Chapter 9. Statistical Inference I.

9.1. Parameter Estimation:

Binomial : n, p, $E(x)=np$. | Exponential : λ $E(X)=\frac{1}{\lambda}$

Poisson : λ $E(x)=\lambda$ | Normal: u, δ. $E(X)=\frac{a+b}{2}$

Geometric : p . $E(x)=\frac{1}{p}$

The k-th population moment is defined as:

$$u_k = E(X^k)$$

The k-th sample moment :

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Method of moments: To estimate k parameters, equate the first k population moment and sample moment:

$$\begin{cases} u_1 = m_1 \\ u_2 = m_2 \\ \vdots \\ u_k = m_k. \end{cases}$$

Method of maximum likelihood:

Requires using joint prob functions, but we can assume independence.

Discrete : $p(x)$.

$$p(x_1, x_2, \cdots, x_n) = p(x_1) \cdots p(x_n)$$

Continue : $f(x)$

$$f(x_1, \cdots, x_n) = f(x_1) \cdot f(x_2) \cdot \cdots \cdot f(x_n)$$

} independent

Then "Maximize" = calculus.

Set $g(p) = \ldots\ldots$

Then set $g'(p) = 0$ to get the max value of p.

9.2. Confidence Interval.

Given $\quad P\{a \le \theta \le b\} = 1-\alpha$,

then we say that $[a,b]$ is a $(1-\alpha)100\%$ confidence interval for θ.

$1-\alpha$ is coverage probability or a confidence level.

Typically, we use 95%, then $1-\alpha = 95\%$

θ is fixed, and $[a,b]$ is random.

Method for Finding $[a,b]$.

1). Assumptions: ①. Estimate θ with $\hat{\theta}$.
e.g.: $\mu \quad \bar{X}$

②. $\hat{\theta}$ is unbiased. (i.e. $E(\hat{\theta}) = \theta$).

③. $\hat{\theta}$ is normally distribution. (or approx normal)

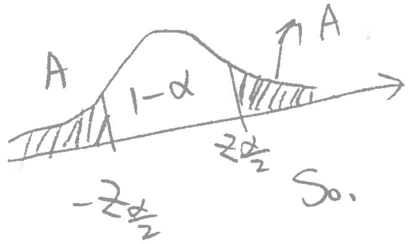
2). We pick $1-\alpha$.

3). We standarlize $\hat{\theta}$:

Assume. $E(\hat{\theta}) = \theta$.

$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sigma(\hat{\theta})} = \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})}$$

$\sigma(\hat{\theta})$ → standard error of $\hat{\theta}$.



$P\{Z > Z_A\} = A$.

So. $\quad P\{-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\} = 1-\alpha$.

$\Downarrow$

$$P\left\{-Z_{\frac{\alpha}{2}} < \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})} < Z_{\frac{\alpha}{2}}\right\} = 1-\alpha$$

Solve for θ.

$$P\{\underbrace{\hat{\theta} - \delta(\hat{\theta})\cdot z_{\frac{\alpha}{2}}}_{a} < \theta < \underbrace{\hat{\theta} + \delta(\hat{\theta})\cdot z_{\frac{\alpha}{2}}}_{b}\} = 1-\alpha$$

$P\{z > z_{\frac{\alpha}{2}}\} = \frac{\alpha}{2}$

$\Downarrow$

$P\{z < z_{\frac{\alpha}{2}}\} = 1 - \frac{\alpha}{2}$.

$\Downarrow$

try to find $z_{\frac{\alpha}{2}}$.

$$\underbrace{\hat{\theta}}_{\text{center}} \pm \underbrace{\delta(\hat{\theta})\cdot z_{\frac{\alpha}{2}}}_{\text{margin}}$$

$$\delta(\hat{\theta}) = \frac{\delta}{\sqrt{n}}$$

Higher confidence = larger Internal.

Margin for population mean:

i). Suppose we want this not more than ε.

$$z_{\frac{\alpha}{2}} \cdot \frac{\delta}{\sqrt{n}} \le \varepsilon$$

$$\Downarrow$$

$$n \ge \left(\frac{z_{\frac{\alpha}{2}} \cdot \delta}{\varepsilon}\right)^2$$

9.3 Confidence Interval: unknown σ.

Two broad situations will be considered:

i). Large samples from any distribution.

ii). Samples of any size from a normal distribution.

i). Large samples:

A large sample should produce a rather accurate estimator of a variance
we can then replace the true standard error $\sigma(\hat{\theta})$ by its estimator $s(\hat{\theta})$:

$$\bar{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \longrightarrow \underbrace{\bar{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}} \rightarrow \text{only depends on sample.}$$

Next case: confidence intervals for proportions.

$$p = \frac{\text{\# of units with same attribute}}{\text{total size}}.$$

$$\text{estimate with } \hat{p} = \frac{\text{\# of observed units with attribute}}{\text{sample size } (n)}.$$

$$\hat{p} = \bar{X},$$

Now we have a sample:

$$S = (x_1, x_2, \cdots, x_n) \text{ with } X_i = \begin{cases} 1, & \text{if } X_i \text{ has the attribute} \\ 0, & \text{otherwise} \end{cases}$$

So, X_i is Bernoulli, we have:

$u = E(X_i) = \underbrace{p}_{}$, $Var(X_i) = p(1-p)$, $\sigma(X_i) = \sqrt{p(1-p)}$.
unknow.

We estimate p using: $\hat{p} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \bar{X}$.

$\hat{p}$ is unbiased.

So. $E(\hat{p}) = E(\bar{X}) = u = E(X) = p$, $Var(\hat{p}) = \dfrac{p(1-p)}{n}$, $\sigma(\hat{p}) = \sqrt{\dfrac{p(1-p)}{n}}$

If n is large, we can replace $\delta(\hat{p})$ with $s(\hat{p})$.

$$s(\hat{p}) = \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

Therefore, Formula for Confidence Interval of proportion:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

gives a $(1-\alpha)\times 100\%$ confidence interval $\boxed{\text{for } \hat{p}}$

Selecting n based on error ε.

$$\varepsilon = z_{\frac{\alpha}{2}} \cdot \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

$$\Rightarrow n \geq \frac{\hat{p}\cdot(1-\hat{p}) \cdot (z_{\frac{\alpha}{2}})^2}{\varepsilon^2}.$$

Confidence interval for the difference of means: Unequal, unknown standard deviations.

$$\bar{X} - \bar{Y} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}.$$

9.4. Hypothesis Testing.

H_0 = hypothesis. (the null hypothesis).

H_A = alternative (the alternative hypothesis).

H_0 and H_A are simply two mutually exclusive statements.

Hypothesis H_0: $u = u_0$.

then Alternative: i). H_A: $u \neq u_0$ is a two-sided alternative.

ii). H_A: $u > u_0$, is one-side alternative

iii) H_A: $u < u_0$, is one-side alternative.

Type I and Type II errors: level of significance.

more dangerous and most costly.

	Reject H_0	Accept H_0
H_0 is true	Type I error	Correct
H_0 is false	Correct.	Type II error.

Probability of a Type I error is the significance level of a test.

$$\alpha = P\{\text{reject } H_0 \mid H_0 \text{ is true}\}.$$

Computing z:

1). Testing a mean: H_0: $u = u_0$.

Use $z = \dfrac{\bar{x} - u_0}{\sigma/\sqrt{n}}$ .... Standarlize $\bar{x}$ assuming H_0 is tru

often approximate with s if n is large.

Proportion: H_0: $p = p_0$.

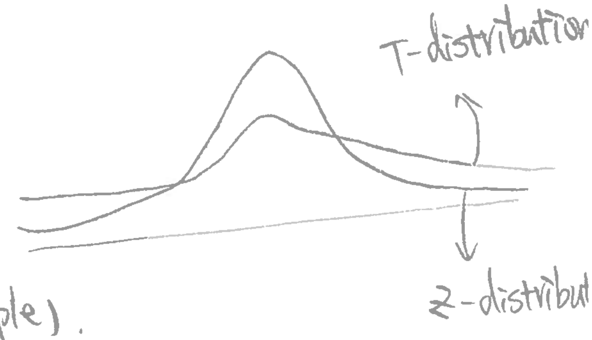
Use $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ .... — ①

Then we find z_α according to significance level α.

2). T test: unknown σ and we don't necessarily have a large sample, so s is not a good estimator of σ.

In the case of estimating μ using $\bar{x}$, we form:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad \ldots \text{ the } t\text{-ratio.}$$



$t_\alpha > Z_\alpha$.

(the cost of a small sample).

As $n \to \infty$, $t_\alpha \to Z_\alpha$.

The T-test for means for unknown σ and small n,

Here, $H_0 : \mu = \mu_0$.

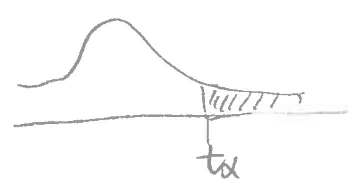
Test statistic: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ $\quad s^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$s = \sqrt{s^2}$.

i). Right-tail alternative:

$H_A : \mu > \mu_0$

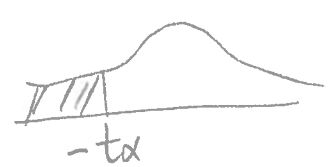
{ reject H_0 if $t \geq t_\alpha$.
  Accept H_0 if $t < t_\alpha$.



ii) Left-tail alternative:

$H_A : \mu < \mu_0$.

{ Reject H_0 if $t \leq -t_\alpha$
  Accept H_0 if $t > -t_\alpha$



iii). Two-sided alternative:

$H_A : \mu \neq \mu_0$.

{ Reject H_0 if $|t| \geq t_{\alpha/2}$.
  Accept H_0 if $|t| < t_{\alpha/2}$