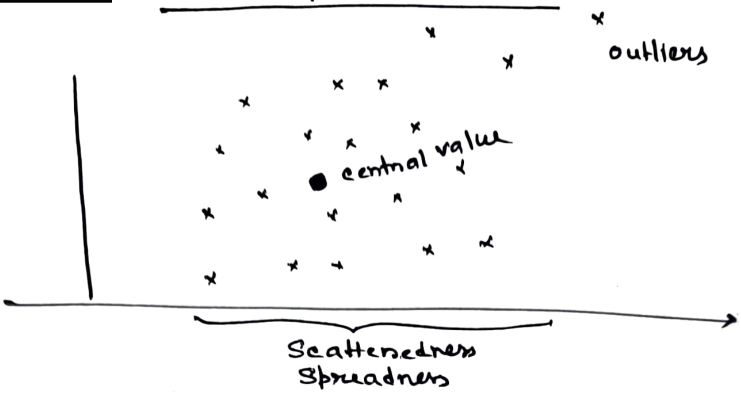
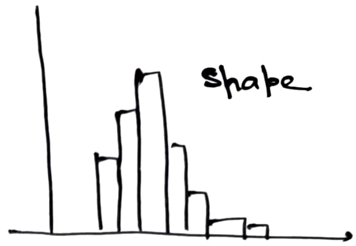
CHAPTER 0: Descriptive Statistics





Central Tendency : Mean (arithmatic, geometric, harmonic, trimmed)
Median
Model

Dispersion : Range, Standard Deviation, QD

Shape : Skewness, Kurtosis

$$AM = \frac{x_1 + \cdots + x_n}{n} \geq GM = \sqrt[n]{x_1 \cdots x_n} \geq \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = HM$$

(Proof!)

Solve (for x, y, z):

$$\frac{x+y+z}{3} = 4$$
$$xyz = 60$$
$$xy + yz + zx = 47$$

Hint: $(\sqrt{x} - \sqrt{y})^2 \geq 0$
$\Rightarrow x + y - 2\sqrt{xy} \geq 0$
$\Rightarrow \frac{x+y}{2} \geq \sqrt{xy}$

Some useful Inequalities:
- Jensen's Inequality
- Cauchy-Schwarz Inequality
- Markov & Chebyshev's Inequality.

Markov Inequality: X : a random variable, $a > 0$, $\mathbb{E}(X)$: Expectation

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Proof: $Y = \begin{cases} a, & X \geq a \\ 0, & \text{ow} \end{cases}$

$$\mathbb{E}(Y) = a \cdot P(X \geq a) + 0 \cdot P(X < a) = aP(X \geq a)$$

$X \geq Y$
$\Leftrightarrow \mathbb{E}(X) \geq \mathbb{E}(Y)$
$\Leftrightarrow P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$

$\mathbb{E}(X) = \mu < \infty$
$\mathbb{V}(X) = \mathbb{E}(X - \mu)^2 > 0.$

Take $Z = (X - \mu)^2$, $a = t^2\sigma^2$

$$\mathbb{P}(Z \geq a) \leq \frac{\mathbb{E}(Z)}{a}$$

$$\Leftrightarrow \mathbb{P}\left((X-\mu)^2 \geq t^2\sigma^2\right) \leq \frac{\mathbb{E}(X-\mu)^2}{t^2\sigma^2} = \sigma^2/t^2\sigma^2 = \frac{1}{t^2}$$

$$\Leftrightarrow \mathbb{P}(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2}$$: Chebyshev's Inequality

- If $Y \sim DF(\mu, \sigma^2)$ then $P[Y \in (\mu + 3\sigma)] \geq \frac{8}{9} \simeq 89\%$. (Six-Sigma Methodology)

For $Y \sim$ Normal
$P[Y \in (\mu \pm 3\sigma)] \doteq 0.9973 = 2\Phi(3) - 1$

Gamma Function:

$$\Gamma(n) = \int_0^\infty e^{-x} x^{n-1}\, dx, \qquad x > 0.$$

$$\Gamma(n) = (n-1)!$$

Using the PDF of standard normal distn., show that $\Gamma(1/2) = \sqrt{\pi}$.

$\rightarrow$ $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$; then

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1 \Rightarrow \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1/2$$

$$\Rightarrow \int_0^\infty e^{-x^2/2}\, dx = \sqrt{\frac{\pi}{2}}$$

Take $x^2/2 = u$

$x^2 = 2u$

$dx = \frac{du}{\sqrt{2u}}$

$$\Rightarrow \frac{1}{\sqrt{2}} \int_0^\infty e^{-u} u^{-1/2}\, du = \sqrt{\frac{\pi}{2}}$$

$$\Rightarrow \int_0^\infty e^{-u} u^{-1/2}\, du = \sqrt{\pi}$$

$$\Rightarrow \boxed{\Gamma(1/2) = \sqrt{\pi}}$$

Proof of theorem on Sampling Distribution:-

S.T. $\mathbb{E}(\bar{X}) = \mu$

$\mathbb{V}(\bar{X}) = \sigma^2/n$.

Pf:

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\left(n\,\mathbb{E}(X_1)\right)$$

$$= \frac{1}{n}\cdot n\mu$$

$$= \mu.$$

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

$$= \frac{1}{n^{\cancel{2}}}\cdot \cancel{n}\,\mathrm{Var}(X_1)$$

$$= \frac{\sigma^2}{n}.$$

Sampling Distribution:

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

IID: independently and identically distributed

Notation: $\bar{x}$: sample mean $\quad s^2$: sample variance

μ : population mean $\quad \sigma^2$: population variance

CDF: cumulative distr. function

PDF: Probability density function

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1); \ \mu = 0, \ \sigma^2 = 1$$

CDF of $Y = Z^2$: $F(y) = P[Y \le y] = P[Z^2 \le y]$

$$= P[-\sqrt{y} \le Z \le \sqrt{y}]$$

$$= 2P[0 \le Z \le \sqrt{y}]; \text{ since } Z \text{ is symmetric with } f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}; z \in \mathbb{R}$$

$$= 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz$$

$z^2 = u$

$z = \sqrt{u}$

$dz = \frac{1}{2} u^{-1/2} du$

z	0	$\sqrt{y}$
u	0	y

$$P[Y \le y] = \frac{2}{2} \int_0^{y} \frac{1}{\sqrt{2\pi}} e^{-u/2} u^{1/2 - 1} du$$

$$\frac{d}{dy} F(y) = \frac{1}{\sqrt{2\pi}} e^{-y/2} y^{1/2-1} dy$$

$$\therefore \boxed{f(y) = \frac{1}{2^{1/2}\Gamma(1/2)} e^{-y/2} y^{1/2-1}}$$

PDF

$y > 0 \quad$ [since $\Gamma(1/2) = \sqrt{\pi}$]

[Newton-Leibnitz formula for integral:

$$\frac{d}{dx}\int_{h(x)}^{g(x)} f(t)\,dt = f(g(x)) \times g'(x) - f(h(x))\,h'(x)]$$

$\therefore Y \sim \chi_1^2$: Chi-squared distribution with degree of freedom=1.

Thus, if $Y_1, \ldots, Y_n \overset{IID}{\sim} \chi_1^2$

Define $V = \sum_{i=1}^{n} Y_i \overset{IID}{\sim} \chi_n^2$

$$\Rightarrow \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi_n^2 \ (\text{How?})$$

NOTE: MGF uniquely defines the distribution of a random variable (proof!)

Moment Generating Function (MGF) Technique:

$\underset{\sim}{X} = (X_1, \ldots, X_n)^T$ be a random vector and $\underset{\sim}{t} = (t_1, \ldots, t_n)^T \in \mathbb{R}$: then the moment generating function is defined by

$$M_{\underset{\sim}{X}}(\underset{\sim}{t}) = \mathbb{E}\left(e^{\underset{\sim}{t}^T \underset{\sim}{X}}\right) = \mathbb{E}\left(e^{\sum_{i=1}^{n} t_i x_i}\right)$$

for all t for which the expectation exists (i.e., finite).

$$M_{\underset{\sim}{X}}(0) = 1.$$

$\mu_n' = \mathbb{E}(X^n)$

$\mu_1' = \mathbb{E}(X)$: mean

$\mu_2' = \mathbb{E}(X^2)$

$\text{Var}(X) = V(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$

$= \mu_2' - \mu_1'$

$\mu_0' = 1.$

Expanding we get

$$M_X(t) = 1 + t\,\mathbb{E}(X) + \frac{t^2}{2!}\mathbb{E}(X^2) + \frac{t^3}{3!}\mathbb{E}(X^3) + \cdots$$
$$= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \cdots = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mu_k'$$

For a r.v. X, we can obtain the k^{th} moment of X:

$$\boxed{\left.\frac{d^k}{dt^k} M_X(t)\right|_{t=0} = \mathbb{E}(X^k)}$$

Example:

$X \sim N(0,1)$

$$M_X(t) = \mathbb{E}(e^{tX})$$
$$= \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 - 2tx)}\,dx$$
$$= e^{t^2/2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2}\,dx}_{N(t,1)\text{ PDF}}$$
$$= e^{t^2/2} \quad \forall t \in \mathbb{R}.$$

Mean and Variance of χ_n^2:

$$M_{\chi_1^2}(t) = \mathbb{E}(e^{tY_1})$$
$$= \int_0^{\infty} e^{ty} f(y)\,dy$$
$$= \int_0^{\infty} e^{ty} \frac{1}{\sqrt{2\pi}} e^{-y/2} y^{1/2-1}\,dy$$
$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y/2(1-2t)} y^{1/2-1}\,dy$$

$y(1-2t) = k$

$dy = \frac{dk}{(1-2t)}$

$y = \frac{k}{(1-2t)} \Rightarrow y^{-1/2} = \frac{k^{-1/2}}{(1-2t)^{-1/2}}$

$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-k/2} \cdot \frac{k^{-1/2}}{(1-2t)^{-1/2}} \cdot \frac{dk}{(1-2t)}$$
$$= \frac{1}{(1-2t)^{1/2}} \underbrace{\int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-k/2} k^{1/2-1}\,dk}_{\text{PDF of } \chi_1^2}$$
$$= (1-2t)^{-1/2}$$

$Y_1, \ldots, Y_n \sim \chi_1^2 \Rightarrow \sum Y_i \sim \chi_n^2$

$$\mathbb{E}\left(e^{tY_1 + tY_2 + \cdots + tY_n}\right) = \mathbb{E}(e^{tY_1})\,\mathbb{E}(e^{tY_2}) \cdots \mathbb{E}(e^{tY_n}) \Rightarrow M_{\sum Y_i}(t) = \prod_{i=1}^{n} M_{Y_i}(t)$$

$$\boxed{\text{MGF of } \chi_n^2 = (1-2t)^{-n/2}}$$

$$\text{Mean} = \left.\frac{d}{dt} M_{\chi_n^2}(t)\right|_{t=0}$$
$$= \left.\frac{d}{dt}\left[(1-2t)^{-n/2}\right]\right|_{t=0} = \left.-2\left(-\frac{n}{2}\right)(1-2t)^{-\frac{n}{2}-1}\right|_{t=0}$$
$$= n.$$

$$\mathbb{E}(\chi_n^2) = \frac{d}{dt}\left[M_{\chi_n^2}(t)\right]_{t=0} = n = \mu_1'$$

$$\mathbb{E}(\chi_n^2)^2 = \frac{d^2}{dt^2}\left[M_{\chi_n^2}(t)\right]_{t=0} = \mu_2' = 2n\left(\frac{n}{2}+1\right) = \frac{d^2}{dt^2}\left[(1-2t)^{-n/2}\right]_{t=0}$$

$$= n(-2)\left(-\frac{n}{2}-1\right)(1-2t)^{-n/2-1-1}\Big|_{t=0}$$

$$= n^2+2n$$

$$\mathrm{Var}(\chi_n^2) = \mu_2' - \mu_1'^2 = n^2+2n-n^2 = 2n.$$

<u>Homework:</u> If $X_1, \ldots, X_n$ follows IID $\mathrm{Bin}(1,p)$; what will be the distribution of $\sum X_i$. Compute mean and variance of the distn. of $\sum X_i$ using MGF.

<u>Product Term:</u>

$$2\sum(x_i-\bar{x})(\bar{x}-\mu)$$

$$= 2\left[\bar{x}\sum x_i - \mu\sum x_i - n\bar{x}^2 + \bar{x}\sum\mu\right]$$

$$= 2\left[\bar{x}\cdot n\bar{x} - \mu\cdot n\bar{x} - n\bar{x}^2 + n\bar{x}\cdot\mu\right] \quad ; \text{ since } \bar{x} = \frac{1}{n}\sum x_i$$

$$= 0.$$

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

Mean test

Z-test — σ known

t-test — σ unknown

$$\cancel{\chi} = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,\, n_2-1}$$

Variance test

χ^2-test — one-sample

F-test — two-sample

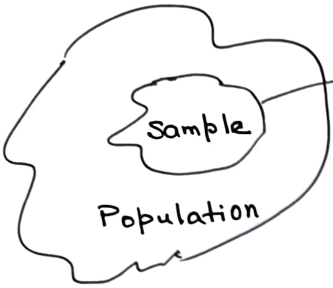
Statistical Inference

Parametric Estimation

- Point
- Interval → CI
- Parameter Estimate validation → TOH, ANOVA

METHODS: MLE/MME ($l(\theta|x)$)

PROPERTIES: UNBIASED ($E(T_1)=\theta$); MIN. VAR [$V(T_1) < V(T_2)$]



mean{Heights} (statistic : $T(X) = \bar{X}$)

Is it a "good" estimator of population heights?

UNBIASED* | CONSISTENT $|T-\theta|<\epsilon$ | SUFFICIENT | MINIMUM VARIANCE* | EFFICIENT | COMPLETE

Let $T(\underset{\sim}{X})$ be an estimator of θ [(μ, σ^2) for Normal, (n, p) for Binomial, λ for Poisson, etc.].

Mean-squared Error (MSE): $\underset{\theta}{\text{Minimize}}$ MSE : Task

$$\begin{aligned}
MSE_\theta(T) &= \mathbb{E}_\theta[T-\theta]^2 \\
&= \mathbb{E}_\theta[\underbrace{T-\mathbb{E}(T)}+\underbrace{\mathbb{E}(T)-\theta}]^2 \\
&= \mathbb{E}[T-\mathbb{E}(T)]^2 + [\mathbb{E}(T)-\theta]^2 \quad \left[\text{Product term} = 0 \text{ (last class)}\right] \\
&= \mathbb{V}\text{ar}(T) + \text{Bias}^2(T)
\end{aligned}$$

[ML: Bias-Variance Trade-off]

Var(T) ↓ Minimum Variance | Bias²(T) ↓ Unbiasedness

Estimator (UE) is $\mathbb{E}(T) = \theta$.

UNBIASEDNESS: (i) $X_i \overset{IID}{\sim} Bin(n, p)$

Ques: UE of p?

$$\mathbb{E}(X) = np$$
$$\Rightarrow p = \frac{\mathbb{E}(X)}{n} = \mathbb{E}\left(\frac{X}{n}\right) = \mathbb{E}(T(X))$$
$$\Rightarrow T(X) = \frac{X}{n} \text{ is an UE of } p.$$

(ii) $X_i \overset{IID}{\sim} N(\underbrace{\mu, \sigma^2}_{\text{popln. mean \& variance}})$

$\bar{X}$: sample mean

S^2: sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

$[\mathbb{E}(\bar{X}) = \mu]$ $\left[\sigma^2 = \frac{1}{n}\sum(X_i-\mu)^2\right]$

QE: S^2 is an UE of σ^2?

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right) &= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right) \\
&= \frac{1}{n-1}\mathbb{E}\left\{\sum_{i=1}^{n}\left(\overline{X_i-\mu} - \overline{\bar{X}-\mu}\right)^2\right\} \\
&= \frac{1}{n-1}\mathbb{E}\left\{\sum_{1}^{n}(X_i-\mu)^2 - 2\sum_{1}^{n}(X_i-\mu)(\bar{X}-\mu) + n(\bar{X}-\mu)^2\right\} \\
&= \frac{1}{n-1}\mathbb{E}\left\{\sum(X_i-\mu)^2 - 2n(\bar{X}-\mu) + n(\bar{X}-\mu)^2\right\} \leftarrow \left[\text{since } \sum(X_i-\mu) = n(\bar{X}-\mu)\right] \\
&= \frac{1}{n-1}\left\{\sum\mathbb{E}(X_i-\mu)^2 - n\mathbb{E}(\bar{X}-\mu)^2\right\} = \frac{1}{n-1}\left\{n\sigma^2 - \frac{n\sigma^2}{n}\right\} = \sigma^2.
\end{aligned}$$

Example on minimum variance unbiased estimator (From Slide):

$X_1, X_2, X_3, X_4, X_5 \overset{\text{IID}}{\sim} N(\mu, \sigma^2)$. Three estimators are given:

$$T_1 = \frac{X_1+X_2+X_3+X_4+X_5}{5} \; ; \quad T_2 = \frac{X_1+X_2}{2} + X_3 \; ; \quad T_3 = \frac{2X_1+X_2+\lambda X_3}{3}$$

Ques: T_1 is unbiased? $\mathbb{E}(T_1) = \mathbb{E}\left(\frac{X_1+X_2+X_3+X_4+X_5}{5}\right) = \frac{5\mu}{5} = \mu$ (Yes)

(due to independence)

T_2 is unbiased? $\mathbb{E}(T_2) = \mathbb{E}\left[\frac{X_1+X_2}{2} + X_3\right] = \frac{\mu+\mu}{2} + \mu = 2\mu \neq \mu$ (No)

T_3 is unbiased, then $\lambda = ?$ $\mathbb{E}(T_3) = \mathbb{E}\left(\frac{2X_1+X_2+\lambda X_3}{3}\right) = \frac{2\mu}{3} + \frac{\mu}{3} + \frac{\lambda\mu}{3} = \mu$

$\Rightarrow \lambda = 0$.

"Best" estimator among T_1, T_2, T_3?

T_2 is not.

$$\mathbb{V}(T_1) = \mathrm{Var}\left(\frac{1}{5}\sum_{i=1}^{5} X_i\right) = \frac{1}{25}\sum_{i=1}^{5} \mathrm{Var}(X_i) = \frac{1}{25}(\sigma^2 \times 5) = \frac{\sigma^2}{5}.$$

Recall
$\mathrm{Var}(aX) = a^2 V(X)$;
$\mathrm{Var}(\sum X_i) = \sum V(X_i)$ if X_i's are independent.

$$<$$

$$\mathbb{V}(T_3) = \mathrm{Var}\left(\frac{2X_1+X_2}{3}\right) = \frac{1}{9}\left[V(2X_1) + V(X_2)\right] = \frac{1}{9}\left[4V(X_1) + V(X_2)\right] = \frac{1}{9}\left[4\sigma^2 + \sigma^2\right] = \frac{5\sigma^2}{9}.$$

Thus, T_1 is better than T_3.

Consistency: [Recall Markov's Inequality (for higher-order): $\mathbb{P}[|X| > \epsilon] < \frac{\mathbb{E}|X|^r}{\epsilon^r} \; \forall r > 0, \epsilon > 0$]

$$|T-\theta| < \epsilon \Rightarrow \lim_{n\to\infty} \mathbb{P}\left[|T-\theta| < \epsilon\right] = 1 \Rightarrow \lim_{n\to\infty} \mathbb{P}\left[|T-\theta| > \epsilon\right] = 0$$

Using Markov's inequality: $\mathbb{P}[|T-\theta| > \epsilon] < \frac{\mathbb{E}(T-\theta)^2}{\epsilon^2} =$

$$= \frac{\mathbb{E}(T - \mathbb{E}(T) + \mathbb{E}(T) - \theta)^2}{\epsilon^2}$$

$$= \frac{\underbrace{\mathbb{E}(T-\mathbb{E}(T))^2}_{V(T)} + \underbrace{(\mathbb{E}(T)-\theta)^2}_{\text{Bias}}}{\epsilon^2}$$

$\to 0$ implies

Thus, sufficient condition for consistency:

(I) $\mathbb{E}(T) = \theta$

(II) $\mathbb{V}(T) \to 0$ as $n \to \infty$

Example: $X_1, \ldots, X_n \overset{\text{IID}}{\sim} \text{Pois}(\lambda)$. Find a consistent estimator for λ. (Show that $\bar{X}$ is a consistent estimator for λ).

$\to$ $\mathbb{E}(X_i) = \lambda = \mathbb{V}(X_i)$

Try with $\bar{X}$:

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\cdot n\lambda = \lambda.$$

$$\mathbb{V}(\bar{X}) = \mathbb{V}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\sum^{n}\mathbb{V}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\lambda = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$$

$\to 0$ as $n \to \infty$.

Method of Parameter Estimation: MAXIMUM LIKELIHOOD ESTIMATORS

Likelihood Function: The probability (likelihood) of the observed sample given the parameter. The likelihood function is a function of the parameter. Suppose θ is the unknown parameter. We write the likelihood function as $L(\theta | x_1, x_2, \ldots, x_n)$.

Note: Likelihood function is not probability. If we sum (or integrate) $L(\theta | x_1, \ldots, x_n)$ over all possible values of θ, it will not become 1.

Maximum Likelihood Principle: Choose as your estimates those values of the parameter that maximizes likelihood of the observed data.

Log likelihood: Likelihood function is $L(\theta | \underset{\sim}{x}) = \prod_{i=1}^{n} p(x_i, \theta)$. The natural logarithm of the likelihood function. It is often preferable to work with the log likelihood for both practical and theoretical reason. The log likelihood converts the product into sum and hence it's easier to handle. If we take logarithm, it results in a large number (since product of probabilities is a tiny value) and $\log L(\theta | \underset{\sim}{x})$ is always negative.

Implication: A likelihood method is a measure of how well a particular model fits a data. They explain how well a parameter explains the observed data.

Advantages of log likelihood: Loglikelihood increase the numerical stability of the estimates. Likelihood functions are product of marginal probabilities are tend to become very small for large samples. Log likelihoods are large negative numbers and hence their usage improves numerical stability.

Kernel Likelihood: $L(\theta | x) = K(\underset{\sim}{x})\, p(\theta | \underset{\sim}{x})$; where $K(\underset{\sim}{x})$ is a function of the observed data and does not involve the parameter to be estimated. **Example:** Suppose $X_i \overset{IND.}{\sim} \text{Pois}(\lambda)$

$$L(\lambda | x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i, \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!} = K(\underset{\sim}{x})\, p(\underset{\sim}{x} | \lambda)$$

where $K(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{x_i!}$ and $p(\theta | x_1, \ldots, x_n) = e^{-n\lambda} \cdot \lambda^{\sum x_i}$.

MLE Example: $X_1,\dots,X_n \overset{IID}{\sim} \text{Pois}(\lambda),\ \lambda>0.$

$$f(x_i) = \mathbb{P}(X_i = x_i) = \frac{e^{-\lambda}.\lambda^{x_i}}{x_i!};\ x_i = 0,1,\dots$$

$$L = L(\lambda|\underset{\sim}{x}) = \prod_{i=1}^{n} f(x_i;\lambda) = \prod_{i=1}^{n}\left(\frac{e^{-\lambda}.\lambda^{x_i}}{x_i!}\right) = e^{-n\lambda}.\frac{\lambda^{\sum_1^n x_i}}{\prod_1^n x_i}$$

$$\ln L = \ln L(\lambda|\underset{\sim}{x}) = -n\lambda \ln e + \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n}\ln x_i$$

$$\frac{\partial}{\partial\lambda}\ln L = -n + \frac{\sum x_i}{\lambda} = -n + \frac{n\bar{x}}{\lambda} = \frac{n(\bar{x}-\lambda)}{\lambda} = \begin{cases} >0 & \text{if } \bar{x}>\lambda \\ <0 & \text{if } \bar{x}<\lambda \end{cases}$$

$L(\lambda|\underset{\sim}{x})$ first increases then achieves its maximum at $\bar{x}=\lambda$ and then decreases.
Hence, $L(\lambda|\underset{\sim}{x})$ is maximum at $\hat{\lambda}=\bar{x}$.

BIVARIATE DATA: – Correlation and Scatter Diagram.

Bivariate data: (x_i, Y_i)

$$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$
$$= \frac{1}{n}\sum(x_i-\bar{x})(Y_i-\bar{Y}) = \frac{1}{n}\sum x_iY_i - \bar{x}\bar{y}$$
$$V(X) = \frac{1}{n}\sum(x_i-\bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2.$$
$$\mu = \mathbb{E}(X)$$

correlation co-efficient, $r_{XY} = \dfrac{cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$; (measures the linear relationship between x (cause) and y (effect))

Taking $u_i = x_i - \bar{x}$, $v_i = y_i - \bar{y}$

$$= \frac{\sum(x_i-\bar{x})(Y_i-\bar{Y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(Y_i-\bar{Y})^2}}$$

$$= \frac{\sum u_i v_i}{\sqrt{\sum u_i^2}\sqrt{v_i^2}}$$

RESULT 1: $-1 \le r_{XY} \le 1$

[Cauchy – Schwarz Inequality: For any sequence of real numbers a_i and b_i $(i=1(1)n)$; we have
$$\left(\sum_{i=1}^{n} a_ib_i\right)^2 \le \left(\sum_1^n a_i^2\right)\left(\sum_1^n b_i^2\right).$$
'=' holds when $a_i = k b_i$ (k constant)]

By CS-inequality:
$$\left[\sum_1^n (x_i-\bar{x})(y_i-\bar{y})\right]^2 \le \left(\sum_1^n (x_i-\bar{x})^2\right)\left(\sum_1^n (y_i-\bar{y})^2\right)$$
$$\Rightarrow -1 \le r_{XY} \le 1$$

'=' holds iff $y_i - \bar{y} = k(x_i - \bar{x})$
iff two variables (x_i and y_i) are exactly linearly related with positive ($r_{XY}=1$) and negative ($r_{XY}=-1$) slope.

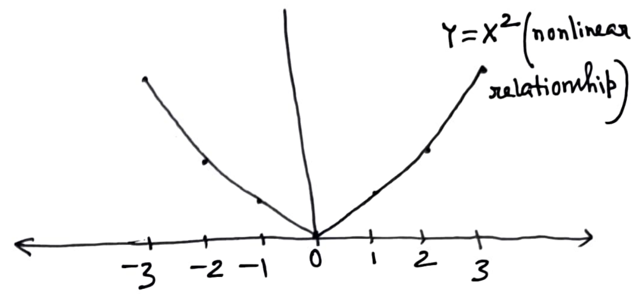
RESULT 2: $r_{XY} = r_{YX}$ (Trivial)

RESULT 3: If X and Y are independent, $r_{XY}=0$.
$\rightarrow \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \Rightarrow cov(X,Y)=0 \Rightarrow r_{XY}=0.$

RESULT 4: $r_{XY} = 0$ does not necessarily imply there is no relationship between X and Y.

Counter Eg.

x	-3	-2	-1	0	1	2	3
y	9	4	1	0	1	4	9



$$Cov(X,Y) = \frac{1}{n}\sum X_iY_i - \bar{x}\,\bar{y} = 0$$

$$\Rightarrow r_{XY} = 0.$$

RESULT 5: r_{XY} is independent of change of origin and scale.

$$u_i = \frac{x_i - a}{c} \quad (c \neq 0) \qquad\qquad v_i = \frac{y_i - b}{d} \quad (d \neq 0)$$

$$\Rightarrow x_i = a + cu_i \qquad\qquad \Rightarrow y_i = b + dv_i$$

$$\Rightarrow \bar{x} = a + c\bar{u} \qquad\qquad \Rightarrow \bar{y} = b + d\bar{v}$$

$$\Rightarrow Var(x_i) = c^2 Var(u_i) \qquad\qquad \Rightarrow Var(y_i) = d^2 Var(v_i)$$

$$Cov(X,Y) = \frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})$$
$$= \frac{cd}{n}\sum (u_i - \bar{u})(u_i - \bar{v})$$
$$= cd\,cov(u,v)$$

$$r_{XY} = \frac{cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{cd\,cov(u,v)}{\sqrt{c^2V(u)}\sqrt{d^2V(v)}}$$

$$= \frac{cd\,cov(u,v)}{|c|\,|d|\sqrt{V(u)}\sqrt{V(v)}}$$

$$= \frac{cd}{|c||d|} r_{uv} = \begin{cases} r_{uv} & \text{if } c \text{ and } d \text{ are of same sign.} \\ -r_{uv} & \text{if } c \text{ and } d \text{ are of opp. sign.} \end{cases}$$