IES302

Engineering Statistics : Lecture 2 - 4 (Probability & Distribution)

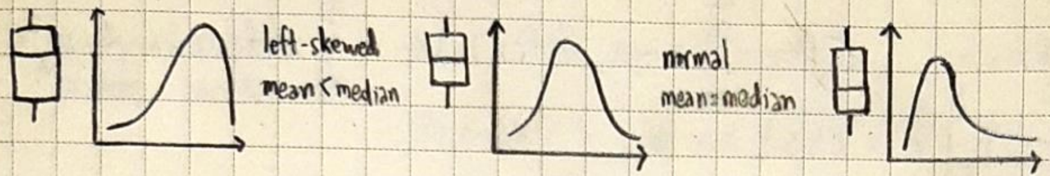
Basics

sample mean $(\bar{x}) = \frac{\sum x_i}{n}$ sample varience $(s^2) = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$

if y_i related to x_i as $y_i = ax_i + b$; $\bar{y} = a\bar{x} + b$, $s_y^2 = a^2 s_x^2$

median = center data point of data set = $Q_2 = P_{50}$ (if position is decimal, take average of 2 nearest)

$IQR = Q_3 - Q_1$ Outlier = $x_i < Q_1 - 1.5IQR$ or $x_i > Q_3 + 1.5IQR$



probability $P(E) = \frac{N(E)}{N(S)}$ conditional probability $P(A|B) = \frac{P(A \cap B)}{P(B)}$ Bayes's theorem $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Independence :

$P(A|B) = P(A)$ & $P(B|A) = P(B)$

$P(A \cap B) = P(A)P(B)$

Probability properties

$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots$

$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

Counting Method

selection of k item out of n items

	w/ repetitions	w/o repetitions
ordered	Tuples n^k	k-perm $_nP_k$
unordered	comb w/ rep $\binom{k+n-1}{n-1}$	Comb. $_nC_k$

permutation $_nP_k = \frac{n!}{(n-k)!}$

combination $_nC_k = \binom{n}{k} = \frac{n!}{(n-k)!\,k!}$

Random variable

discrete random variable comes from a discrete set

- probability mass function (pmf) / probability distribution $p(x) = P(X = x)$
- cumulative distribution function (cmf) $F(x) = P(X \leq x)$

$$F(x) = \sum_{t \leq x} p(t) = \sum_{t \leq x} P(X = x)$$; cumulate probability less than or equal to x

$$\therefore \sum_x p(x) = 1$$

$\mu_x = \sum x p(x)$ $\sigma_x^2 = \sum x^2 p(x) - \mu_x^2 = \sum (x - \mu_x)^2 p(x)$

↳ mean / expected value $E(x)$ ↳ population varience $V(x)$

continuous random variable presented as histogram with very small interval ($\rightarrow +\infty$)

- probability density function (pdf) as $f(x)$

$$P(X \leq b) = P(X < b) = \int_{-\infty}^{b} f(x)\,dx \;,\; P(X \geq a) = P(X > a) = \int_{a}^{\infty} f(x)\,dx$$

• probability density function (continue)

$$P(a \le X \le b) = P(a < X < b) = \int_a^b f(x)\,dx \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

• cumulative distribution function of continuous random variable

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$; t as intermediate variable

population
• mean & varience of continuous random variable

$$\mu_x = \int_{-\infty}^{\infty} x f(x)\,dx \qquad \sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)\,dx = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu_x^2$$

• median & percentile

$$F(x_m) = P(X \le x_m) = \int_{-\infty}^{x_m} f(t)\,dt = 0.5$$; solve this equation x_m as median

or $$F(x_p) = P(X \le x_p) = \int_{-\infty}^{x_p} f(t)\,dt = p/100$$; as solve for percentile x_p as percentile p-th

• Chebyshev's inequality

$$P(|x-\mu_x| \geq k\sigma_x) \le \frac{1}{k^2}$$

; providing a bound on the probability that random variable takes on value differ from its mean by more than a given multiple of its standard deviation

• probability that random variable differs from means by k stdev. or more is never greater than $1/k^2$

• should be used only when distribution of random variable is unknown b/c it is generally much larger cbound of chebyshev

■ Linear function of random variable

$$\mu_{ax+b} = a\mu_x + b, \quad \sigma_{ax+b}^2 = a^2\sigma_x^2, \quad \sigma_{ax+b} = |a|\sigma_x$$

$$\mu_{ax+by} = \mu_{ax} + \mu_{by} = a\mu_x + b\mu_y$$; true even more than 2

• independent •

$$\sigma^2_{c_1x_1+c_2x_2+\dots} = c_1^2\sigma_{x_1}^2 + c_2^2\sigma_{x_2}^2 + \dots \quad ; \quad \sigma^2_{X\pm Y} = \sigma_X^2 + \sigma_Y^2$$

■ independent random variable

$$P(X \in S \text{ and } Y \in T) = P(X \in S)\,P(Y \in T)$$

■ sample mean & varience.

$$\mu_{\bar{X}} = \mu \qquad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

b/c treated as linear combination $\bar{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 \cdots + \frac{1}{n}X_n$

$X_1 \dots X_n$ as SRS.

$$\sigma_{\bar{X}} = \frac{\sigma_x}{\sqrt{n}}$$

Jointly distributed Random variable

- two or more random variable is associated with each item in population they are jointly distributed. → jointly discrete / → jointly continuous

• jointly discrete random variable

$p(x,y) = P(X=x \text{ and } Y=y)$; joint pmf of X and Y

→ marginal PMF of X and Y

$p_X(x) = P(X=x) = \sum_y p(x,y)$, $p_Y(y) = P(Y=y) = \sum_x p(x,y)$

$\sum_x \sum_y p(x,y) = 1$

• jointly continuous random variable

joint pdf as $f(x,y)$, $f(x,y) \geq 0$ for all x & y

$P(a \leq X \leq b \text{ and } c \leq Y \leq d) = \int_a^b \int_c^d f(x,y)\,dy\,dx$

$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 1$

→ marginal pdf of X and Y

$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

• mean function

X as random variable, $h(x)$ is function of X

→ discrete $\mu_{h(X)} = \sum_x h(x)p(x)$

→ continuous $\mu_{h(X)} = \int_{-\infty}^{\infty} h(x)f(x)\,dx$

μ_{XY} :→ $\mu_{h(X,Y)} = \sum_x \sum_y h(x,y)\,p(x,y)$; jointly discrete.

→ $\mu_{h(X,Y)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) f(x,y)$, jointly continuous

• conditional distribution

$P(Y=y \mid X=x) = \frac{P(X=x \text{ and } Y=y)}{P(X=x)} = \frac{p(x,y)}{p_X(x)}$; conditional pmf $p_{Y|X}(y|x)$

$f_{Y|X}(Y|X) = \frac{f(x,y)}{f_X(x)}$; conditional pdf

- conditional expectation → mean calculated by conditional pmf/pdf

$E(Y \mid X=x) = \mu_{Y|X=x} = \sum_y y\,p_{Y|X}(y|x)$; jointly discrete

$= \int_{-\infty}^{\infty} y\,f_{Y|X}(y|x)$; jointly continuous

- Independent random variable

X and Y are jointly discrete, pmf
→ $p(x,y) = p_X(x)\, p_Y(y)$

X and Y are jointly continuous, pdf
→ $f(x,y) = f_X(x)\, f_Y(y)$

→ if X and Y are independent random variable

→ jointly discrete $p_{Y|X}(y|x) = p_Y(y)$ → jointly continuous $f_{Y|X}(y|x) = f_Y(y)$

- Covarience

→ two random variable are not independent, covarience is measurement of linear relationship by me. as STRENGTH, unit : XY

$$Cov(X,Y) = \mu_{(X-\mu_X)(Y-\mu_Y)} = \mu_{XY} - \mu_X \mu_Y$$

- Correlation

→ measurement of linear relationship, unitless

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \le \rho_{X,Y} \le 1$$

$X\uparrow Y\uparrow$

$X\uparrow Y\downarrow$

→ If $Cov(X,Y) = \rho_{X,Y} = 0$, then X & Y are uncorrelated

→ X & Y are independent, then X & Y are uncorrelated

☆ if X & Y are random variable

$$\sigma^2_{X+Y} = \sigma_X^2 + \sigma_Y^2 + 2\,Cov(X,Y)$$

$$\sigma^2_{X-Y} = \sigma_X^2 + \sigma_Y^2 - 2\,Cov(X,Y)$$

- Uncertainties / propagation of error

$$\sigma_U = \left|\frac{dU}{dX}\right| \sigma_X$$ ☆

relative uncertainty $\sigma_{\ln U} = \frac{\sigma_U}{U}$

- Distribution

1) Bernoulli Distribution

- only 2 outcomes in experiment
  one is success, another is failure
  this is called bernoulli trial
  → success as p → $X=1$
  → failure as $1-p$ → $X=0$

denoted as $X \sim Bernoulli(p)$

- $\begin{cases} p(x) = 0 & ; x \text{ is not } 0,1 \\ p(1) = P(X=1) = p \\ p(0) = P(X=0) = 1-p \end{cases}$

- $\mu_X = p$

- $\sigma_X^2 = p(1-p)$

☆ think like discrete random variable

2) Binomial Distribution

- n bernouli trial conducted
- trial are independent
- each trial has same probability
- X is number of success in n trial

denoted as $X \sim Bin(n,p)$

* sample size is no more than 5% of population

- $p(x) = P(X=x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & ; x = 0,1,\dots n \\ 0 & ; \text{otherwise} \end{cases}$
- $\mu_X = np$
- $\sigma_X^2 = np(1-p)$

→ sample propotion to estimate success probability $(\hat{p})$

$$\hat{p} = \frac{\text{number of success}}{\text{number of trial}} = \frac{X}{n}$$

$\hat{p}$ is unbiased.

uncertainty ; $\sigma_{\hat{p}} = \frac{\sigma_X}{n} = \sqrt{\frac{p(1-p)}{n}}$

3) Poisson Distribution

→ approx of binomial distribution when n is large, p is small

- approx. binomial mass function with $\lambda = np$; λ is parameter in poisson distribution

denoted as $X \sim Poisson(\lambda)$

* X is discrete, possible value are $\mathbb{Z}^+$
* λ is positive constant

- $p(x) = P(X=x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & ; x = 0,1,\dots \\ 0 & ; \text{other wise} \end{cases}$
- $\mu_X = \lambda$
- $\sigma_X^2 = \lambda$

- Estimate rate by Poisson distribution

→ λ denote the mean number of events that occurs in one unit of time
X denote the number of events that are occurs in t unit of time

then, if $X \sim Poisson(\lambda t)$

λ is estimated as $\hat{\lambda} = \frac{X}{t}$, $\hat{\lambda}$ is unbiased

uncertainty in $\hat{\lambda}$ is $\sigma_{\hat{\lambda}} = \sqrt{\frac{\lambda}{t}}$, in practice we substitute $\hat{\lambda}$ in λ in equation since λ is unknown

4) Hypergeometric Distribution

→ each item sampled constitute a bernouli trial

→ each item selected, probability of success in population is increase or decrease

* trials are not independent *

denoted as $X \sim H(N,R,n)$

- population N, R success, N-R failure, n sampled

pmf of X

$$p(x) = P(X=x) = \begin{cases} \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} & ; \max(0, R+n-N) \le x \le \min(n,R) \\ 0 & ; \text{otherwise} \end{cases}$$

- $\mu_X = \frac{nR}{N}$
- $\sigma_X^2 = n\left(\frac{R}{N}\right)\left(1 - \frac{R}{N}\right)\left(\frac{N-n}{N-1}\right)$

5) Geometric Distribution

- sequence of independent bernoulli trials is conducted with each same prob success.

denoted as $X \sim Geom(p)$

- pmf

$$p(x) = P(X=x) = \begin{cases} p(1-p)^{x-1} & ; x = 1, 2, \dots \\ 0 & ; \text{otherwise} \end{cases}$$

- $\mu_X = \frac{1}{p}$
- $\sigma_X^2 = \frac{1-p}{p^2}$

6) Negative Binomial Distribution

- extension of geom(p)

denoted as $X \sim NB(r, p)$

r as $\mathbb{Z}^+$, probability p

pmf ; $$p(x) = P(X=x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & ; x = r, r+1, \dots \\ 0 & ; \text{otherwise} \end{cases}$$

- $\mu_X = \frac{r}{p}$
- $\sigma_X^2 = \frac{r(1-p)}{p^2}$

7) Multinomial Distribution

- n trials, k outcomes with same propability $p_1, p_2, p_3, \dots p_k$
- $X_1, \dots X_k$ as discrete RV.

denoted as $X_1, X_2, \dots X_k \sim MN(n, p_1, \dots p_k)$

- pmf

$$p(x_1, \dots x_k) = P(X_1 = x_1, \dots, X_k = x_k)$$

$$= \begin{cases} \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} & ; x_i = 0, 1, \dots n \text{ and } \sum x_i = n \\ 0 & ; \text{otherwise} \end{cases}$$

8) Normal Distribution / Gaussian distrb.

- most common distribution for Continuous population
- any value for μ and σ^2

denoted as $X \sim N(\mu, \sigma^2)$

- no need to remember

↳ pdf $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$; $-\infty < x < \infty$

$\mu_X = \mu$, $\sigma_X^2 = \sigma^2$; given.

- 68 - 95 - 99.7 % Rule

→ 68% of population is in interval of $\mu \pm \sigma$

95% of population is in interval of $\mu \pm 2\sigma$

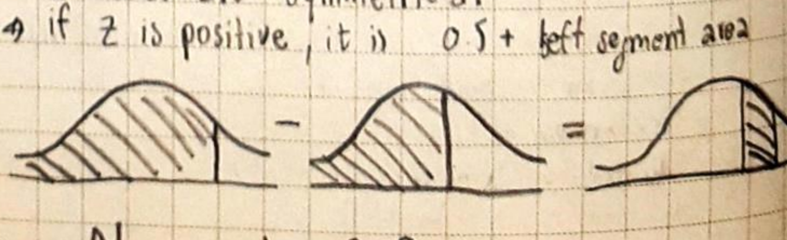
99.7% of population is in interval of $\mu \pm 3\sigma$

- standard unit tells how many stdev an observation is from population mean

↳ z-score , $z = \frac{x - \mu}{\sigma}$

↳ standard normal population

- finding area under normal curve
  - use standard normal table / z table
  - 2-sides are symmetrical
  - if z is positive, it is 0.5 + left segment area



- $X \sim N(\mu, \sigma^2)$ then $aX + b \sim N(a\mu + b, a^2\sigma^2)$

- let $X_1, \dots X_n$ independent & normally distributed with μ and σ^2 then $\bar{X} \sim N(\mu, \sigma^2/n)$

- let X & Y independent then $X+Y \sim N(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$

  $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ $\quad X-Y \sim N(\mu_X-\mu_Y, \sigma_X^2+\sigma_Y^2)$

- lognormal distribution

  $X \sim N(\mu, \sigma^2)$ and $Y = e^X$ then $X = \ln Y \rightarrow N(\mu, \sigma^2)$
  → for data with highly skewed and/or contain outlier

9) Exponential Distribution

- continuous distribution model time elapse before event occurs, waiting time.
- pdf involves λ, determine location & shape
  denoted as $X \sim Exp(\lambda)$
- memory less property
  → not remember how long it been waiting
  → if $T \sim Exp(\lambda)$ then,

  $P(T>t+s \mid T>s) = P(T>t)$

- pdf $f(x) = \begin{cases} \lambda e^{-\lambda x} & ; x>0 \\ 0 & , \text{otherwise} \end{cases}$

- cdf $F(x) = \begin{cases} 0 & ; x \le 0 \\ 1-e^{-\lambda x} & , x>0 \end{cases}$

- $\mu_X = \frac{1}{\lambda}$ $\quad$ • $\sigma_X^2 = \frac{1}{\lambda^2}$

- if events follow poisson process with parameter λ
  if T is the waiting time from any point
  $T \sim Exp(\lambda)$

- Exponential to estimate rate

  - estimator is biased! $\hat{\lambda} = 1/\bar{X}$

    uncertainty $\sigma_{\hat{\lambda}} \approx \frac{1}{\bar{X}\sqrt{n}}$; reasonably good when size more than 20

  - correct the bias
    $\mu_{\hat{\lambda}} = \mu_{\frac{1}{\bar{X}}} \approx \lambda + \lambda/n = (n+1)\lambda/n$

    that $n/[(n+1)\lambda]$ has smaller bias → bias corrected value.

10) Uniform Distribution

- 2 parameters a & b with $a<b$
- uniformly distributed on interval (a,b)
  denoted as $X \sim U(a,b)$

pdf $f(x) = \begin{cases} \frac{1}{b-a} & , a<x<b \\ 0 & , \text{otherwise} \end{cases}$

- $\mu_X = \frac{a+b}{2}$ $\quad$ • $\sigma_X^2 = \frac{(b-a)^2}{12}$

Principle of Point Estimation

- collect data for purpose of estimating some numerical characteristic of population from which they come

Questions of interest

1) Given a point estimator, how do we determine how good it is?
2) what methods can be used to construct good point estimator.

Notation : θ is used to denote an unknown parameter

$\hat{\theta}$ to denote an estimator of θ

- accuracy measure by its bias, precision by stdev or uncertainty

MSE

- overall goodness measured by MSE (mean squared error), combined both bias and uncertainty

$$MSE_{\hat{\theta}} = (\mu_{\hat{\theta}} - \theta)^2 + \sigma_{\hat{\theta}}^2 \equiv MSE_{\hat{\theta}} = \mu_{(\hat{\theta}-\theta)^2}$$

as estimate value - true value squared

//error

Maximum Likelihood (MLE)

- estimate a parameter with value that makes the observed data most likely
- pdf or pmf is considered to be function of parameter
  → called likelihood function

- maximum likelihood estimate is the value of estimators that when substituted in for the parameters maximizes the likelihood f^n

- Properties : 1) most cases, as n increase (sample size) bias of MLE converges to 0

  2) most cases, as n increase (sample size) varience of MLE converges to theoretical minimum

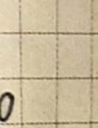
Probability Plots

- ways to determine an appropriate distribution

1) arrange data in increasing / ascending order

2) assign evenly spaced value b/w 0 to 1 to each X_i $\left(\frac{1-0.5}{n}\right)$

distribution should have μ_X and σ_X^2 match the X we want

normal probability plots, closed to the slope line 

Central Limit Theorem (CLT)

let $X_1, \dots, X_n$ be a random sample w/ μ and σ^2

$\bar{X}$ is sample mean, S_n = sum of sample observation

* if n is sufficiently large

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ and } S_n \sim N(n\mu, n\sigma^2)$$

- for most population, $n > 30$, CLT approx is good

Binomial CLT

* if $X \sim Bin(n,p)$ if $np > 10$ and $n(1-p) > 10$ then

$X \sim N(np, np(1-p))$ approx.

and $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ approx.

Poisson CLT

* if $X \sim Poisson(\lambda)$ where $\lambda > 10$ then

$X \sim N(\lambda, \lambda)$ approx.

note : continuity correction, determine precisely which rectangles of the discrete probability histogram you wish to include, then compute the area under the normal curve corresponding to those rectangle it improves accuracy of normal approx. for most cases

IES302
Engineering Statistic (point estimation)

Point Estimation

→ point estimate is a reasonable value of a population parameter.

→ data collected are random variable.

→ Statistics is function of random variables ($\bar{X}$ & S)

→ statistic have their unique distributions that are called sampling distributions

→ definitions

→ point estimation of some population parameter θ is a single numerical value $\hat{\theta}$
$\hat{\theta}$ is point estimator

→ random variables $X_1, X_2 \dots X_n$ are random sample of size n if:
X_i are independent random variables & X_i has the same probability distribution

→ statistic is any function of the observations in a random sample.
→ The probability distribution of a statistic is called sampling distribution

→ two populations

two independent normal population
sampling distribution of $\bar{X}_1 - \bar{X}_2$ is

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2 \qquad \sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

if n_1 & $n_2 \geq 30$
→ distribution of $\bar{X}_1 - \bar{X}_2$ is normal

sampling distribution
↳ 2 independent population
$\bar{X}_1$ & $\bar{X}_2$ sample mean

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

→ general concepts

we want point estimator that "UNBIASED" & "MINIMAL VARIANCE"
→ use standard error of estimator to calculate MSE

$$\text{bias} = E(\hat{\theta}) - \theta$$ ⇒ if $=0$ then unbiased
$\neq 0$ then called biased

MSE ⇒ $E(\hat{\theta} - \theta)^2$
$= E(\hat{\theta} - E(\hat{\theta}))^2 + [\theta - E(\hat{\theta})]^2$
$= V[\hat{\theta}] + \text{biased}^2$ // varience + biased square

$$\text{relative efficiency} = \frac{MSE(\theta_1)}{MSE(\theta_2)}$$ if < 1 → 1st is superior to 2nd

→ Method of point estimation
- Method of moment
- Method of maximum likelihood
- Bayesian estimation of parameter

Method of moments

→ Moment is a kind of expected value of a random variable
→ Population moment relates to the entire population or its representative function
→ Sample moment is calculated like associated population moments

k^{th} population moment $\Rightarrow E(X^k)$, $k = 1, 2, \dots$

k^{th} sample moment $\Rightarrow \frac{1}{n}\Sigma(X^k)$

$k=1$ (called first moment) → population moment μ, sample moment $\bar{X}$

→ exponential estimator. $E(X) = \bar{X} \Rightarrow E(X) = \frac{1}{\lambda} = \bar{X} \rightarrow \lambda = \frac{1}{\bar{X}}$ is moment estimator

→ normal moment estimator
$E(X) = \mu$
$E(X^2) = \mu^2 + \sigma^2$
$\mu = \frac{1}{n}\Sigma x_i$ $\quad \mu^2 + \sigma^2 = \frac{1}{n}\Sigma x_i^2$
$\rightarrow \sigma^2 = \frac{\Sigma(x_i - \bar{X})^2}{n}$ (biased)

Method of MLE

θ is single unknown parameter
likelihood function : $L(\theta) = f(x_1 : \theta)\, f(x_2 : \theta) \cdots f(x_n : \theta)$

maximum likelihood estimator of θ is the value of θ that maximize the likelihood function $L(\theta)$

→ Bernouli MLE $\quad f(x : p) = p^x (1-p)^{1-x} \rightarrow \hat{p} = \frac{\Sigma x_i}{n} = \bar{X}$

→ Normal MLE $\quad X$ random var w/ unknown μ and known σ^2
when both unknown
$\left(\sigma^2 = \frac{\Sigma(x_i - \bar{X})^2}{n}\right)$ $\quad \rightarrow \mu = \frac{\Sigma x_i}{n} = \bar{X}$

→ Exponential MLE $\quad \lambda = \frac{n}{\Sigma x_i} = \frac{1}{\bar{X}}$

→ steps for gain estimator
① set $f(x : \theta)$
② simplify
③ take $\ln$ to take exponent down
④ diff equation once
⑤ viola!

→ Properties of MLE $\quad$ when sample n is large if θ is MLE of parameter

1) θ is approx unbiased estimator
2) varience of $\hat{\theta}$ is nearly small as variance obtained w/ other estimator
3) $\hat{\theta}$ approx normal distribution

ex normal MLE when both unknown

$$E(\sigma^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-\sigma^2}{n}$$ → underestimate σ^2

→ Complication of MLE

1) not easy to maximize likelihood due to derivatives
2) maybe impossible to solve

■ Bayesian Estimation

Moment & Likelihood ⇒ objective frequency

Bayesian method combine sample info w/ prior information

$f(\theta)$ is called prior distribution w/ μ_0 & σ_0^2

$f(x_1, x_2, x_3, \cdots x_n, \theta)$ is the joint distribution $\begin{cases} \sum_{\theta} f(x_1, \ldots \theta) & \text{for discrete} \\ \int_{-\infty}^{\infty} f(x_1, \ldots \theta) & \text{for continuous} \end{cases}$

$f(\theta \mid x_1, x_2, \ldots, x_n)$ is the posterior distribution

thus, $f(\theta \mid x_1, x_2, \ldots x_n) = \dfrac{f(x_1, x_2, \ldots x_n, \theta)}{f(x_1, x_2, \ldots x_n)}$

and $f(x_1, \ldots x_n, \theta) = f(x_1, x_2, \ldots x_n \mid \theta) f(\theta)$

→ Normal Bayesian

$$E(\mu) = \mu = \frac{\left(\frac{\sigma^2}{n}\right)\mu_0 + \sigma_0^2 \bar{x}}{\sigma_0^2 + \frac{\sigma^2}{n}}$$

→ estimation is weight average of μ_0 and $\bar{x}$
→ $\bar{x}$ is MLE of μ
→ important of μ_0 decrease as n increase

$$V(\mu) = \frac{\sigma_0^2 \left(\frac{\sigma^2}{n}\right)}{\sigma_0^2 + \frac{\sigma^2}{n}}$$

note that $\mu_0, \sigma_0^2 \Rightarrow$ prior

$\frac{\sigma^2}{n}, \bar{x} \Rightarrow$ sample