— Overview of statistics   Ch 8

• — Point Estimation
  - Statistic and estimator + Examples

• — Bias & Mean Square Error
  - unbiasedness, Bias, MSE($\hat{\theta}$)
    decomposition of MSE (# 8.8 ($\hat{\theta}_1, \hat{\theta}_5$), p 394, # 8.6, p 394)

• — Common unbiased estimators
  - $\mu, p, \mu_1 - \mu_2, p_1 - p_2$ & $\sigma^2 \rightsquigarrow S^2_{n-1} \to S^2$ in the textbook

• — Error of the estimation
  - $\varepsilon = |\hat{\theta} - \theta| \to$ Tchebyscheff's Theorem if $\hat{\theta}$ is an unbiased estimator (Example 8.2, p 401)

• — Confidence Interval
  - pivotal Quantities

C.I.
- large n → C.L.T. & Asymptotic Normality of the MLE
- Small n
  - pivotal Quantity probab. Integral transform
  - Normal dist.

Pivotal:

$$X_i \sim F_\theta(x) \Rightarrow Y_i = F_\theta^{-1}(X_i) \sim Unif(0,1) \Rightarrow -\log Y_i \sim Exp(1)$$

$$\Rightarrow \sum_{i=1}^{n} -\log F_\theta^{-1}(X_i) = \sum_{i=1}^{n} Y_i \sim G(n,1)$$

Ex. $X_i \sim Exp(\lambda)$

large n: $\frac{\hat{\theta}_n - \theta}{\sigma_{\hat{\theta}}} \overset{approx}{\sim} N(0,1)$ for large n

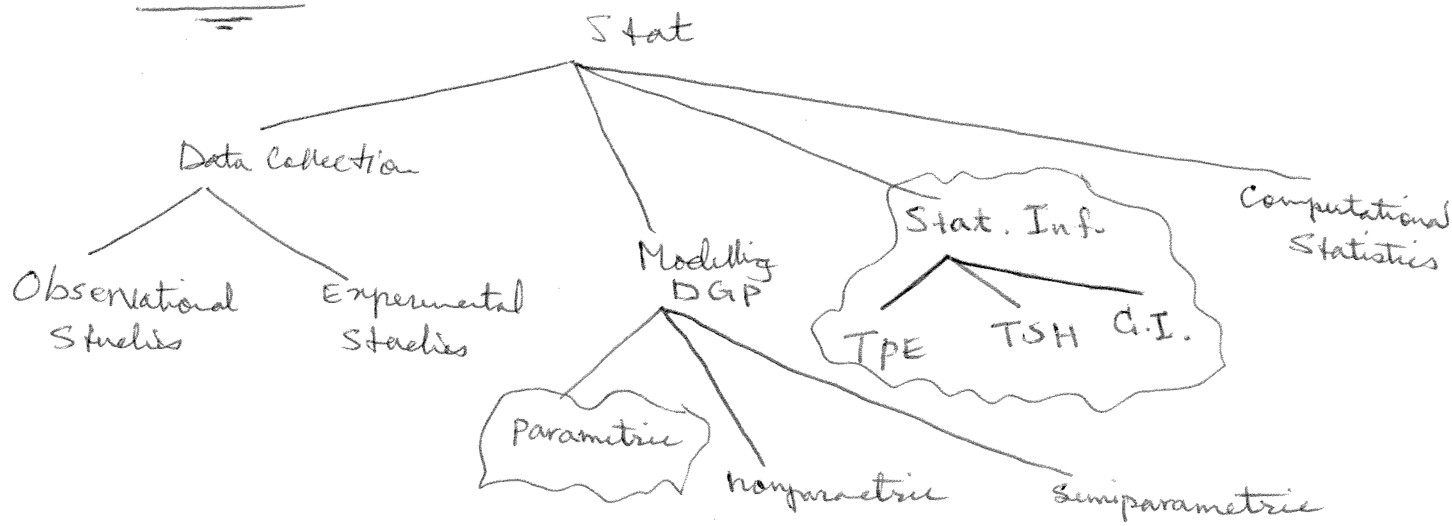
Ex. $X_i \sim N(\mu, 1)$

• Sample Size determination: Use the notes for 203

Statistics (Math 324)
Winter 2018

→ First Lecture: Jan. 9th, 2018 (Tuesday)

- Overview



- In Math 324 we cover Statistical Inference (TPE, TSH and C.I) for parametric models. Note:
  - TPE: Theory of point estimation
  - TSH: Testing Statistical Hypotheses
  - C.I.: Confidence Intervals
  - DGP: Data generating Process
- parametric models: Model is known up to finitely many unknown parameters
  - e.g $X_i \overset{iid}{\underset{i=1,2,\ldots,n}{\sim}} N(\mu, \sigma^2)$ where μ & σ^2 are unknown.
  - iid: Independent and identically distributed
  - "$\sim$": Distributed according to

- Nonparametric models: $X_i \overset{iid}{\sim} F_X(x)$ where the cdf F is completely unknown, but we may assume that F is smooth, for instance continuous or differentiable.

In the nonparametric setting $F(x)$ should be estimated for every x. Thus for a r.v. X that can assume infinitely many values, we need to estimate $F(x)$ at infinitely values of x. This is, particularly, the case when X is a continuous r.v. Recall that $F_X(x) = P(X \leq x)$. Then the sample counterpart of $F_X(x)$ is $\dfrac{\# X_i \leq x}{n}$ for a sample $X_1, \ldots, X_n$. Define

$$\mathcal{E}(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{, o.w.} \end{cases}$$

Then $$\frac{\# X_i \leq x}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{E}(x - X_i).$$

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{E}(x - X_i)$$

is the Empirical Cumulative Distribution Function (ECDF)

- Point Estimation

Suppose $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$ where

$$N(\mu, 1): \quad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}, \quad x \in \mathbb{R}, \ \mu \in \mathbb{R}$$

We want to have an estimate of μ; i.e. a scientific guess, based on the observations, $X_1, \ldots, X_n$. Recall that $E(X_i) = \mu$, $i = 1, 2, \ldots, n$ (μ is the population mean).

- What is an "estimate"?

  Statistic: A function of observations that does not depend on any unknown parameter.

Estimator: An estimator is a statistic that aims at estimating a function of the population unknown parameters

Example: $X_i \overset{iid}{\sim} N(\mu, 1)$

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a Statistic and as estimator of μ

$(\bar{X}_n - \mu)$ is NOT a statistic since it depends on μ, an unknown parameter.

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ is a Statistic, but not an estimator of μ. Note that $\dim(S^2) = (\dim \mu)^2$. For instance, if X_i's are returns of a fund and measured in dollars (\$), then dim of μ is \$ while the dim of S^2 is \$^{2}. Besides μ can be negative while S^2 is always ≥ 0.

Estimation Error:

Going back to our example
$$X_i \overset{iid}{\sim} N(\mu, 1) \quad i = 1, 2, -, n$$
and choosing $\bar{X}_n$ as the estimator of μ. We often want to study $\varepsilon = |\bar{X}_n - \mu|$ or a function of ε. Starting with ε itself, the first thing that comes to mind is $P(\varepsilon \geq \delta)$ for a prespecified δ. Or perhaps $E(\varepsilon)$. A well known tool for studying the former is Tchbyshev's inequality

Statistics (Math 324)

→ 2nd Lecture: Thursday Jan. 11th, 2018

- Tchbyshv's Inequality

Tchbyshev's inequality is a special case of Markov's Inequality.

Markov's Inequality:

Let X be a r.v. and h a nonnegative function, i.e. $h: \mathbb{R} \to \mathbb{R}^+ \cup \{0\} = [0, +\infty)$. Suppose $E(h(X)) < \infty$. Then for any $\lambda > 0$, we have

$$P(h(X) \geq \lambda) \leq \frac{E[h(X)]}{\lambda} \qquad (1)$$

Proof: Suppose X is a continuous r.v.

$$E[h(X)] = \int_x h(x) f_X(x)\,dx$$

$$= \left(\int_{x:\, h(x) \geq \lambda} + \int_{x:\, h(x) < \lambda} \right) h(x) f_X(x)\,dx$$

Given that $h \geq 0$

$$\geq \int_{x:\, h(x) \geq \lambda} h(x) f_X(x)\,dx$$

$$\geq \lambda \int_{x:\, h(x) \geq \lambda} f_X(x)\,dx = \lambda\, P(h(X) \geq \lambda)$$

Thus

$$P(h(X) \geq \lambda) \leq \frac{E[h(X)]}{\lambda}. \qquad \blacksquare$$

The proof for the discrete case is similar.

Now consider $h(x) = (x-\mu)^2$. Then

$$P(|X-\mu| \geq \lambda) = P((X-\mu)^2 \geq \lambda^2)$$
$$\leq \frac{E[(X-\mu)^2]}{\lambda^2} \quad \text{if } E[(X-\mu)^2] < \infty$$

Let $\mu = E(X)$. Then $E[(X-\mu)^2] = Var(X)$, denoted by σ_X^2. We therefore have

$$P(|X-\mu_X| \geq \lambda) \leq \frac{\sigma_X^2}{\lambda^2} \quad (2)$$

where $\mu_X = E(X)$. Now consider $\lambda = K\sigma_X$ where K is a known number. Then

$$P(|X-\mu_X| \geq K\sigma_X) \leq \frac{\cancel{\sigma_X^2}}{K^2\cancel{\sigma_X^2}} = \frac{1}{K^2} \quad (3)$$

This is called <u>Tchbyshev's Inequality</u>. Suppose $K=3$, then

$$P(|X-\mu_X| \geq 3\sigma_X) \leq \frac{1}{9}$$

In other words, at least 88% of the observations are within 3 standard deviation from the population mean

- Going back to our example:

$$X_i \overset{iid}{\sim} N(\mu, 1), \quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

We want to study $P(\varepsilon \geq \delta) = P(|\bar{X}_n - \mu| \geq \delta)$

First we note that $E(X_i) = \mu$, $i = 1, 2, \ldots, n$. Then

$$E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{\cancel{n}} \cdot (\cancel{n}\mu) = \mu \quad (\dagger)$$

Thus using (2)

$$P(|\bar{X}_n - \mu| \geq \delta) \leq \frac{Var(\bar{X}_n)}{\delta^2}$$

$$Var(\bar{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)$$

Using Thm 5.12 (b) on page 271

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n} Var(X_i) + 2\sum_{1\leq i<j\leq n}\sum Cov(X_i, X_j)\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \quad \text{since } \coprod_{i=1}^{n} X_i$$

$$= \frac{1}{n^{\cancel{2}}}\cancel{n} Var(X) = \frac{Var(X)}{n} \quad \text{since } X_i\text{s are identically distributed}$$

$$= \frac{\sigma_X^2}{n} \quad (\ddagger)$$

In our case $X \sim N(\mu, 1)$ so $Var(X) = \sigma_X^2 = 1$. Thus

$$Var(\bar{X}_n) = \frac{1}{n}$$

- Remark: $X \perp\!\!\!\perp Y \Rightarrow Cov(X, Y) = 0$. Note that

$X \perp\!\!\!\perp Y \Rightarrow E[g_1(X)\, g_2(Y)] = E[g_1(X)]\, E[g_2(Y)]$. In particular

$X \perp\!\!\!\perp Y \Rightarrow E[XY] = E[X]\, E(Y)$.

On the other hand,

$$Cov(X, Y) = E[XY] - E(X)E(Y)$$

Thus $X \perp\!\!\!\perp Y \Rightarrow Cov(X, Y) = 0$.

Recall that $X \perp\!\!\!\perp Y$ means X and Y are independent, i.e. $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ where $f_{X,Y}$, f_X, f_Y represent, resp., the joint and marginal dists.

We therefore have

$$P(|\bar{X}_n - \mu| \geq \delta) \leq \frac{1}{n\delta^2} \quad (4)$$

- Using (4) and the sample size, n, we can find an upper bound for the proportion of deviations which are greater than a given threshold δ

- We can also use (4) for <u>sample size determination</u>. Suppose δ is given and we want

$$P(|\bar{X}_n - \mu| \geq \delta) \leq \beta$$

where β is also given. Then setting $\frac{1}{n\delta^2} = \beta$ we can estimate $n \simeq \frac{1}{\beta\delta^2}$. In fact $n \geq \frac{1}{\beta\delta^2}$

– Application to voting

Define $X_i = \begin{cases} 1 & \text{NDP} \\ 0 & \text{o.w.} \end{cases}$. Associated to each eligible voter in Canada we have a binary variable X. Let $p = P(X=1)$. So p represents the proportion of eligible voter who favor NDP. Of interest is often estimation of p. Suppose we have a sample of size n, $X_1, \ldots, X_n$.

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample proportion; the counterpart of p which may be denoted by $\hat{p}_n$. Note that

$\mu_X = E(X) = 1 \cdot P(X=1) + 0 \cdot P(X=0) = 1 \cdot p + 0 \cdot (1-p) = p$

and $E(X^2) = 1^2 \cdot P(X=1) + 0^2 \cdot P(X=0) = 1 \cdot p + 0 \cdot (1-p) = p$

$\sigma_X^2 = Var(X) = E(X^2) - [E(X)]^2 = p - [p]^2 = p - p^2 = p(1-p)$

From (†) and (‡) we find that

$E(\hat{p}_n) = E(\bar{X}_n) = \mu_X = p$

$Var(\hat{p}_n) = Var(\bar{X}_n) = \frac{Var(X)}{n} = \frac{\sigma_X^2}{n} = \frac{p(1-p)}{n}$

Thus using 2 we have

$$P(|\hat{p}_n - p| \geq \delta) \leq \frac{Var(\hat{p}_n)}{\delta^2} = \frac{p(1-p)}{n\delta^2}$$

Note that the above bound on the probab. of deviation by δ depends on p which is <u>unknown</u>. We, however, notice that

$$p(1-p) \leq \frac{1}{4}$$

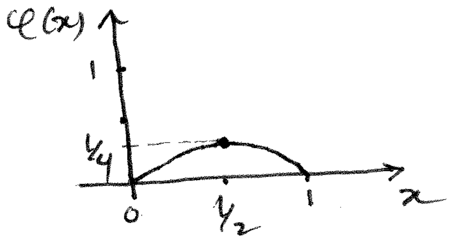
Define $\varphi(x) = x(1-x)$ for $0 < x < 1$. Then

$$\varphi'(x) = 1 - 2x \Rightarrow \varphi'(x) = 0 \Rightarrow x = \frac{1}{2}$$

$$\varphi''(\tfrac{1}{2}) = -2 \Rightarrow x = \tfrac{1}{2} \text{ is a } \underline{\text{maximizer}}$$

$$\varphi(\tfrac{1}{2}) = \tfrac{1}{2}(1 - \tfrac{1}{2}) = \tfrac{1}{4}$$

(Note that $\varphi''(x) = -2$ for all $0 < x < 1$.)



We therefore find

$$P(|\hat{p}_n - p| \geq \delta) \leq \frac{1}{4n\delta^2} \quad (5)$$

- Using 5 and a given sample size n we can find an upper bound for the probab. of deviation by δ amount for any given δ.

- We can also use 5 for <u>sample size determination</u> for a given bound β and deviation δ as follows

$$\frac{1}{4n\delta^2} = \beta \Rightarrow n \geq \frac{1}{4\beta\delta^2}$$. This is, of course, conservative since $p(1-p) \leq \frac{1}{4}$.

(D)