α — fix this before checking the sample or finding the value of T to avoid biasing yourself.

Rejection region }
- $D_c: \{(X_1, X_2, \dots, X_n) : |T - \mu_0| > \varepsilon\}$
  - If sample $\in D_c$ reject H_0.
  - ε Depends on $\alpha, n, \dots$



test $\implies$ $P\Big[\underbrace{|\bar{X} - \mu| > \varepsilon}_{\text{(rejecting } H_0\text{)}} \,\Big|\, \underbrace{\mu = \mu_0}_{\text{(}H_0\text{ is true)}}\Big] = \alpha$

$$\implies P\left[\,|\bar{X} - \mu_0| > \varepsilon\,\right] = \alpha \qquad \bar{X} - \mu_0 \sim N(0, \sigma^2)$$

$$\implies P\left[\,\left|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right| > \frac{\varepsilon}{\sigma/\sqrt{n}}\right] = \alpha$$

$$\rightarrow Z_{\alpha/2} = \frac{\varepsilon}{\sigma/\sqrt{n}} \implies \varepsilon = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

d) find the value of T $\rightsquigarrow$ e) Conclusion

If $|\bar{X} - \mu_0| > \varepsilon = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ reject H_0

If $|\bar{X} - \mu_0| \leq \varepsilon$ donot reject H_0.

} @ given Level of significance.

⚠ Another strategy }
- fix test statistic

T as $Z := \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ $\rightsquigarrow$ If $Z_{calc} > Z_{\alpha/2}$ reject H_0.

* Critical region depends on
  a) Level of significance. ($\alpha = 1\%$ is stricter than $\alpha = 5\%$)
  b) Type of test / hypotheses.

$$\left\{\begin{array}{l} \text{if } H_a : \mu > \mu_0 \ , \text{ one-sided test} \\ \quad\ \ H_a : \mu < \mu_0 \ , \text{ use } z_\alpha \end{array}\right\}$$

Eg 2: Test for μ, σ^2 unknown.

Use t-test instead of z., s instead of σ

T: $\dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = t_{calc}$ (say).

If $t_{calc} < \underbrace{t_{(n-1),\alpha/2} \text{ or } t_{(n-1),\alpha}}$ donot reject H_0.

depends on H_0.

called the "table value" t_{tab}.

Eg 3: Test for variance of $N(\mu, \sigma^2)$

H_0: $\sigma^2 = \sigma_0^2$

H_1: $\sigma^2 \neq \sigma_0^2$

Rejection Region } reject H_0 if $\sigma_0 \notin \left(\dfrac{(n-1)s^2}{\chi^2_{\alpha/2,(n-1)}}, \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2,(n-1)}} \right)$

Ex 4: Test for equality of means.

$\equiv$ Do both samples come from the same population / inherent distribution. (keeping sampling errors apart)

Given: $\{X_i\}_{i=1}^{n} \overset{iid}{\sim} N(\mu_x, \sigma_x^2)$, variance σ_x^2 & σ_y^2 known.

$\{Y_i\}_{i=1}^{m} \overset{iid}{\sim} N(\mu_y, \sigma_y^2)$, $X \perp\!\!\!\perp Y$ (independent)

H_0: $\mu_x = \mu_y$

H_a: $\mu_x \neq \mu_y$

Test statistic } T: $\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}\right)$

$$P\left[\left| \bar{x} - \bar{y} - (\mu_x - \mu_y) \right| \le z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \ \middle| \ \mu_x = \mu_y \right] = 1 - \alpha$$

Reject H_0 if: $|\bar{x} - \bar{y}| > z_{\alpha/2} \sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}$

Given: $\{X_i\}_{i=1}^{n} \overset{iid}{\sim} N(\mu_x, \sigma_x^2)$, variance σ_x^2 & σ_y^2 unknown.

$\{Y_i\}_{i=1}^{m} \overset{iid}{\sim} N(\mu_y, \sigma_y^2)$, $X \perp\!\!\!\perp Y$ (independent)

$H_0: \mu_x = \mu_y$

$H_a: \mu_x \neq \mu_y$

- $\dfrac{(n-1)S_x^2}{\sigma_x^2} \sim \chi^2_{n-1}$; $\dfrac{(m-1)S_y^2}{\sigma_y^2} \sim \chi^2_{m-1}$

$$\Rightarrow \quad \frac{(n-1)S_x^2}{\sigma_x^2} + \frac{(m-1)S_y^2}{\sigma_y^2} \sim \chi^2_{n+m-2}$$

Another important assumption for simplification $\sigma_x^2 = \sigma_y^2 = \sigma^2$ (same population)

$\Rightarrow \quad \sim \chi^2_{n+m-2}$

$$\frac{\overbrace{\dfrac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{\sigma^2\left(\frac{1}{m}+\frac{1}{n}\right)}}}^{\text{\& std normal}}}{\underbrace{\sqrt{\dfrac{(n-1)S_x^2+(m-1)S_y^2}{\sigma^2}\left(\dfrac{1}{n+m-2}\right)}}_{\sqrt{\frac{\chi^2_{n+m-2}}{n+m-2}}}} \sim t_{n+m-2}$$

Test Statistic $\Big\}$ T: $$\frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\left(\frac{1}{m}+\frac{1}{n}\right)\underbrace{\sqrt{\dfrac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}}}}$$ $\Big\}$ the pooled estimator S_p^2 of variance. σ^2

$$P\left[\,\left|\bar{X}-\bar{Y}-(\mu_x-\mu_y)\right| \le t_{\alpha/2,\,(n+m-2)}\sqrt{S_p^2\left(\frac{1}{m}+\frac{1}{n}\right)}\;\middle|\; \mu_x=\mu_y\right] = 1-\alpha$$

Reject H_0 if:

$$t_{calc} := \frac{|\bar{X}-\bar{Y}|}{\sqrt{S_p^2\left(\frac{1}{m}+\frac{1}{n}\right)}} > t_{(n+m-2),\alpha/2} \;(= t_{tab})$$

- The other case when $\sigma_x^2 \neq \sigma_y^2$ is complicated.
  - ↳ If n, m are large, $\dfrac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}}$ is approximately normal, use z test.