Probability, Statistics & Mathematics
(Cheat Sheet)

Author: **Anik Chakraborty**

Contents

1 Probability **1**
1.1 Theory of Probability . . . 1
1.1.1 Sigma Field . . . 1
1.1.2 Properties . . . 1
1.1.3 Conditional Probability . . . 2
1.1.4 Stochastic Independence . . . 3
1.2 Random Variable . . . 3
1.2.1 Univariate . . . 3
1.2.2 Bivariate . . . 5
1.2.3 Results . . . 5
1.3 Generating Functions . . . 6
1.3.1 Moments . . . 6
1.3.2 Cumulants . . . 6
1.3.3 Characteristic Function . . . 7
1.3.4 Probability Generating Function . . . 7
1.4 Inequalities . . . 7
1.4.1 Markov & Chebyshev . . . 7
1.4.2 Cauchy-Schwarz . . . 8
1.4.3 Jensen . . . 8
1.4.4 Lyapunov . . . 8
1.5 Theoretical Distributions . . . 9
1.5.1 Discrete . . . 9
1.5.2 Continuous . . . 10
1.5.3 Multivariate . . . 12
1.5.4 Truncated Distribution . . . 14
1.6 Sampling Distributions . . . 15
1.6.1 Chi-square, t, F . . . 15
1.6.2 Order Statistics . . . 16
1.7 Distribution Relationships . . . 17
1.7.1 Binomial . . . 17
1.7.2 Negative Binomial . . . 17
1.7.3 Poisson . . . 17
1.7.4 Normal . . . 18
1.7.5 Gamma . . . 18
1.7.6 Beta . . . 18
1.7.7 Cauchy . . . 18
1.7.8 Others . . . 18

1.8 Transformations . . . 19
1.8.1 Orthogonal . . . 19
1.8.2 Polar . . . 19
1.8.3 Special Transformations . . . 19

2 Statistics 21
2.1 Point Estimation . . . 21
2.1.1 Minimum MSE . . . 21
2.1.2 Consistency . . . 22
2.1.3 Sufficiency . . . 23
2.1.4 Completeness . . . 23
2.1.5 Exponential Family . . . 24
2.1.6 Methods of finding UMVUE . . . 25
2.1.7 Cramer-Rao Inequality . . . 28
2.1.8 Methods of Estimation . . . 29
2.2 Testing of Hypothesis . . . 30
2.2.1 Tests of Significance . . . 30
2.3 Interval Estimation . . . 32
2.3.1 Methods of finding C.I. . . . 32
2.3.2 Wilk's Optimum Criteria . . . 32
2.3.3 Test Inversion Method . . . 32
2.4 Large Sample Theory . . . 33
2.4.1 Modes of Convergence . . . 33

3 Mathematics 35
3.1 Basics . . . 35
3.1.1 Combinatorial Analysis . . . 35
3.1.2 Difference Equation . . . 36
3.2 Linear Algebra . . . 37
3.2.1 Vectors & Vector Spaces . . . 37
3.2.2 Matrices . . . 38
3.2.3 Determinants . . . 40
3.2.4 System of Linear Equation . . . 42

Chapter 1

Probability

1.1 Theory of Probability

1.1.1 Sigma Field

Ω: Universal Set. A non-empty class $\mathcal{A}$ of few subsets of Ω is said to form a sigma field on Ω if it satisfies the following properties-

(**i**) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$ (Closed under complementation)

(**ii**) $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{A} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ (Closed under countable union)

Theorems

(**1**) A σ-field is closed under finite unions.

(**2**) A σ-field must include the null set, ϕ and the whole set, Ω.

- (a) $\mathcal{A} = \{\emptyset, \Omega\}$ is the smallest/minimal σ-field on Ω.
- (b) If $A \in \Omega$, then $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$ is the minimal σ-field containing A, on Ω.
- (c) The power set of Ω (the set of all subsets of Ω) is the largest σ-field on Ω.

(**3**) A σ-field is closed under countable intersections.

1.1.2 Properties

(**1**) For two sets $A, B \in \mathcal{A}$ -

- (a) **Monotonic Property:** If $A \subseteq B$, $P(A) \leq P(B)$
- (b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) \implies P(A \cup B) \leq P(A) + P(B)$
- (c) $P(A \cup B) = P(A - B) + P(B - A) + P(A \cap B)$
- (d) $P(A \cap B) \leq \min\{P(A), P(B)\} \implies P(A \cap B) \leq \sqrt{P(A) \cdot P(B)}$
- (e) $P(A \cap B) \geq P(A) + P(B) - 1$
- (f) $P(A) = P(A \cap B) + P(A \cap B^c) \implies P(A - B) = P(A) - P(A \cap B)$

(2) For any n events $A_1, A_2, \ldots, A_n \in \mathcal{A}$ -

(a) **Boole's inequality:** $P\left(\bigcup\limits_{i=1}^{n} A_i\right) \leq \sum\limits_{i=1}^{n} P(A_i)$

(b) **Bonferroni's inequality:** $P\left(\bigcap\limits_{i=1}^{n} A_i\right) \geq \sum\limits_{i=1}^{n} P(A_i) - (n-1)$

(c) $\sum\limits_{i=1}^{n} P(A_i) - \sum_{1\leqslant i_1<i_2\leqslant n} P(A_{i_1} \cap A_{i_2}) \leq P\left(\bigcup\limits_{i=1}^{n} A_i\right) \leq \sum\limits_{i=1}^{n} P(A_i)$

(d) **Poincare's theorem:**

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \;-\; \sum_{1\leqslant i_1<i_2\leqslant n} P(A_{i_1} \cap A_{i_2}) \;+\; \sum_{1\leqslant i_1<i_2<i_3\leqslant n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \;-\; \cdots$$

$$\cdots \;+\; (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

(e) **Jordan's theorem:**

i. The probability that exactly m of the n events will occur is -

$$P_{(m)} = S_m - \binom{m+1}{m} S_{m+1} + \binom{m+2}{m} S_{m+2} - \cdots + (-1)^{n-m} \binom{n}{m} S_n$$

ii. The probability that atleast m of the n events will occur is -

$$\begin{aligned} P_m &= P_{(m)} + P_{(m+1)} + \cdots + P_{(n)} \\ &= S_m - \binom{m}{m-1} S_{m+1} + \binom{m+1}{m-1} S_{m+2} - \cdots + (-1)^{n-m} \binom{n-1}{m-1} S_n \end{aligned}$$

where, $S_r = \sum_{1\leqslant i_1<i_2<\cdots<i_r\leqslant n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r})$, $r = 1(1)n$

1.1.3 Conditional Probability

Consider, a probability space $(\Omega, \mathcal{A}, \mathcal{P})$.

(1) Compound probability: n events $A_1, A_2, \ldots, A_n \in \mathcal{A}$ are such that $P\left(\bigcap\limits_{i=1}^{n-1} A_i\right) > 0$. Then,

$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \cdots P(A_n/\bigcap\limits_{i=1}^{n-1} A_i)$

(2) Total Probability Theorem: If $(B_1, B_2, \ldots, B_n)$ is a partition of Ω with $P(B_i) > 0\ \forall i$, then for any event $A \in \mathcal{A}$, $P(A) = \sum\limits_{i=1}^{n} P(B_i) \cdot P(A/B_i)$

(3) Bayes' Theorem: $P(B_i/A) = \frac{P(B_i)P(A/B_i)}{\sum\limits_{k=1}^{n} P(B_k)P(A/B_k)}$, $i = 1(1)n$, if $P(A) > 0$

(**4**) **Bayes' theorem with future events:** Let, $C \in \mathcal{A}$ be an event under the previous conditions with $P(A/B_i) > 0, i = 1(1)n$. Then,

$$P(C/A) = \frac{\sum_{i=1}^{n} P(B_i)P(A/B_i)P(C/A \cap B_i)}{\sum_{i=1}^{n} P(B_i)P(A/B_i)}$$

1.1.4 Stochastic Independence

(**1**) For two independent events A, B -

$$P(A/B) = P(A/B^c) = P(A) \iff P(A \cap B) = P(A) \cdot P(B)$$

(**2**) **Pairwise independence:** For n events $A_1, A_2, \ldots, A_n \in \mathcal{A}$ are said to be 'pairwise' independent if -

$$P(A_{i_1} \cap A_{i_2}) = P(A_{i_1}) \cdot P(A_{i_2}), \forall i_1 < i_2$$

(**3**) **Mutual independence:** The above events are 'mutually' independent if -

$$\begin{aligned}
&P(A_{i_1} \cap A_{i_2}) = P(A_{i_1}) \cdot P(A_{i_2}),\ \forall i_1 < i_2 \\
&P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot P(A_{i_3}),\ \forall i_1 < i_2 < i_3 \\
&\qquad\vdots \\
&P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n)
\end{aligned}$$

1.2 Random Variable

1.2.1 Univariate

$X : \Omega \to \mathbb{R}$, such that $\{\omega : X(\omega) \leq x\} \in \mathcal{A}$, $\forall x \in \mathbb{R}$ is a random variable on $\{\Omega, \mathcal{A}\}$.

(**1**) The same variable $X(\cdot)$ is a R.V. for a particular choice of σ-field but may not for another choice of σ-field.

(**2**) X is a R.V. on $(\Omega, \mathcal{A}) \implies f(X)$ is also a R.V. on $(\Omega, \mathcal{A})$. (for any f)

(**3**) **Continuity theorem of Probability:** $(A_1 \subset A_2 \subset \ldots)$ or $(A_1 \supset A_2 \supset \ldots)$

$$\implies \lim_{n\to\infty} P(A_n) = P\left(\lim_{n\to\infty} A_n\right)$$

(**4**) **CDF:** $F_X(x) = P[\{\omega : X(\omega) \leq x\}]$, $\forall x \in \mathbb{R}$

- (a) **Non-decreasing:** $-\infty < x_1 < x_2 < \infty \implies F(x_1) \leq F(x_2)$
- (b) **Normalized:** $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$
- (c) **Right Continuous:** $\lim_{x\to a^+} F(x) = F(a)$, $\forall a \in \mathbb{R}$

For any R.V. X with CDF $F(\cdot)$ -

$$P(a < X < b) = F(b-0) - F(a) \qquad P(a \le X \le b) = F(b) - F(a-0)$$

$$P(a < X \le b) = F(b) - F(a) \qquad P(a \le X < b) = F(b-0) - F(a-0)$$

(**5**) **Decomposition theorem:** $F(x) = \alpha F_c(x) + (1-\alpha)F_d(x)$ where, $0 \le \alpha \le 1$ and $F_c(x), F_d(x)$ are continuous and discrete D.F., respectively.

(a) $\alpha = 0 \implies X$ is purely discrete.

(b) $\alpha = 1 \implies X$ is purely continuous.

(c) $0 < \alpha < 1 \implies X$ is mixed.

(**6**) X is non-negative with $E(X) = 0 \implies P(X = 0) = 1$

(**7**) $P(a \le X \le b) = 1 \implies Var(X) \le \frac{(b-a)^2}{4}$

(**8**) $P(|X| \le M) = 1$ for some $0 \le M < \infty \implies \mu_r'$ exists $\forall\, r$

(**9**) $P\left(X \in \{0, \mathbb{N}\}\right) = 1 \implies E(X) = \sum\limits_{x=0}^{\infty} \{1 - F(x)\}$

(**10**) $P(X \in [0,\infty)) = 1 \implies \lim\limits_{x \to \infty} x\{1 - F(x)\} = 0$, if $E(X)$ exists.

(**11**) $E(X) = \int\limits_0^{\infty} \{1 - F(x)\}dx$ for any non-negative R.V. X.

$$E(X^r) = \int_0^{\infty} r\, x^{r-1} \{1 - F(x)\} dx$$

(**12**) $\ln(\mathrm{GM}_X) = E(\ln X)$

(**13**) $\boldsymbol{p^{th}}$ **quantile:** ξ_p such that $F(\xi_p - 0) \le p \le F(\xi_p)$. For continuous case, $F(\xi_p) = p$

Symmetry

X has a symmetric distribution about 'a' if any of the following, holds -

(a) $P(X \le a - x) = P(X \ge a + x), \forall x \in \mathbb{R}$

(b) $F(a-x) + F(a+x) = 1 + P(a+x)$

Again, if X is continuous then $F(a-x) + F(a+x) = 1$ or $f(a-x) = f(a+x), \ \forall x \in \mathbb{R}$

- $E(X) = a$, if it exists
- $\mathrm{Med}\,(X) = a$

1.2.2 Bivariate

$\binom{X}{Y} : \Omega \to \mathbb{R}^2$, such that $\{\omega : X(\omega) \le x, Y(\omega) \le y\} \in \mathcal{A}$, $\forall (x,y) \in \mathbb{R}^2$ is a bivariate random variable on $\{\Omega, \mathcal{A}\}$.

(1) **CDF:** $F(x,y) = P[\{\omega : X(\omega) \le x, Y(\omega) \le y\}]$, $\forall (x,y) \in \mathbb{R}^2$

(a) $F(x,y)$ is non-decreasing and right continuous w.r.t. each of the arguments x and y.

(b) $F(-\infty, y) = F(x, -\infty) = 0$, $F(+\infty, +\infty) = 1$

(c) For $x_1 < x_2, y_1 < y_2$ -

$$P(x_1 < X < x_2,\ y_1 < Y < y_2) = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1) \ge 0$$

Marginal CDF

$F_X(x) = \lim\limits_{y\to\infty} F_{X,Y}(x,y),\ F_Y(y) = \lim\limits_{x\to\infty} F_{X,Y}(x,y)$

- $F_X(x) + F_Y(y) - 1 \le F_{X,Y}(x,y) \le \sqrt{F_X(x)\cdot F_Y(y)}$, $\forall (x,y) \in \mathbb{R}^2$

(2) Joint distribution cannot be determined uniquely from the marginals.

(3) $f_{X,Y}(x,y) = f_X(x)\cdot f_{Y|X}(y|x) = f_Y(y)\cdot f_{X|Y}(x|y)$

(4) $f_{X,Y}(x,y;\alpha) = f_X(x)\, f_Y(y) \left\{1 + \alpha \cdot \overline{2F_X(x) - 1} \cdot \overline{2F_Y(y) - 1}\right\},\ \alpha \in [-1,1]$

Stochastic Independence

(5) $F_{X,Y}(x,y) = F_X(x)\cdot F_Y(y),\ \forall (x,y) \in \mathbb{R}^2 \implies f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y),\ \forall (x,y)$

(6) $X \perp\!\!\!\perp Y \implies f(X) \perp\!\!\!\perp g(Y)$ (converse is true when f, g is 1-1)

(7) $X \perp\!\!\!\perp Y$ iff $f_{X,Y}(x,y) = k\cdot f_1(x)\cdot f_2(y),\ \forall\, x,y \in \mathbb{R}$ for some $k > 0$.

1.2.3 Results

(1) **Sum Law:** $E(X+Y) = E(X) + E(Y)$, if all exists

(2) **Product Law:** $X \perp\!\!\!\perp Y \implies E(XY) = E(X)\cdot E(Y)$
$\text{Cov}\,(X,Y) = 0 \not\Rightarrow X \perp\!\!\!\perp Y$

(3) X, Y identically distributed $\not\Rightarrow P(X = Y) = 1$

(4) $X_1, X_2, \ldots, X_n$ are *iid* and continuous R.V.s $\implies n!$ arrangements are equally likely

(5) $X \overset{iid}{\sim} Y \implies (X - Y)$ is symmetric

(6) PDF of $\max\{X,Y\} : f_U(u) = \int\limits_{-\infty}^{u} \{f(u,t) + f(t,u)\}dt$ $\quad f$: Joint PDF of (X,Y)

Conditional Distribution

(7) $X \perp\!\!\!\perp Y \implies E(Y|X=x) = k$, some constant $\forall\, x$

(8) $X \perp\!\!\!\perp \left(Y - \rho\frac{\sigma_Y}{\sigma_X}\right) \implies E(Y|X=x)$ is linear in x

(9) $E(Y) = E[E(Y|X)]$ or $E(X) = E[E(X|Y)]$

(10) $Var\,(Y) = Var\,\{E(Y|X)\} + E\{Var\,(Y|X)\}$

(11) **Correlation ratio:** $\eta^2_{YX} = \frac{Var\,\{E(Y|X)\}}{Var\,(Y)}$

(12) **Wald's equation:** $\{X_n\}$: sequence of *iid* R.V.s, $P(N \in \mathbb{N}) = 1$. Define, $S_N = \sum\limits_{i=1}^{N} X_i$

$$\implies E(S_N) = E(X_1)\,E(N)$$

$$\implies Var\,(S_N) = Var\,(X_1) \cdot E(N) + E^2(X_1) \cdot Var\,(N)$$

1.3 Generating Functions

1.3.1 Moments

(1) **MGF:** $M_X(t) = E\left(e^{tX}\right)$, $|t| < t_0$, for some $t_0 > 0$ [if $E\left(e^{tX}\right) < \infty$]
It determines a distribution uniquely.

(2) $\boldsymbol{\mu'_r}$ **:** coefficient of $\frac{t^r}{r!}$ in the expansion of $M_X(t)$, $r = 0, 1, 2, \ldots$

(3) If the power series: $\sum\limits_{r=0}^{\infty} \frac{t^r \mu'_r}{r!}$ converges absolutely, then a sequence of moments $\{\mu'_r\}$ determine a distribution uniquely. For a bounded R.V this always holds.

(4) X_i are independent with MGF $M_i(t) \implies M_S(t) = \prod_{i=1}^{n} M_i(t)$, where $S = \sum_{i=1}^{n} X_i$

(5) **Bivariate MGF:** $M_{X,Y}(t_1, t_2) = E\left(e^{t_1 X + t_2 Y}\right)$ for $|t_i| < h_i$ for some $h_i > 0,\ i = 1,\, 2$

(6) $\boldsymbol{\mu'_{r,s}}$ **:** coefficient of $\frac{t_1^r t_2^s}{r!s!}$ in the expansion of $M_{X,Y}(t_1, t_2)$

(7) Also, $\frac{\partial^{r+s}}{\partial t_1^r \partial t_2^s} M_{X,Y}(t_1, t_2)\ |_{(t_1=0, t_2=0)} = \mu'_{r,s}$

(8) **Marginal MGF:** $M_X(t) = M_{X,Y}(t, 0)$ & $M_Y(t) = M_{X,Y}(0, t)$

(9) X & Y are independent 'iff' $M_{X,Y}(t_1, t_2) = M_{X,Y}(t_1, 0) \cdot M_{X,Y}(0, t_2),\ \forall (t_1, t_2)$

1.3.2 Cumulants

(1) **CGF:** $K_X(t) = \ln\{M_X(t)\}$, provided the expansion is a <u>convergent power series</u>.

(2) $k_1 = \mu'_1$ (mean), $k_2 = \mu_2$ (variance), $k_3 = \mu_3$ & $k_4 = \mu_4 - 3k_2^2$

(3) For two independent R.V. X & Y, $k_r(X + Y) = k_r(X) + k_r(Y)$

1.3.3 Characteristic Function

(1) **CF:** $\phi_X(t) = E\left(e^{itX}\right)$

(2) $\phi_X(0) = 1$, $|\phi_X(t)| \leq 1$

(3) $\phi_X(t)$ is continuous on $\mathbb{R}$ and always exists for $t \in \mathbb{R}$

(4) $\phi_X(-t) = \overline{\phi_X(t)}$

(5) If X has a symmetric distribution about '0' then $\phi_X(t)$ is real valued and an even function of t.

(6) Uniqueness property and independence as of MGF.

(7) **Inversion theorem:** If $\int_{-\infty}^{\infty} \phi_X(t)dt < \infty$, then pdf of the distribution is -

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itX}\phi_X(t)dt$$

1.3.4 Probability Generating Function

(1) **PGF:** $P_X(t) = E(t^X)$, if $|t| < 1$

(2) It generates probability and factorial moments. It also determines a distribution uniquely.

(3) r^{th} order factorial moment: $\mu_{[r]} = \frac{d^r}{dt^r}P_X(t)\,|_{(t=1)}$, $r = 0, 1, \ldots$

(4) $X_1, X_2, \ldots, X_n$ are independent with PGF $P_i(t) \implies P_S(t) = \prod_{i=1}^n P_i(t)$, where $S = \sum_{i=1}^n X_i$

1.4 Inequalities

1.4.1 Markov & Chebyshev

(1) **Markov:** For a non-negative R.V. X, $P(X \geqslant a) \leqslant \frac{E(X)}{a}$, for $a > 0$.
'=' holds if X has a two-point distribution.

(2) **Chebyshev:** $P(|X - \mu| \geqslant t\sigma) \leqslant \frac{1}{t^2}$, $t > 0$ where $\mu = E(X)$ & $\sigma^2 = Var(X) < \infty$.
'=' holds if X is such that -

$$f(x) = \begin{cases} \dfrac{1}{2t^2} & \text{, if } x = \mu \pm t\sigma \\ 1 - \dfrac{1}{t^2} & \text{, if } x = \mu \end{cases} \qquad (t > 1)$$

(3) **One-sided Chebyshev:** $E(X) = 0$, $Var(X) = \sigma^2 < \infty$

$$P(X \geqslant t) \begin{cases} \leqslant \dfrac{\sigma^2}{\sigma^2 + t^2} & \text{, if } t > 0 \\ \geqslant \dfrac{t^2}{\sigma^2 + t^2} & \text{, if } t < 0 \end{cases}$$

(4) If also $\mu_4 < \infty$ then,

$$P(|X-\mu| \geqslant t\sigma) \leqslant \frac{\mu_4 - \sigma^4}{\mu_4 - \sigma^4 + (t^2-1)^2\sigma^4}$$

It is an improvement over Chebyshev's inequality if $t^2 \geq \frac{\mu_4}{\sigma^4}$

(5) **Bivariate Chebyshev:** (X_1, X_2) is a bivariate R.V. with means (μ_1, μ_2), variances (σ_1^2, σ_2^2) & correlation ρ. Then for $t > 0$,

$$P(|X_1-\mu_1| \geqslant t\sigma_1 \text{ or } |X_2-\mu_2| \geqslant t\sigma_2) \leqslant \frac{1+\sqrt{1-\rho^2}}{t^2}$$

1.4.2 Cauchy-Schwarz

If a bivariate R.V. (X, Y) has finite variances and $E(XY)$ exists, then -

$$E^2(XY) \leq E(X^2)E(Y^2)$$

'=' holds iff X & Y are linearly related passing through the origin i.e. $P(X + \lambda Y = 0) = 1$, for any λ.

1.4.3 Jensen

$f(\cdot)$ is convex function and $E(X)$ exists, then $E\left[f(X)\right] \geq f\left[E(X)\right]$

Note: A function, $f(\cdot)$ is said to be **convex** on an interval I, if for $x_1, x_2 \in I$ and for some $\lambda \in [0,1]$, if

$$f[\lambda x_1 + (1-\lambda)x_2] \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

If $f(\cdot)$ is twice differentiable then $f''(x) \geq 0$ is the condition for convexity.

1.4.4 Lyapunov

For a R.V. X, define $\beta_r = E\left(|X|^r\right)$ (assuming it exists) then -

$$\left\{\beta_r^{\frac{1}{r}}\right\} \text{ is non decreasing i.e. } \beta_r^{\frac{1}{r}} \leq \beta_{r+1}^{\frac{1}{r+1}}$$

1.5 Theoretical Distributions

1.5.1 Discrete

X	CDF	PDF	E(X)	Var(X)	MGF
U $\{x_1, \ldots, x_N\}$	$\frac{\#\{i:x_i \leqslant x\}}{N}$	$\frac{1}{N}$	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$	$\frac{e^t}{N}\left(\frac{e^{Nt}-1}{e^t-1}\right)$
Bernoulli (p)	$(1-p)^{1-x}$	$p^x(1-p)^{1-x}$	p	$p(1-p)$	$(1-p+pe^t)$
Bin (n,p)	$I_{1-p}(n-x, x+1)$[1]	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$	$(1-p+pe^t)^n$
Hyp (N,n,p)	-	$\frac{\binom{Np}{x}\binom{N-Np}{n-x}}{\binom{N}{n}}$	np	$np(1-p)\left(\frac{N-n}{N-1}\right)$	-
Geo (p)	$1-(1-p)^{x+1}$	$p(1-p)^x$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{1-(1-p)e^t}$
NB (n,p)	$I_p(n, x+1)$[1]	$\binom{n+x-1}{n-1}p^n(1-p)^x$	$\frac{n(1-p)}{p}$	$\frac{n(1-p)}{p^2}$	$\left(\frac{p}{1-(1-p)e^t}\right)^n$
Poisson (λ)	$\int\limits_{\lambda}^{\infty} \frac{e^{-t}t^x}{\Gamma(x+1)}dt$	$e^{-\lambda}\frac{\lambda^x}{x!}$	λ	λ	$e^{\lambda\left(e^t-1\right)}$

[1] $I_p(k, n-k+1) = \int_0^p \frac{t^{k-1}(1-t)^{n-k}}{B(k,n-k+1)}dt$ (Incomplete Beta Function)

Properties

Binomial

(1) **Mode:** $[(n+1)\,p]$ if $(n+1)\,p$ is not an integer, else $\{(n+1)\,p-1\}$ and $(n+1)\,p$.

(2) **Factorial Moment:** $\mu_{(r)} = (n)_r\,p^r$

(3) Bin (n,p) is symmetric iff $p = \frac{1}{2}$

(4) Variance of Bin (n,p) is minimum iff $p = \frac{1}{2}$ and minimum variance $= \frac{n}{4}$.

(5) $X, Y \overset{iid}{\sim} \text{Bin}\left(n, \frac{1}{2}\right) \implies P(X=Y) = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}$

Geometric

(1) X : number of trials required to get the 1^{st} success, then $E(X) = \frac{1}{p}$ and $Var(X) = \frac{1-p}{p^2}$

(2) **Lack of Memory:** $X \sim \text{Geo}\,(p) \iff P(X > m+n \,|\, X > m) = P(X \geqslant n),\ \forall m, n \in \mathbb{N}$

Negative Binomial

(1) **Mode:** $\left[\frac{(n-1)(1-p)}{p}\right]$ if $\frac{(n-1)(1-p)}{p}$ is not an integer, else $\left(\frac{(n-1)(1-p)}{p}-1\right)$, $\frac{(n-1)(1-p)}{p}$.

(2) NB $(n,p) \equiv$ Bin $(-n, P)$ where, $P = -\frac{1-p}{p}$

(**3**) Y : number of trials required to get the r^{th} success. Then -

$$P(Y=y)=\binom{y-1}{r-1}p^r\,(1-p)^{y-r},\ y=r,r+1,\ldots$$

Here, Y is discrete waiting time R.V. (Pascal Distribution)

(**4**) $X\sim \text{Bin}\,(n,p), Y\sim \text{NB}\,(r,p) \implies P(X\geq r)=P(Y\leq n)$

Poisson

(**1**) **Mode:** $[\lambda]$ if λ is not an integer, else $(\lambda-1)$ and λ.

1.5.2 Continuous

X	CDF	PDF	E(X)	Var(X)	MGF
$\text{U}\,(a,b)$	$\frac{x-a}{b-a}$	$\frac{I\{a<x<b\}}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
$\text{Gamma}\,(n,\theta)$	$\Gamma_x\,(n,\theta)^{[2]}$	$\frac{e^{-\frac{x}{\theta}}\,x^{n-1}}{\theta^n\Gamma(n)}$	$n\theta$	$n\theta^2$	$\frac{1}{(1-t\theta)^n}$
$\text{Exp}\,(\theta)$	$1-e^{-\frac{x}{\theta}}$	$\frac{1}{\theta}\,e^{-\frac{x}{\theta}}$	θ	θ^2	$\frac{1}{(1-t\theta)}$
$\text{Beta}\,(m,n)$	$I_x(m,n)$	$\frac{x^{m-1}(1-x)^{n-1}}{B(m,n)}$	$\frac{m}{m+n}$	$\frac{mn}{(m+n)^2(m+n+1)}$	-
$\text{Beta}_2\,(m,n)$	-	$\frac{1}{B(m,n)}\cdot\frac{x^{m-1}}{(1+x)^{m+n}}$	$\frac{m}{n-1}\ (n>1)$	$\frac{m(m+n-1)}{(n-2)(n-1)^2}\ (n>2)$	-
$\mathcal{N}\,(\mu,\sigma^2)$	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	$\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{t\mu+\frac{t^2\sigma^2}{2}}$
$\Lambda\,(\mu,\sigma^2)$	$\Phi\left(\frac{\ln x-\mu}{\sigma}\right)$	$\frac{1}{x\sigma\sqrt{2\pi}}\,e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$	$e^{\mu+\frac{1}{2}\sigma^2}$	$e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$	$\times$
$\mathcal{C}\,(\mu,\sigma)$	$\frac{1}{2}+\frac{1}{\pi}\tan^{-1}\left(\frac{x-\mu}{\sigma}\right)$	$\frac{\sigma}{\pi\{\sigma^2+(x-\mu)^2\}}$	$\times$	$\times$	$\times$
$\text{SE}\,(\mu,\sigma)$	$1-e^{-\left(\frac{x-\mu}{\sigma}\right)}$	$\frac{1}{\sigma}\,e^{-\left(\frac{x-\mu}{\sigma}\right)}$	$\mu+\sigma$	σ^2	$\frac{e^{t\mu}}{(1-t\sigma)}$
$\text{DE}\,(\mu,\sigma)$	$\begin{cases}\dfrac{1}{2}\,e^{\frac{x-\mu}{\sigma}} & ,x\leq\mu\\ 1-\dfrac{1}{2}\,e^{-\frac{x-\mu}{\sigma}} & ,x>\mu\end{cases}$	$\frac{1}{2\sigma}\,e^{-\frac{\lvert x-\mu\rvert}{\sigma}}$	μ	$2\sigma^2$	$\frac{e^{t\mu}}{(1-t^2\sigma^2)}$
$\text{Pareto}\,(x_0,\theta)$	$1-\left(\frac{x_0}{x}\right)^\theta$	$\frac{\theta x_0^\theta}{x^{\theta+1}}$	$\frac{\theta x_0}{\theta-1}\ (\theta>1)$	$\frac{\theta x_0^2}{(\theta-2)(\theta-1)^2}\ (\theta>2)$	-
$\text{Logistic}\,(\alpha,\beta)$	$\frac{1}{1+e^{-\left(\frac{x-\alpha}{\beta}\right)}}$	$\frac{1}{\beta}\,\frac{e^{\frac{x-\alpha}{\beta}}}{\left\{1+e^{\frac{x-\alpha}{\beta}}\right\}^2}$	α	$\frac{\beta^2\pi^2}{3}$	$\frac{\pi\beta t\;e^{t\alpha}}{\sin(\pi\beta t)}$

[2] $\Gamma_x\,(n,\theta)=\int_0^x\frac{e^{-\frac{t}{\theta}}\,t^{n-1}}{\theta^n\,\Gamma(n)}dt$ (Incomplete Gamma Function)

Properties

Uniform

(1) $\mu'_r = \frac{b^{r+1}-a^{r+1}}{(r+1)(b-a)}$

(2) $X \sim \mathrm{U}\,(0,n), n \in \mathbb{N} \implies X - [X] \sim \mathrm{U}\,(0,1)$

(3) *Classical & Geometric* definition of probability is based on 'Uniform distribution' over <u>discrete</u> & <u>continuous</u> space, respectively.

Gamma

(1) **Moments:** $\mu'_r = \theta^r \frac{\Gamma(n+r)}{\Gamma(n)}$, if $r > -n$

(2) **HM:** $(n-1)\,\theta$, if $n > 1$

(3) **Mode:** Mode is at $(n-1)\,\theta$, if $n > 1$; 0, if $n = 1$ and for $0 < n < 1$ no mode.

Exponential

(1) $\mu'_r = \theta^r r!$

(2) $\xi_p = -\theta \ln(1-p) \implies \text{Median} = \theta \ln 2$

(3) Mode is at $x = 0$, $\mathrm{MD}_\theta = \frac{2\theta}{e}$

(4) **Lack of Memory:** $X \sim \mathrm{Exp}\,(\theta) \iff P(X > m+n \mid X > m) = P(X > n),\ \forall m, n > 0$

(5) $\frac{F'(x)}{1-F(x)} = \text{constant } \forall x > 0 \iff X \sim \text{Exponential}$

(6) $X \sim \mathrm{Exp}\,(\lambda) \implies [X] \sim \mathrm{Geo}\left(p = 1 - e^{-\frac{1}{\lambda}}\right)$ and $X \perp\!\!\!\perp [X]$

(7) $X \sim \mathrm{DE}\,(\theta, 1) \implies P[X_{(1)} \le \theta \le X_{(n)}] = 1 - \left(\frac{1}{2}\right)^{n-1}$

Beta

(1) $\mu'_r = \frac{B(r+m,n)}{B(m,n)}$, if $r + m > 0$

(2) **HM:** $\frac{m-1}{m+n-1}$, if $m > 1$

(3) **Mode:** $\frac{m-1}{m+n-2}$, if $m > 1, n > 1$

(4) If $m = n$, median $= \frac{1}{2}$, $\forall n > 0$ and mode $= \frac{1}{2}$, if $n > 1$, else no mode.

(5) $\mathrm{Beta}\,(1,1) \overset{D}{\equiv} \mathrm{U}\,(0,1)$

Beta$_2$

(1) $\mu'_r = \frac{B(r+m,n-r)}{B(m,n)}$, if $-m < r < n$

(2) **HM:** $\frac{n}{m-1}$, if $m > 1$

(3) **Mode:** $\frac{m-1}{n+1}$, if $m > 1$, for $0 < m < 1$, no mode.

Normal

(**1**) median = mode = μ and bell-shaped (unimodal)

(**2**) $\mu_{2r-1} = 0,\ \mu_{2r} = (2\sigma^2)^r \frac{\Gamma(r+\frac{1}{2})}{\sqrt{\pi}} = \{(2r-1)\cdot(2r-3)\cdots 5\cdot 3\cdot 1\}\,\sigma^{2r}$

(**3**) $\text{MD}_\mu = \sigma\sqrt{\frac{2}{\pi}}$

(**4**) $\int t\,\phi(t)\,dt = -\phi(t) + c$

(**5**) For $x > 0,\ \left\{\frac{1}{x} - \frac{1}{x^3}\right\} < \frac{1-\Phi(x)}{\phi(x)} < \frac{1}{x} \implies 1-\Phi(x) \simeq \frac{\phi(x)}{x}$, for large x ($x > 3$)

(**6**) $X \sim \mathcal{N}(0,1) \implies E[X] = -\frac{1}{2}$

Lognormal

(**1**) $\mu_r' = e^{r\mu + \frac{1}{2}r^2\sigma^2}$

(**2**) **HM:** $e^{\mu - \frac{1}{2}\sigma^2}$, **GM:** e^μ, **Median:** e^μ, **Mode:** $e^{\mu-\sigma^2}$

$$\implies \text{Mean} > \text{Median} > \text{Mode} \implies \underline{\text{Positively skewed}}$$

(**3**) $X_i \overset{iid}{\sim} \Lambda(\mu, \sigma^2) \implies \text{GM}(\underset{\sim}{X}) \sim \Lambda(\mu, \frac{\sigma^2}{n})$

Cauchy

(**1**) μ_r' exists for $-1 < r < 1$

(**2**) Median = Mode = μ

1.5.3 Multivariate

A 'p'-component (dimensional) **Random Vector** (R.V.), $\underset{\sim}{X}^{p\times 1} = \begin{pmatrix} X_1 & X_2 & \cdots & X_p\end{pmatrix}'$ defined on $(\Omega, \mathcal{A})$ is a vector of p real-valued functions $X_1(\cdot), X_2(\cdot), \ldots, X_p(\cdot)$ defined on 'Ω' such that - $\{\omega : X_1(\omega) \le x_1, X_2(\omega) \le x_2, \cdots, X_p(\omega) \le x_p\} \in \mathcal{A},\ \forall \underset{\sim}{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_p\end{pmatrix}' \in \mathbb{R}^p$ is a random vector.

(**1**) **CDF:** $F_{\underset{\sim}{X}}(\underset{\sim}{x}) = P\big[\{\omega : X_1(\omega) \le x_1, X_2(\omega) \le x_2, \cdots, X_p(\omega) \le x_p\}\big],\ \forall \underset{\sim}{x} \in \mathbb{R}^p$

(a) $F_{\underset{\sim}{X}}(\underset{\sim}{x})$ is <u>non-decreasing</u> and <u>right continuous</u> w.r.t. each of $x_1, x_2, \ldots, x_p$.

(b) $F_{\underset{\sim}{X}}(+\infty, +\infty, \ldots, +\infty) = 1,\ \lim_{x_i \to -\infty} F_{\underset{\sim}{X}}(\underset{\sim}{x}) = 0,\ \forall i = 1(1)p$

(c) For $h_1, h_2, \ldots, h_p > 0$ -

$$P(x_1 < X_1 < x_1 + h_1, x_2 < X_2 < x_2 + h_2, \ldots, x_p < X_p < x_p + h_p) \ge 0$$

(**2**) $\sum_{i=1}^{p} F_{\underset{\sim}{X}}(x_i) - (p-1) \le F_{\underset{\sim}{X}}(\underset{\sim}{x}) \le \sqrt[p]{F_{X_1}(x_1)\,F_{X_2}(x_2)\cdots F_{X_p}(x_p)}$

(**3**) The distribution of any sub-vector is a marginal distribution. There are $\mathbf{(2^p - 1)}$ marginals.

(**4**) **Independence:** $\underset{\sim}{X}^{p\times 1} = \begin{pmatrix} \underset{\sim}{X}_{(1)} \\ \underset{\sim}{X}_{(2)} \end{pmatrix},\ \underset{\sim}{X}_{(1)} \perp\!\!\!\perp \underset{\sim}{X}_{(2)} \iff F_{\underset{\sim}{X}}(\underset{\sim}{x}) = F_{\underset{\sim}{X}_{(1)}}(\underset{\sim}{x}_{(1)})\cdot F_{\underset{\sim}{X}_{(2)}}(\underset{\sim}{x}_{(2)}),\ \forall \underset{\sim}{x} \in \mathbb{R}^p$

(5) $E(\underset{\sim}{a}'\underset{\sim}{X}) = \underset{\sim}{a}'\underset{\sim}{\mu},\ Var(\underset{\sim}{a}'\underset{\sim}{X}) = \underset{\sim}{a}'\,\Sigma\,\underset{\sim}{a},\ Cov(\underset{\sim}{a}'\underset{\sim}{X}, \underset{\sim}{b}'\underset{\sim}{X}) = \underset{\sim}{a}'\,\Sigma\,\underset{\sim}{b}$ for non-stochastic vectors $\underset{\sim}{a}, \underset{\sim}{b} \in \mathbb{R}^p$

(6) $E(A\underset{\sim}{X}) = A\underset{\sim}{\mu},\ D(A\underset{\sim}{X}) = A\Sigma A',\ Cov(A\underset{\sim}{X}, B\underset{\sim}{X}) = A\Sigma B'$ for non-stochastic matrices $A^{q\times p}, B^{r\times p}$

(7) $E\big[(\underset{\sim}{X} - \underset{\sim}{\alpha})'A(\underset{\sim}{X} - \underset{\sim}{\alpha})\big] = \text{trace}(A\Sigma) + (\underset{\sim}{\mu} - \underset{\sim}{\alpha})'A(\underset{\sim}{\mu} - \underset{\sim}{\alpha})$

(8) A matrix $\Sigma = (\sigma_{ij})$ is a dispersion matrix if and only if it is **n.n.d.**

(9) **Generalized variance:** $\det(\Sigma)$, where $\Sigma = E\left\{(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})'\right\} = E(\underset{\sim}{X}\underset{\sim}{X}') - \underset{\sim}{\mu}\underset{\sim}{\mu}' = D\left(\underset{\sim}{X}^{p\times 1}\right)$

(10) Σ is **p.d.** iff there is no $\underset{\sim}{a} \neq \underset{\sim}{0}$ for which $P(\underset{\sim}{a}'\underset{\sim}{X} = c) = 1$
Σ is **p.s.d.** iff there is a vector $\underset{\sim}{a} \neq \underset{\sim}{0}$ for which $P\big[\underset{\sim}{a}'(\underset{\sim}{X} - \underset{\sim}{\mu}) = 0\big] = 1$

(11) $\det(\Sigma) > 0 \implies$ Non-singular, $\det(\Sigma) = 0 \implies$ Singular Distribution

(12) $\Sigma = BB'$ for any dispersion matrix Σ, where B is n.n.d.

(13) Σ is p.d. $\implies \Sigma = BB',\ B$ is non-singular and let, $\underset{\sim}{Y} = B^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu}) \implies E(\underset{\sim}{Y}) = \underset{\sim}{0}, D(\underset{\sim}{Y}) = I_p$

(14) $\rho_{12\cdot 3} = \dfrac{\rho_{12} - \rho_{23}\rho_{31}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}}$

Multinomial

PMF: $(X_1, X_2, \ldots, X_k) \sim \text{Multinomial}\,(n; p_1, p_2, \ldots, p_k)$

(a) Singular -

$$f_{\underset{\sim}{X}}(x_1, x_2, \ldots, x_k) = \begin{cases} \dfrac{n!}{x_1!\,x_2!\cdots x_k!} p_1^{x_1} p_k^{x_k} \cdots p_k^{x_k} & , \text{if } \displaystyle\sum_{i=1}^{k} x_i = n, \sum_{i=1}^{k} p_i = 1 \\ 0 & , \text{otherwise} \end{cases}$$

(b) Non-singular - $\sum\limits_{i=1}^{k-1} x_i \leq n,\ \sum\limits_{i=1}^{k-1} p_i < 1 \qquad \left(x_k = n - \sum\limits_{i=1}^{k-1} x_i,\ p_k = 1 - \sum\limits_{i=1}^{k-1} p_i\right)$

Properties

(1) $E(X_i) = np_i,\ Cov(X_i, X_j) = \begin{cases} np_i(1-p_i) & , \text{if } i = j \\ -np_ip_j & , \text{if } i \neq j \end{cases}, \quad i, j = 1, 2, \ldots, k-1$

(2) $\rho_{ij} = \rho(X_i, X_j) = -\sqrt{\dfrac{p_i\,p_j}{(1-p_i)(1-p_j)}},\ i \neq j \quad \left(\text{As } \sum\limits_{i=1}^{k} X_i = n,\ X_i \uparrow \implies X_j \downarrow \text{ on an average}\right)$

(3) $\det(\Sigma) = n^{k-1}\det(D - \underset{\sim}{P}\underset{\sim}{P}') = n^{k-1}\det(D)(1 - \underset{\sim}{P}'D^{-1}\underset{\sim}{P}) \qquad \begin{array}{l} D = \text{diag}\,(p_1, p_2, \ldots, p_{k-1}) \\ \underset{\sim}{P} = (p_1, p_2, \ldots, p_{k-1})' \end{array}$

(4) $\underset{\sim}{X}^{\overline{k-1}\times 1} \sim \text{Multinomial}\,(n; p_1, \ldots, p_{k-1}),\ \sum\limits_{i=1}^{k-1} p_i < 1$

$$\implies X_1 \mid (X_2 = x_2, \ldots, X_{k-1} = x_{k-1}) \sim \text{Bin}\left(n - \sum_{i=2}^{k-1} x_i, \frac{p_1}{1 - \sum\limits_{i=2}^{k-1} p_i}\right)$$

$\implies$ the regression of X_1 on $X_2, X_3, \ldots, X_{k-1}$ is linear and the distribution is <u>heteroscedastic</u>.

(**5**) **MGF:** $E\left(e^{\underset{\sim}{t}'\underset{\sim}{X}}\right) = \left[1 + \sum_{i=1}^{k-1} p_i \left(e^{t_i} - 1\right)\right]^n$

Multiple Correlation

(**6**) For singular case, $\rho_{1\cdot 23\cdots k} = 1$

(**7**) For non-singular case, $\rho^2_{1\cdot 23\cdots\overline{k-1}} = \dfrac{p_1 \cdot \sum_{i=2}^{k-1} p_i}{(1-p_1)\left(1 - \sum_{i=2}^{k-1} p_i\right)}$

$$\rho_{12\cdot 34\cdots\overline{k-1}} = -\frac{\sqrt{p_1\, p_2}}{\sqrt{(1-p_2-p_3-\cdots-p_{k-1})}\sqrt{(1-p_1-p_3-\cdots-p_{k-1})}}$$

Bivariate Normal

$(X,Y) \sim \text{BN}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$

(**1**) $X \sim \mathcal{N}(\mu_1, \sigma_1^2),\ Y \mid X = x \sim \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_2}(x-\mu_1), \sigma_2^2(1-\rho^2)\right)$

(**2**) $Q(X,Y) = \frac{1}{1-\rho^2}\left\{\left(\frac{X-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{X-\mu_1}{\sigma_1}\right)\left(\frac{Y-\mu_2}{\sigma_2}\right) + \left(\frac{Y-\mu_2}{\sigma_2}\right)^2\right\} = U^2 + V^2 \sim \chi^2_2$

where, $U = \dfrac{Y - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(X-\mu_1)}{\sigma_2\sqrt{1-\rho^2}},\ V = \dfrac{X-\mu_1}{\sigma_1} \overset{iid}{\sim} \mathcal{N}(0,1)$

(**3**) (X,Y) is independent $\iff \rho = 0$

(**4**) $(X,Y) \sim \text{BN}(0,0,1,1,\rho)$

(a) $(X+Y) \perp\!\!\!\perp (X-Y) \implies \frac{X+Y}{X-Y} \sim \mathcal{C}\left(0, \sqrt{\frac{1+\rho}{1-\rho}}\right)$

(b) $E\{\max(X,Y)\} = \sqrt{\frac{1-\rho}{\pi}}$, PDF: $f_U(u) = 2\phi(u)\Phi\left(u\sqrt{\frac{1-\rho}{1+\rho}}\right)$

(c) $\rho(X^2, Y^2) = \rho^2$

(**5**) $(X,Y) \sim \text{BN}(0,0,1,1,0),\ Y_1 = X_1 \operatorname{sgn}(X_2),\ Y_2 = X_2 \operatorname{sgn}(X_1)$, where $\operatorname{sgn}(X) = \begin{cases} -1, & X < 0 \\ 1, & X > 0 \end{cases}$

$\implies (Y_1, Y_2) \nsim \text{BN},\ \rho(Y_1, Y_2) = \frac{2}{\pi}$

1.5.4 Truncated Distribution

Univariate

$F(x)$ be the CDF of X over the sample space $\mathfrak{X}$. Let, $A = (a, b] \subset \mathfrak{X}$, then the CDF of X over truncated space A is -

$$G(x) = P(X \le x \,|\, X \in A) = \begin{cases} 0 & , x \le a \\ \dfrac{F(x) - F(a)}{F(b) - F(a)} & , a < x \le b \\ 1 & , x > b \end{cases}$$

PMF/PDF: $\frac{f(x)}{P(X\in A)}$, $x \in A$

Results

- $E(X) = E(X|A) \cdot P(X \in A) + E(X|A^c) \cdot P(X \in A^c)$
- $X \sim \text{Geo}\,(p) \implies (X-k)\,|\,X \geq k \sim \text{Geo}\,(p)$
- Truncated Normal distribution is <u>platykurtic</u>.
- Truncated Cauchy distribution has <u>finite moments</u>.

Bivariate

(X,Y): bivariate R.V. with PDF, $f(x,y)$ over the sample space, $\mathfrak{X} \subseteq \mathbb{R}^2$. Let, $A \subset \mathfrak{X}$, then the PDF over the truncated space is -

$$g(x,y) = \frac{f(x,y)}{P\big[(X,Y) \in A\big]}, \text{ if } (x,y) \in A$$

- $\mu'_{r,s}(A) = E(X^r Y^s|A) = \iint\limits_A x^r y^s \, \frac{f(x,y)}{P\big[(X,Y)\in A\big]} \, dx\, dy$

1.6 Sampling Distributions

1.6.1 Chi-square, t, F

$\boldsymbol{\chi^2_n}$

(1) $E(X) = n$, $\text{Var}\,(X) = 2n$

(2) $\mu'_r = 2^r \cdot \dfrac{\Gamma(\frac{n}{2}+r)}{\Gamma(\frac{n}{2})}$, if $r > -\frac{n}{2}$

(3) $\chi^2_n \overset{D}{\equiv} \text{Gamma}\left(\frac{n}{2}, 2\right),\ n \in \mathbb{N}$

$\boldsymbol{t_n}$

(1) $E(X) = 0$ $(n > 1)$, $\text{Var}\,(X) = \frac{n}{n-2}$ $(n > 2)$

(2) $\mu'_{2r} = n^r \cdot \dfrac{\Gamma(\frac{1}{2}+r)}{\sqrt{\pi}} \cdot \dfrac{\Gamma(\frac{n}{2}-r)}{\Gamma(\frac{n}{2})}$, if $-1 < 2r < n$

(3) $t_1 \overset{D}{\equiv} \mathcal{C}\,(0,1)$

(4) $t_n^2 \overset{D}{\equiv} F_{1,n}$

$\boldsymbol{F_{n_1,n_2}}$

(1) $E(X) = \dfrac{n_2}{n_2-2}$ (if $n_2 > 2$), $\text{Var}\,(X) = \dfrac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ (if $n_2 > 4$)

(2) **Mode:** $\dfrac{n_2(n_1-2)}{n_1(n_2+2)}$ (if $n_1 > 2$) $\implies$ Mean $> 1 >$ Mode (if $n_1, n_2 > 2$)

(3) $\mu_r' = \left(\frac{n_2}{n_1}\right)^r \cdot \frac{\Gamma(\frac{n_1}{2}+r)}{\Gamma(\frac{n_1}{2})} \cdot \frac{\Gamma(\frac{n_2}{2}-r)}{\Gamma(\frac{n_2}{2})}$, if $-n_1 < 2r < n_2$

(4) ξ_p and ξ_p' are p^{th} quantile of F_{n_1,n_2} and F_{n_2,n_1} respectively $\implies \xi_p \, \xi_{1-p}' = 1$

(5) $F \sim F_{n,n} \implies F \stackrel{D}{\equiv} \frac{1}{F}$ and median $(F) = 1$

(6) $F \sim F_{n_1,n_2} \implies \frac{n_1}{n_2} F \sim \text{Beta}_2\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$

(7) Points of inflexion are equidistant from mode (if $n > 4$)

1.6.2 Order Statistics

Order Statistics: $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$

(1) $F_{X_{(r)}}(x) = \sum\limits_{k=r}^{n} \binom{n}{k} \{F(x)\}^k \{1-F(x)\}^{n-k} = I_{F(x)}(r, n-r+1)$

(2) $F_{X_{(1)},X_{(n)}}(x_1, x_2) = \{F(x_2)\}^n - \{F(x_2) - F(x_1)\}^n, \; x_1 < x_2$

Only for Absolutely Continuous Random Variable - CDF: $F(x)$, PDF: $f(x)$

(3) $f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} \{F(x)\}^{r-1} f(x) \{1-F(x)\}^{n-r}, \; x \in \mathbb{R}$

(4) Joint PDF: $\frac{n!}{(r-1)!(s-r-1)!(n-r)!} \{F(x)\}^{r-1} f(x) \{F(y)-F(x)\}^{s-r-1} f(y) \{1-F(y)\}^{n-s}, \quad \begin{matrix} x < y \\ (r < s) \end{matrix}$

(5) **Sample Range:** $f_R(r) = n(n-1) \int\limits_{-\infty}^{\infty} \{F(r+s) - F(s)\}^{n-2} f(r+s) f(s)\, ds, \; 0 < r < \infty$

Results

(6) $X_i \stackrel{iid}{\sim} \text{U}(0,1) \implies X_{(r)} \sim \text{Beta}\,(r, n-r+1)$, $r = 1, 2, \ldots, n$

(7) $X_1, X_2 \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2) \implies E\big[X_{(1)}\big] = \mu - \frac{\sigma}{\sqrt{\pi}}, \; E\big[X_{(2)}\big] = \mu + \frac{\sigma}{\sqrt{\pi}}$

(8) $X_1, X_2, X_3 \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2) \implies$ Sample Range: $\frac{1}{2}\left(|X_1 - X_2| + |X_2 - X_3| + |X_3 - X_1|\right)$

(9) $X_i \stackrel{iid}{\sim} \text{Exp}\,(\theta) \implies E[X_{(n)}] = \theta\left(1 + \frac{1}{2} + \frac{1}{3} \cdots + \frac{1}{n}\right)$

(10) $X_i \stackrel{iid}{\sim} \text{Exp}\,(\theta) \implies U_i = X_{(i)} - X_{(i-1)} \stackrel{\perp\!\!\!\perp}{\sim} \text{Exp}\left(\frac{\theta}{n-i+1}\right), \; X_{(0)} = 0$

(11) $X_1, X_2, \ldots, X_{2k+1}$: random sample from a **continuous** distribution, **symmetric about $\boldsymbol{\mu}$**
$\implies$ Distribution of $\tilde{X}$ is also symmetric about μ
$\implies E(\tilde{X}) = \mu, \; E\left(\frac{X_{(1)}+X_{(n)}}{2}\right) = \mu$ (if exists)

(12) $X_i \stackrel{iid}{\sim} \text{Shifted Exp}\,(\theta, 1) \implies (n-i+1)\left(X_{(i)} - X_{(i-1)}\right) \stackrel{iid}{\sim} \text{Exp}\,(1)$
$\implies 2n\left[X_{(1)} - \theta\right] \sim \chi^2_2 \perp\!\!\!\perp 2\sum\limits_{i=2}^{n}\left[X_{(i)} - X_{(1)}\right] \sim \chi^2_{2n-2}$

1.7 Distribution Relationships

1.7.1 Binomial

(1) $X_i \overset{iid}{\sim} \text{Bernoulli}\,(p) \implies \sum\limits_{i=1}^{n} X_i \sim \text{Bin}\,(n,p)$

(2) $X_i \overset{\perp\!\!\!\perp}{\sim} \text{Bin}\,(n_i,p) \implies \sum\limits_{i=1}^{k} X_i \sim \text{Bin}\left(\sum\limits_{i=1}^{k} n_i, p\right)$

(3) $X_1, X_2 \overset{iid}{\sim} \text{Bin}\,(n, \frac{1}{2}) \implies X_1 - X_2$ is symmetric about '0'.

(4) $X_i \overset{\perp\!\!\!\perp}{\sim} \text{Bin}\,(n_i,p) \implies X_k \Big| \sum\limits_{i=1}^{m} X_i = t \sim \text{Hyp}\left(N = \sum\limits_{i=1}^{k} n_i, t, \frac{n_k}{N}\right), k = 1(1)m$

(5) $\text{Bin}\,(n,p) \to \text{Poisson}\,(\lambda = np)$, for $n \to \infty$ and $p \to 0$ such that np is finite.

(6) $\text{Bin}\,(n,p) \to \mathcal{N}\big(np, np(1-p)\big)$, for large n and moderate p.

1.7.2 Negative Binomial

(1) $X_i \overset{iid}{\sim} \text{Geo}\,(p) \implies X_{(1)} \sim \text{Geo}\,(1-q^n)$, where $q = 1-p$

(2) $X, Y \overset{iid}{\sim} \text{Geo}\,(p) \iff X \big| X + Y = t \sim \text{U}\,\{0,1,2,\ldots,t\}$

(3) $X, Y \overset{iid}{\sim} \text{Geo}\,(p) \implies \min\{X,Y\} \perp\!\!\!\perp (X - Y)$

(4) $X_i \overset{iid}{\sim} \text{Geo}\,(p) \implies \sum\limits_{i=1}^{n} X_i \sim \text{NB}\,(n,p)$

(5) $X_i \overset{\perp\!\!\!\perp}{\sim} \text{NB}\,(n_i,p) \implies \sum\limits_{i=1}^{k} X_i \sim \text{NB}\left(\sum\limits_{i=1}^{k} n_i, p\right)$

(6) $\text{NB}\,(n,p) \to \text{Poisson}\,(\lambda = n(1-p))$, for $n \to \infty$ and $p \to 1$ such that $n(1-p)$ is finite.

1.7.3 Poisson

(1) $X_i \overset{\perp\!\!\!\perp}{\sim} \text{Poisson}\,(\lambda_i) \implies \sum\limits_{i=1}^{k} X_i \sim \text{Poisson}\left(\sum\limits_{i=1}^{k} \lambda_i\right)$

(2) $X_i \overset{\perp\!\!\!\perp}{\sim} \text{Poisson}\,(\lambda_i) \implies X_k \Big| \sum\limits_{i=1}^{m} X_i = t \sim \text{Bin}\left(t, p = \frac{\lambda_k}{\lambda}\right), k = 1(1)m$, where $\lambda = \sum\limits_{i=1}^{m} \lambda_i$

(3) $X_i \overset{\perp\!\!\!\perp}{\sim} \text{Poisson}\,(\lambda_i) \implies (X_1, X_2, \ldots, X_k) \Big| \sum\limits_{i=1}^{k} X_i = t \sim \text{Multinomial}\,(t, p_1, p_2, \ldots, p_k)$, where $p_i = \dfrac{\lambda_i}{\sum\limits_{i=1}^{k} \lambda_i},\ i = 1, 2, \ldots, k$

1.7.4 Normal

(1) $X \sim \mathcal{N}(0, \sigma^2) \implies X \overset{D}{\equiv} -X$

(2) $X_i \overset{\perp\!\!\!\perp}{\sim} \mathcal{N}(\mu_i, \sigma_i^2) \implies \sum\limits_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum\limits_{i=1}^{n} a_i \mu_i, \sum\limits_{i=1}^{n} a_i^2 \sigma_i^2\right)$

(3) $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2),\ \sum\limits_{i=1}^{n} a_i X_i \perp\!\!\!\perp \sum\limits_{i=1}^{n} b_i X_i \iff \underset{\sim}{a}.\underset{\sim}{b} = 0$

$\bar{X}$ and $(X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X})$ are independently distributed

1.7.5 Gamma

(1) $X_i \overset{iid}{\sim} \text{Exp}\,(\theta) \implies X_{(1)} \sim \text{Exp}\left(\frac{\theta}{n}\right)$

(2) $X, Y \overset{iid}{\sim} \text{Exp}\,(\theta) \implies X \Big| X + Y = t \sim \text{U}(0, t)$

(3) $X \sim \text{Shifted Exp}\,(\mu, \theta) \implies (X - \mu) \sim \text{Exp}\,(\theta)$

(4) $X \sim \text{DE}\,(\mu, \sigma) \implies \left|\frac{X-\mu}{\sigma}\right| \sim \text{Exp}\,(\theta = 1)$ and $|X| \sim \text{Shifted Exp}\,(\mu, \sigma)$

(5) $X_i \overset{iid}{\sim} \text{Exp}\,(\theta) \equiv \text{Gamma}\,(n = 1, \theta) \implies \sum\limits_{i=1}^{n} X_i \sim \text{Gamma}\,(n, \theta)$

(6) $X_i \overset{\perp\!\!\!\perp}{\sim} \text{Gamma}\,(n_i, \theta) \implies \sum\limits_{i=1}^{n} X_i \sim \text{Gamma}\left(\sum\limits_{i=1}^{k} n_i, \theta\right)$

1.7.6 Beta

(1) $X \sim \text{Beta}\,(m, n) \implies \frac{X}{1-X} \sim \text{Beta}_2\,(m, n)$

(2) $X \sim \text{Beta}_2\,(m, n) \implies \frac{X}{1+X} \sim \text{Beta}\,(m, n)$

(3) $X_1 \sim \text{Beta}\,(n_1, n_2)$ & $X_2 \sim \text{Beta}\,(n_1 + \frac{1}{2}, n_2)$, independently $\implies \sqrt{X_1 X_2} \sim \text{Beta}\,(2n_1, 2n_2)$

1.7.7 Cauchy

(1) $X_i \overset{iid}{\sim} \mathcal{C}\,(\mu, \sigma) \implies \bar{X}_n \sim \mathcal{C}(\mu, \sigma)$

(2) $X_i \overset{iid}{\sim} \mathcal{C}\,(0, \sigma) \implies \frac{1}{X_i} \overset{iid}{\sim} \mathcal{C}\left(0, \frac{1}{\sigma}\right) \implies HM_{\underset{\sim}{X}} \sim \mathcal{C}(0, \sigma)$

(3) $X_i \overset{\perp\!\!\!\perp}{\sim} \mathcal{C}\,(\mu_i, \sigma_i) \implies \sum\limits_{i=1}^{n} X_i \sim \mathcal{C}\left(\sum\limits_{i=1}^{n} \mu_i, \sum\limits_{i=1}^{n} \sigma_i\right)$

1.7.8 Others

(1) $X_i \overset{iid}{\sim} \mathcal{N}\,(0, 1) \implies \sum\limits_{i=1}^{n} X_i^2 \sim \chi_n^2$

(2) $U \sim \mathcal{N}\,(0, 1),\ V \sim \chi_n^2$, independently $\implies \frac{U}{\sqrt{V/n}} \sim t_n$

(3) $U_i \overset{\perp\!\!\!\perp}{\sim} \chi_{n_i}^2 \implies \frac{U_1/n_1}{U_2/n_2} \sim F_{n_1, n_2}$

(4) X is symmetric about '0' $\implies X \overset{D}{\equiv} -X$

1.8 Transformations

1.8.1 Orthogonal

$\underset{\sim}{y} = T(\underset{\sim}{x}) = A^{n\times n}\underset{\sim}{x}^{n\times 1} \rightarrow$ Linear Transformation. [If $\det(A) \neq 0$, Jacobian: $J = \det(A^{-1})$].

(1) If $T(\underset{\sim}{x})$ is orthogonal transformation then $A^TA = I_n \implies \det(A) = \pm 1$ & $|J| = 1$

(2) $\underset{\sim}{y}^T\underset{\sim}{y} = \underset{\sim}{x}^T\underset{\sim}{x} \implies |\underset{\sim}{y}|^2 = |\underset{\sim}{x}|^2$ (length is preserved)

(3) **Cochran's theorem:** $X_i \overset{iid}{\sim} \mathcal{N}(0,1)$ & $\sum\limits_{i=1}^{n} X_i^2 = \underset{\sim}{X}^T A_1 \underset{\sim}{X} + \underset{\sim}{X}^T A_2 \underset{\sim}{X}$, where A_1, A_2 are n.n.d. matrices with ranks $r_1, r_2, r_1 + r_2 = n$

$$\implies \underset{\sim}{X}^T A_1 \underset{\sim}{X} \sim \chi^2_{r_1} \text{ and } \underset{\sim}{X}^T A_2 \underset{\sim}{X} \sim \chi^2_{r_2}, \text{ independently.}$$

1.8.2 Polar

(1) For a point with Cartesian coordinates $(x_1, x_2, \dots, x_n)$ in $\mathbb{R}^n$ -

$$\begin{aligned}
x_1 &= r\cos\theta_1 \\
x_2 &= r\sin\theta_1\cos\theta_2 \\
x_3 &= r\sin\theta_1\sin\theta_2\cos\theta_3 \\
&\vdots \\
x_{n-1} &= r\sin\theta_1\cdots\sin\theta_{n-2}\cos\theta_{n-1} \\
x_n &= r\sin\theta_1\cdots\sin\theta_{n-2}\sin\theta_{n-1}
\end{aligned}$$

where, $r^2 = \sum\limits_{i=1}^{n} x_i^2$, $\quad 0 < r < \infty$ and $0 < \theta_1, \theta_2, \dots, \theta_{n-2} < \pi,\ 0 < \theta_{n-1} < 2\pi$

Jacobian: $|J| = r^{n-1}(\sin\theta_1)^{n-2}(\sin\theta_2)^{n-3}\cdots\sin\theta_{n-2}$

(2) $X = R\cos\theta, Y = R\sin\theta,\ 0 < R < \infty, 0 < \theta < 2\pi$

$$X, Y \overset{iid}{\sim} \mathcal{N}(0,1) \iff \begin{matrix} \theta \sim \mathrm{U}(0, 2\pi) \\ R^2 \sim \mathrm{Exp}(2) \overset{D}{\equiv} \chi^2_2 \end{matrix}, \text{ independently.}$$

(3) $\theta \sim \mathrm{U}(0, 2\pi) \perp\!\!\!\perp R^2 \sim \chi^2_2 \implies R\sin(\theta + \theta_0) \sim \mathcal{N}(0,1),\ \theta_0$ is a fixed quantity

1.8.3 Special Transformations

(1) $X \sim \mathrm{U}(a,b) \implies -\ln\left(\frac{X-a}{b-a}\right) \sim \mathrm{Exp}(1)$

(2) $X_1, X_2 \overset{iid}{\sim} \mathrm{U}(0,1) \implies X_1 + X_2 \sim$ Triangular $(0,2)$, $|X_1 - X_2| \sim$ Beta $(1,2)$

(3) **Box-Muller Transformation:**

$$X_1, X_2 \overset{iid}{\sim} \mathrm{U}(0,1) \implies \begin{matrix} Y_1 = \sqrt{-2\ln X_1}\cos(2\pi X_2) \\ Y_2 = \sqrt{-2\ln X_1}\sin(2\pi X_2) \end{matrix} \quad \overset{iid}{\sim} \mathcal{N}(0,1)$$

(4) $X \sim \text{Gamma}\,(n_1, \theta), Y \sim \text{Gamma}\,(n_2, \theta)$

$\implies X + Y \sim \text{Gamma}\,(n_1 + n_2, \theta),\ \frac{X}{X+Y} \sim \text{Beta}\,(n_1, n_2)$, independently

$\implies X + Y \sim \text{Gamma}\,(n_1 + n_2, \theta),\ \frac{X}{Y} \sim \text{Beta}_2\,(n_1, n_2)$, independently

(5) $X, Y \overset{iid}{\sim} \mathcal{N}(0,1) \implies \frac{X}{Y}, \frac{X}{|Y|} \sim \mathcal{C}(0,1)$

(6) $X_1, X_2 \overset{iid}{\sim} \text{Exp}\,(\theta),\ R \sim \text{Bernoulli}\left(\frac{1}{2}\right) \implies (X_1 - X_2), RX_1 - (1-R)X_2 \sim \text{DE}\left(0, \frac{1}{\theta}\right)$

(7) $X \sim \text{Beta}\,(a, b) \perp\!\!\!\perp Y \sim \text{Beta}\,(a+b, c) \implies XY \sim \text{Beta}\,(a, b+c)$

(8) $X \sim \text{U}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \iff \tan X \sim \mathcal{C}(0,1)$

(9) **Dirichlet Transformation:** $X_i \overset{iid}{\sim} \text{Exp}\,(\theta)$

$$\implies Y_1 = \sum_{i=1}^{n} X_i \sim \text{Gamma}\,(n, \theta),\ Y_k = \frac{\sum\limits_{i=1}^{n-k+1} X_i}{\sum\limits_{i=1}^{n-k+2} X_i} \sim \text{Beta}\,(n-k+1, 1),\ k = 2, 3, \ldots, n$$

$Y_1, Y_2, \ldots, Y_n$ are independently distributed.

(10) $X_1, X_2, X_3, X_4 \overset{iid}{\sim} \mathcal{N}\,(0,1) \implies X_1X_2 \pm X_3X_4 \sim \text{DE}\,(0,1)$ (valid for any combination)

(11) $X_i \overset{iid}{\sim} \text{Exp}\,(\theta) \implies \frac{2}{\theta} \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$

(12) $X \sim \text{Beta}\,(\theta, 1) \implies -\ln X \sim \text{Exp}\,\left(\frac{1}{\theta}\right)$

(13) $X \sim \text{Pareto}\,(\theta, x_0) \implies \ln\left(\frac{X}{x_0}\right) \sim \text{Exp}\,\left(\frac{1}{\theta}\right)$

(14) $X_1, X_2 \sim \chi^2_2 \implies \dfrac{aX_1 + bX_2}{X_1 + X_2} \sim \text{U}(a, b)$ $(a < b)$

Chapter 2

Statistics

2.1 Point Estimation

2.1.1 Minimum MSE

(**1**) **Measures of Closeness:** T: Statistic/Estimator, $\psi(\theta)$: Parametric function

Destroying the randomness, general measures of closeness are -

(a) $E|T-\theta|^r$, for some $r>0$ (smaller value is better)

(b) $P\big[|T-\theta|<\epsilon\big]$, for $\epsilon>0$ (higher value is better)

(**2**) **Mean Square Error:** $\text{MSE}_{\psi(\theta)}(T) = E[T-\psi(\theta)]^2 = Var(T) + \big[b\left(\psi(\theta),T\right)\big]^2$

T can be said a 'good estimator' of $\psi(\theta)$ if it has a small variance.

(**3**) $E(m'_r) = \mu'_r$, if μ'_r exists $\implies E(\bar{X}) = \mu$, provided $\mu = E(X_1)$ exists

(**4**) $E\left(s^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2\right) = \sigma^2$, population variance (if exists) but $E(m_2)\neq\sigma^2=\mu_2$

(**5**) $X_i \overset{iid}{\sim} \mathcal{N}(0,\sigma^2) \implies E\left(T_1 = \sqrt{\frac{\pi}{2}}\cdot\frac{1}{n}\sum\limits_{i=1}^{n}|X_i|\right) = \sigma^2 = E\left(T_2 = C_n\sqrt{\sum\limits_{i=1}^{n}X_i^2}\right)$.

Here, T_1, T_2 are two UEs of σ^2 based on $\sum\limits_{i=1}^{n}|X_i|$ and $\sum\limits_{i=1}^{n}X_i^2$, respectively. $\left(C_n = \dfrac{\Gamma(\frac{n}{2})}{\sqrt{2}\,\Gamma(\frac{n+1}{2})}\right)$

(**6**) $X_i \overset{iid}{\sim} \text{Exp}\,(\theta) \implies E(\bar{X}) = \theta, E\left(\frac{n-1}{n\bar{X}} = \frac{1}{\theta}\right)$

(**7**) $X_i \overset{iid}{\sim} \mathcal{N}(\mu,\sigma^2) \implies T' = \frac{n-1}{n+1}\cdot s^2$ has the smallest MSE in the class $\{bs^2 : b>0\}$ i.e. a biased estimator T' is better than an UE s^2, in terms of MSE.

(**8**) $X \sim \text{Poisson}\,(\lambda) \implies T(X) = (-1)^X$ is the UMVUE of $e^{-2\lambda}$ which is an absurd UE.

Note: Absurd unbiased estimator is that unbiased estimator which can take values outside the parameter space.

(9) $X_i \overset{iid}{\sim} \text{Poisson}\,(\lambda) \implies T_\alpha = \alpha\bar{X} + (1-\alpha)s^2$ is an UE of λ for any $\alpha \in [0,1]$
$\implies$ There may be infinitely many UEs

(10) Estimable Parametric Functions

(a) $X \sim \text{Bin}\,(n,p) \implies E[(X)_r] = (n)_r\, p^r,\ r = 1, 2, \ldots, n$
$\implies$ Only polynomials of degree $\leq$ n are estimable.

(b) $X \sim \text{Bernoulli}\,(p) \implies$ Only $\psi(p) = a + bp$ is estimable.

(c) $X \sim \text{Poisson}\,(\lambda) \implies E[(X)_r] = \lambda^r,\ r = 1, 2, \ldots \implies e^{-\lambda}$ is estimable but not $\frac{1}{\lambda}, \sqrt{\lambda}$.

(11) $X \sim \text{Bernoulli}\,(\theta), T_1(X) = X$ and $T_2(X) = \frac{1}{2}$
$\implies$ Between T_1 and T_2 none are uniformly better than the other, in terms of MSE.

(12) $X_i \overset{iid}{\sim} f(x;\theta), E\big[T(X_1)\big] = \theta, Var\big[T(X_1)\big] < \infty$
$\implies \lim\limits_{n\to\infty} Var(S_n) = 0$, where S_n is the UMVUE of θ

(13) Best Linear Unbiased Estimator (BLUE)

$T_1, T_2, \ldots, T_k$ be UEs of $\psi(\theta)$ with known variances $v_1, v_2, \ldots, v_k$ and are independent

$$\implies \text{BLUE of } \psi(\theta) : T = \frac{1}{\sum\limits_{i=1}^{k} \frac{1}{v_i}} \sum_{i=1}^{k} \frac{T_i}{v_i}$$

2.1.2 Consistency

$$T_n \text{ is consistent for } \theta \iff \begin{matrix} P\big[|T_n - \theta| < \epsilon\big] \to 1 \\ \text{or} \\ P\big[|T_n - \theta| > \epsilon\big] \to 0 \end{matrix} \quad \text{as } n \to \infty,\ \forall\theta \in \Omega \text{ for every } \epsilon > 0$$

(1) Sufficient Condition

$E(T_n - \theta)^2 \to 0 \iff E(T_n) \to \theta, Var(T_n) \to 0$ as $n \to \infty \implies T_n \xrightarrow{P} \theta$

(2) $m'_r = \frac{1}{n}\sum\limits_{i=1}^{n} x_i^r \xrightarrow{P} \mu'_r = E(X_1^r),\ r = 1, 2, \ldots, k$ (if k^{th} order moment exists)

(3) If $T_n \xrightarrow{P} \theta$ then -

(a) $b_n T_n \xrightarrow{P} \theta$, if $b_n \to 1$ as $n \to \infty$

(b) $a_n + T_n \xrightarrow{P} \theta$, if $a_n \to 0$ as $n \to \infty$

This also shows that, 'unbiasedness' and 'consistency' are <u>not interrelated</u>.

(4) Invariance Property: $T_n \xrightarrow{P} \theta \implies \psi(T_n) \xrightarrow{P} \psi(\theta)$, provided $\psi(\cdot)$ is <u>continuous</u>

2.1.3 Sufficiency

S is sufficient for $\theta \iff (X_1, X_2, \ldots, X_n) \mid S = s$ is independent of θ, $\forall\, s$
S is sufficient for $\theta \iff T \mid S = s$ is independent of θ, $\forall\, s$, for all statistic T.

(**1**) Any one-to-one function of a sufficient statistic is also sufficient for a parameter.

(**2**) **Factorization Theorem**

$$\prod_{i=1}^{n} f(x_i;\theta) = g\big(T(\underset{\sim}{x});\theta\big)\cdot h(\underset{\sim}{x}) \iff T(\underset{\sim}{X}) \text{ is sufficient for } \theta$$

where, $g\big(T(\underset{\sim}{x});\theta\big)$ depends on θ and on $\underset{\sim}{x}$ only through $T(\underset{\sim}{x})$ and $h(\underset{\sim}{x})$ is independent of θ.

(**3**) **Trivial Sufficient Statistic:** $(X_1, X_2, \ldots, X_n)$ and $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$.

Sufficiency means **"space reduction without losing any information"**. In this aspect, the order statistics, $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is better as a sufficient statistic than the whole sample i.e. $(X_1, X_2, \ldots, X_n)$, with respect to data summarization.

(**4**) T_1, T_2 are two sufficient statistic for $\theta \implies$ they are related

(**5**) $X_i \overset{iid}{\sim} \text{DE}\,(\mu, \sigma)$, $\exists$ non-trivial sufficient statistic if μ is known (say, μ_0) and that is $\sum\limits_{i=1}^{n} |X_i - \mu_0|$.

Minimal Sufficient Statistic

(**6**) T_0 is a minimal sufficient of θ if,

(a) T_0 is sufficient

(b) T_0 is a function of every sufficient statistic

(**7**) **Theorem:** *For two sample points $\underset{\sim}{x}$ and $\underset{\sim}{y}$, the ratio $\frac{f(\underset{\sim}{x};\theta)}{f(\underset{\sim}{y};\theta)}$ is independent of θ if and only if $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$, then $T(\underset{\sim}{X})$ is minimal sufficient for θ.*

2.1.4 Completeness

T is complete for $\theta \iff$ "$E[h(T)] = 0, \forall\theta \in \Omega \implies P[h(T) = 0] = 1, \forall\theta \in \Omega$"

Remark

If a two component statistic (T_1, T_2) is minimal sufficient for a single component parameter θ, then in general (T_1, T_2) is not complete.

It is possible to find $h_1(T_1)$ and $h_2(T_2)$ such that,

$$E[h_1(T_1)] = \psi(\theta) = E[h_2(T_2)],\ \forall\theta$$

$$\implies E[h(T_1, T_2)] = 0,\ \forall\theta \quad \text{where, } h(T_1, T_2) = h_1(T_1) - h_2(T_2) \neq 0$$

$$\implies (T_1, T_2) \text{ is not complete.}$$

2.1.5 Exponential Family

One Parameter

An one parameter family of PDFs or PMFs, $\{f(x;\theta):\theta\in\Omega\}$ that can be expressed in the form -

$$f(x;\theta)=\exp\left[T(x)u(\theta)+v(\theta)+w(x)\right],\ x\in\mathcal{S}$$

with the following regularity conditions -

$\boldsymbol{C_1}$ **:** The support, $\mathcal{S}=\{x:f(x;\theta)>0\}$ is independent of θ

$\boldsymbol{C_2}$ **:** The parameter space, Ω is an open interval in $\mathbb{R}$ i.e. $\Omega=\{\theta:a<\theta<b\}$

$\boldsymbol{C_3}$ **:** $\{1,T(x)\}$ and $\{1,u(\theta)\}$ are linearly independent i.e. $T(x)$ and $u(\theta)$ are non-constant functions

is called One Parameter Exponential Family (OPEF)

K Parameter

A K-parameter family of PDFs or PMFs, $\{f(x;\underset{\sim}{\theta}):\underset{\sim}{\theta}\in\Omega\subseteq\mathbb{R}^k\}$ satisfying the form -

$$f(x;\underset{\sim}{\theta})=\exp\left[\sum_{i=1}^{k}T_i(x)u_i(\underset{\sim}{\theta})+v(\underset{\sim}{\theta})+w(x)\right],\ x\in\mathcal{S}$$

with the following regularity conditions -

$\boldsymbol{C_1}$ **:** The support, $\mathcal{S}=\{x:f(x;\underset{\sim}{\theta})>0\}$ is independent of $\underset{\sim}{\theta}$

$\boldsymbol{C_2}$ **:** The parameter space, $\Omega\subseteq\mathbb{R}^k$ is an open rectangle in $\mathbb{R}^k$ i.e. $a_i<\theta_i<b_i,\ i=1(1)k$

$\boldsymbol{C_3}$ **:** $\{1,T_1(x),\ldots,T_k(x)\}$ and $\{1,u_1(\underset{\sim}{\theta}),\ldots,u_k(\underset{\sim}{\theta})\}$ are linearly independent

is called K-parameter Exponential Family

Theorem

(a) $\underset{\sim}{X}\overset{iid}{\sim}f(x;\theta)\in\text{OPEF}\implies\sum\limits_{i=1}^{n}T(X_i)$ is complete and sufficient for the family.

(b) $\underset{\sim}{X}\overset{iid}{\sim}f(x;\underset{\sim}{\theta})\in$ K-parameter Exponential Family $\implies\left(\sum\limits_{i=1}^{n}T_1(X_i),\ldots,\sum\limits_{i=1}^{n}T_k(X_i)\right)$ is complete and sufficient for the family.

Distributions in Exponential Family

(a) $f(x;\theta)=\frac{a(x)\,\theta^x}{g(\theta)},\ x=0,1,2,\ldots;\ 0<\theta<\rho,\ a(x)\geq 0,\ g(\theta)=\sum\limits_{x=0}^{\infty}a(x)\,\theta^x$ (Power Series)

$\implies$ Binomial (n known), Poisson, Negative Binomial (n known) are in OPEF.

(b) Normal, Exponential, Gamma, Beta, Pareto (x_0 known) are in the Exponential family.

(c) Uniform, Cauchy, Laplace, Shifted Exponential, $\{\mathcal{N}(\theta,\theta^2):\theta\neq 0\}$, $\{\mathcal{N}(\theta,\theta):\theta>0\}$ are not in the Exponential Family.

The last two families are identified by Lehmann as 'Curved Exponential Family'.

2.1.6 Methods of finding UMVUE

Theorem 2.1.6.1 (Necessary & Sufficient Condition for UMVUE) *Let X has a distribution from $\{f(x;\theta) : \theta \in \Omega\}$. Define, $U_\psi = \Big\{T(X) : E_\theta\big[T(X)\big] = \psi(\theta), Var_\theta\big[T(X)\big] < \infty, \forall\theta \in \Omega\Big\}$ and $U_0 = \Big\{u(X) : E_\theta\big[T(X)\big] = 0, Var_\theta\big[u(X)\big] < \infty, \forall\theta \in \Omega\Big\}$. Then, $T_0 \in U_\psi$ is UMVUE of θ if and only if $Cov_\theta(T_0, u) = 0, \forall\theta \in \Omega, \forall u \in U_0$*

Results

- UMVUE if exists, is unique
- T_i is UMVUE of $\psi(\theta) \implies \sum\limits_{i=1}^{k} a_i T_i$ is UMVUE of $\sum\limits_{i=1}^{k} a_i \psi_i(\theta)$
- T is UMVUE $\implies T^k$ is UMVUE $\implies$ any polynomial function, $f(T)$ is UMVUE of their expectations

Theorem 2.1.6.2 (Rao-Blackwell) *Let X has a distribution from $\{f(x;\theta) : \theta \in \Omega\}$ and h be a statistic from $U_\psi = \{h : E(h) = \psi(\theta), Var(h) < \infty, \forall\theta \in \Omega\}$. Let, T be a sufficient statistic for θ. Then-*

(a) $E(h \mid T)$ is an UE of $\psi(\theta)$

(b) $Var\big[E(h \mid T)\big] \le Var(h), \forall\theta \in \Omega$

Implication: UMVUE is necessarily a function of minimal sufficient statistic

Theorem 2.1.6.3 (Lehmann-Scheffe) *Let X has a distribution from $\{f(x;\theta) : \theta \in \Omega\}$ and T be a complete sufficient statistic for θ. Then-*

(a) If $E\big[h(T)\big] = \psi(\theta)$, then UMVUE of $\psi(\theta)$ is the unique UE, $h(T)$

(b) If h^* is an UE of $\psi(\theta)$, then $E(h^* \mid T)$ is the UMVUE of $\psi(\theta)$

UMVUE of Different Families

Binomial

$X_i \overset{iid}{\sim} \text{Bernoulli}\,(p) \implies$ Complete Sufficient: $T = \sum\limits_{i=1}^{n} X_i$

(1) $p = E(X_1) : \frac{T}{n}$

(2) $p(1-p) = Var(X_1) : \frac{T(n-T)}{n(n-1)}$

(3) $p^r : \frac{(T)_r}{(n)_r} = \frac{T(T-1)\cdots(T-r+1)}{n(n-1)\cdots(n-r+1)},\ r = 1, 2, \ldots, n$

$X_i \overset{iid}{\sim} \text{Bin}\,(n,p) \implies$ Complete Sufficient: $T = \sum\limits_{i=1}^{n} X_i$

$\rightarrow\ p : \dfrac{\bar{X}_n}{n} = \dfrac{T}{n^2}$

Poisson

$X_i \overset{iid}{\sim} \text{Poisson}\,(\lambda) \implies$ Complete Sufficient: $T = \sum_{i=1}^{n} X_i$

(**1**) $\lambda^r : \frac{(T)_r}{n^r},\ r = 1, 2, \ldots$

(**2**) $e^{-k}\frac{\lambda^k}{k!} = P(X_1 = k) : \binom{T}{k}\frac{(n-1)^{T-k}}{n^T}$

(**3**) $e^{-k\lambda} = P(X_1 = 0, X_2 = 0, \ldots, X_k = 0) : \left(1 - \frac{k}{n}\right)^T,\ 1 \le k < n$

Geometric

$X_i \overset{iid}{\sim} \text{Geo}\,(p) \implies$ Complete Sufficient: $T = \sum_{i=1}^{n} X_i$

$\to p = P(X_1 = 0) : \frac{n-1}{n-1+T}$

Uniform

(**1**) **Discrete:** $X_i \overset{iid}{\sim} \text{U}\{1, 2, \ldots, N\} \implies$ Complete Sufficient: $T = X_{(n)}$

$\to N : \frac{T^{n+1}-(T-1)^{n+1}}{T^n-(T-1)^n}$

(**2**) **Continuous:** $X_i \overset{iid}{\sim} \text{U}(0, \theta) \implies$ Complete Sufficient: $T = X_{(n)}$

$\to \psi(\theta) : \left\{\frac{T\psi'(T)}{n} + \psi(T)\right\}$ $\quad \left[\psi(\theta) = \theta^r : \left(\frac{n+r}{n}\right) T^r\right]$

Also if, $X_i \overset{iid}{\sim} \text{U}(\theta_1, \theta_2) \implies$ Complete Sufficient: $T = \left(X_{(1)}, X_{(n)}\right)$

$\to \theta_1 : \frac{nX_{(1)}-X_{(n)}}{n-1} \qquad \theta_2 : \frac{nX_{(n)}-X_{(1)}}{n-1}$

Gamma

$X_i \overset{iid}{\sim} \text{Exp}\,(\theta) \implies$ Complete Sufficient: $T = \sum_{i=1}^{n} X_i$

(**1**) $\theta = E(X_1) : \frac{T}{n}$

(**2**) $\frac{1}{\theta} : \frac{n-1}{T}$

(**3**) $P(X_1 > k) = e^{-\frac{k}{\theta}} : \left(1 - \frac{k}{T}\right)^{n-1}$, if $k < T$

(**4**) $f(k;\theta) = \frac{1}{\theta}e^{-\frac{k}{\theta}} : \frac{(n-1)}{T}\left(1 - \frac{k}{T}\right)^{n-2}$, if $k < T$

$X_i \overset{iid}{\sim} \text{Gamma}\,(p, \theta) \implies$ Complete Sufficient: $\left(\sum_{i=1}^{n} \ln X_i, \sum_{i=1}^{n} X_i\right)$ $\qquad$ (mean $= p\,\theta$)

For known p, $\theta^r : \dfrac{\Gamma(np)}{\Gamma(np+r)}T^r,\ r > -np$

$X_i \overset{iid}{\sim} \text{Shifted Exp}(\theta, \sigma_0) \implies$ Complete Sufficient: $X_{(1)} \to \theta : X_{(1)} - \dfrac{\sigma_0}{n}$

$X_i \overset{iid}{\sim} \text{DE}(\mu_0, \sigma) \implies$ Complete Sufficient: $T = \sum_{i=1}^{n} |X_i - \mu_0| \to \sigma^r : \dfrac{\Gamma(n)}{\Gamma(n+r)}T^r,\ r > -n$

Beta

$X_i \overset{iid}{\sim} \text{Beta}\,(\theta_1, \theta_2) \implies$ Complete Sufficient: $\left(\sum\limits_{i=1}^{n} \ln X_i, \sum\limits_{i=1}^{n} \ln(1-X_i)\right)$

For $\theta_2 = 1$, UMVUE of θ_1 is $\dfrac{n-1}{-\sum\limits_{i=1}^{n} \ln X_i}$

Normal

$X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma_0^2) \implies$ Complete Sufficient: $\sum\limits_{i=1}^{n} X_i$ or $\bar{X}$ $\quad \to \mu : \bar{X} \quad \mu^2 : \bar{X}^2 - \dfrac{\sigma_0^2}{n}$

$X_i \overset{iid}{\sim} \mathcal{N}(\mu_0, \sigma^2) \implies$ Complete Sufficient: $\sum\limits_{i=1}^{n} (X_i - \mu_0)^2$ or S_0^2 $\quad \to \sigma^r : S_0^r K_{n,r},\ r > -n$

$X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2) \implies$ Complete Sufficient: $\left(\sum\limits_{i=1}^{n} X_i, \sum\limits_{i=1}^{n} X_i^2\right)$ or $\left(\bar{X}, S^2\right)$

(1) $\mu = E(X_1) : \bar{X}$

(2) $\sigma^2 : S^2, \quad \sigma^r : S^r K_{n-1,r},\ r > -(n-1)$ $\qquad \left[K_{n-1,r} = \dfrac{(n-1)^{\frac{r}{2}}\Gamma(\frac{n-1}{2})}{2^{\frac{r}{2}}\Gamma(\frac{n-1+r}{2})}\right]$

(3) $\dfrac{\mu}{\sigma^r} : \bar{X} \cdot K_{n-1,-r}\, S^{-r},\ r < (n-1)$

(4) p^{th} quantile of $X_1 = \xi_p = \mu + \sigma\,\Phi^{-1}(p) : \bar{X} + K_{n-1,1}\, S\, \Phi^{-1}(p)$

$X_i \overset{iid}{\sim} \mathcal{N}(\theta, 1)$ $\qquad \left[\phi(x; \mu, \sigma^2) : \text{PDF of } \mathcal{N}(\mu, \sigma^2)\right]$

(1) $\Phi(k - \theta) = P(X_1 \le k) : \Phi\left[(k - \bar{X})\sqrt{\dfrac{n}{n-1}}\right]$

(2) $\phi\left(k; \theta, 1\right) : \phi(k; \bar{X}, \frac{n-1}{n})$

(3) $e^{h\theta} : e^{h\bar{X} - \frac{h^2}{2n}}$

Others

$X_i \overset{iid}{\sim} \text{Pareto}(\theta, x_0) \implies$ Complete Sufficient: $\prod\limits_{i=1}^{n} X_i$ or $\sum\limits_{i=1}^{n} \ln X_i$ $\qquad (x_0$ known$)$

$\to \dfrac{1}{\theta^r} : \dfrac{\Gamma(n)}{\Gamma(n+r)} \left\{\sum\limits_{i=1}^{n} \ln\left(\frac{X_i}{x_0}\right)\right\}^r,\ r > -n$

Special case: $r = -1 \implies \theta : \dfrac{n-1}{\sum\limits_{i=1}^{n} \ln\left(\frac{X_i}{x_0}\right)}$

$X_i \overset{iid}{\sim} \text{Pareto}(\theta_0, \alpha) \implies$ Complete Sufficient: $X_{(1)}$ $\qquad (\theta_0$ known$)$

$\to \alpha^r : \left(1 - \dfrac{r}{n\theta_0}\right) X_{(1)}^r \quad \left[\text{if } r < n\theta_0\right]$

2.1.7 Cramer-Rao Inequality

Let X has a distribution from $\{f(x;\theta) : \theta \in \Omega\}$ satisfying the following regularity conditions -

(i) The parameter space, Ω is an open interval in $\mathbb{R}$ i.e. $\Omega = \{\theta : a < \theta < b\}$

(ii) The support, $\mathcal{S} = \{x : f(x;\theta) > 0\}$ is independent of θ

(iii) For each $x \in \mathcal{S}$, $\frac{\partial}{\partial\theta}\big[\ln f(x;\theta)\big]$ exists and finite

(iv) The identity "$\sum\limits_{x\in\mathcal{S}} f(x;\theta) = 1$" or "$\int\limits_{\mathcal{S}} f(x;\theta)dx = 1$" can be differentiated under the summation or integral sign.

(v) $T \in U_\psi = \Big\{T(X) : E_\theta\big[T(X)\big] = \psi(\theta), Var_\theta\big[T(X)\big] < \infty, \forall\theta \in \Omega\Big\}$ is an UE of $\psi(\theta)$ such that the derivative of $\psi(\theta) = E\big[T(X)\big]$ with respect to θ can be evaluated by differentiating under the summation or integral sign.

Then, $Var\big[T(X)\big] \geq \frac{\{\psi'(\theta)\}^2}{I(\theta)}$ where $I(\theta) = E\left[\frac{\partial}{\partial\theta}\big\{\ln f(x;\theta)\big\}\right]^2 > 0$

Equality Case

'=' holds in CR Inequality iff -

$$\frac{\partial}{\partial\theta}\big[\ln f(x;\theta)\big] = \pm\frac{I(\theta)}{\psi'(\theta)}\{T - \psi(\theta)\} \text{ a.e. } \ldots\ldots(*)$$

$\iff$ the family $\{f(x;\theta) : \theta \in \Omega\}$ belongs to OPEF

$\rightarrow$ $(*)$ is the **necessary and sufficient** condition for attaining CRLB by an UE, $T(X)$ of $\psi(\theta)$.

Remarks

(**1**) Even in OPEF, the only parametric function for which $T(X)$ attains CRLB, is that $E\big[T(X)\big]$

(**2**) If MVBUE $T(X)$ of $\psi(\theta)$ exists, then it is given by, $T(X) = \psi(\theta) \pm \frac{\psi'(\theta)}{I(\theta)} \cdot \frac{\partial}{\partial\theta}\big\{\ln f(X;\theta)\big\}$
MVBUE is also the UMVUE but UMVUE may not be MVBUE always -

- **Non-regular case:** one of the regularity conditions does not hold, eg. $\{\text{U}(0,\theta) : \theta > 0\}$
- If all the regularity conditions hold but CRLB is not attainable, then there may exist UMVUE but that is not the MVBUE

(**3**) **Fisher's Information**

(a) $I(\theta) = E\left[\frac{\partial}{\partial\theta}\big\{\ln f(X;\theta)\big\}\right]^2 = E\left[-\frac{\partial^2}{\partial\theta^2}\big\{\ln f(X;\theta)\big\}\right]$

(b) $I_{\underset{\sim}{X}}(\theta) = n \cdot I_{X_1}(\theta)$, if the regularity conditions hold

(c) $\underset{\sim}{X} \overset{iid}{\sim} \{f(x;\theta) : \theta \in \Omega\} \implies$ for any statistic $T(\underset{\sim}{X})$, $I_{T(\underset{\sim}{X})}(\theta) \leq I_{\underset{\sim}{X}}(\theta)$
'=' holds if and only if $T(\underset{\sim}{X})$ is sufficient

(**4**) **Lower bound for the MSE of any estimator:** $\text{MSE}_{\psi(\theta)}(T) \geq \dfrac{\left\{\frac{\partial}{\partial\theta}E(T)\right\}^2}{I(\theta)} = \dfrac{\{\psi'(\theta) + b'(\theta)\}^2}{I(\theta)}$

(**5**) $\{\mathcal{C}(\theta, 1) : \theta \in \mathbb{R}\}$ is a regular family as the CR inequality holds, but CRLB is not attainable

2.1.8 Methods of Estimation

Method of Moments

If the sample drawn is a good representation of the population, then this method is quite reasonable. Equate 'k' sample moments m'_r with corresponding population moments μ'_r and solve for k unknowns for a k-parameter family.

Method of Least Squares

Here we minimize the sum of squares of errors with respect to the parameter $(\theta_1, \theta_2, \ldots, \theta_k)$

Model: $y_i = E(Y \mid X = x_i) + z_i$

Assumptions: Conditional distribution of $Y \mid X = x_i$ is homoscedastic.

Method of Maximum Likelihood

(**1**) **Bernoulli (p)**

(a) $p \in (0,1) \implies$ No MLE of p when $\underset{\sim}{x} = \underset{\sim}{0}$ or $\underset{\sim}{x} = \underset{\sim}{1}$, else $\bar{X}$

(b) $\Omega = \left\{p : p \in \{\mathbb{Q}' \cap [0,1]\}\right\} \implies$ No MLE of $p \in \Omega$

(**2**) $X_i \overset{iid}{\sim} \mathrm{U}(0,\theta),\ \theta > 0 \implies \hat{\theta} = X_{(n)}$

(**3**) $X_i \overset{iid}{\sim} \mathrm{U}(\alpha,\beta),\ \alpha < \beta \implies \hat{\underset{\sim}{\theta}} = \left(\hat{\alpha}, \hat{\beta}\right) = \left(X_{(1)}, X_{(n)}\right)$

(**4**) **MLE is not unique**

$$X_i \overset{iid}{\sim} \mathrm{U}(\theta - k_1, \theta + k_2) \implies \hat{\theta} = \alpha(X_{(n)} - k_2) + (1-\alpha)(X_{(1)} + k_1),\ \alpha \in [0,1]$$

(**5**) $X_i \overset{iid}{\sim} \mathrm{U}(-\theta,\theta),\ \theta > 0 \implies \hat{\theta} = \max\limits_{i=1(1)n} \{|X_i|\} = \max\left\{-X_{(1)}, X_{(n)}\right\}$

(**6**) $X_i \overset{iid}{\sim} \mathcal{N}(\mu,\sigma^2) \implies \left(\hat{\mu}, \widehat{\sigma^2}\right) = \left(\bar{X}, \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2\right)$

(**7**) $X_i \overset{iid}{\sim} \mathrm{Gamma}\,(p_0,\theta) \implies \hat{\theta} = \frac{\bar{X}}{p_0}$ $\qquad (p_0$ known$)$

(**8**) $X_i \overset{iid}{\sim} \mathrm{Beta}\,(\theta,1) \implies \hat{\theta} = \dfrac{n}{-\sum\limits_{i=1}^{n} \ln X_i}$

(**9**) $X_i \overset{iid}{\sim} \mathrm{Pareto}\,(x_0,\theta) \implies \left(\hat{x_0}, \hat{\theta}\right) = \left(X_{(1)}, \dfrac{n}{\sum\limits_{i=1}^{n} \ln\left(\frac{X_i}{X_{(1)}}\right)}\right)$

(**10**) $X_i \overset{iid}{\sim} \mathrm{DE}\,(\mu,\sigma) \implies (\hat{\mu}, \hat{\sigma}) = \left(\tilde{X}, \frac{1}{n}\sum\limits_{i=1}^{n}|X_i - \tilde{X}|\right)$

(**11**) $X_i \overset{iid}{\sim} \text{Shifted Exp}\,(\mu,\sigma) \implies (\hat{\mu}, \hat{\sigma}) = \left(X_{(1)}, \bar{X} - X_{(1)}\right)$

In particular if $\mu = \sigma > 0$, then $\hat{\mu} = X_{(1)}$

(12) **Truncated parameter:** $X_i \overset{iid}{\sim} \text{Bernoulli}(p),\ p \in \left[\frac{1}{4}, \frac{3}{4}\right]$. Here, the MLE of p is -

$$\hat{p}(\underset{\sim}{X}) = \begin{cases} \dfrac{1}{4}, & \text{if } \bar{X} < \dfrac{1}{4} \\ \bar{X}, & \text{if } \dfrac{1}{4} \le \bar{X} \le \dfrac{3}{4} \\ \dfrac{3}{4}, & \text{if } \bar{X} > \dfrac{3}{4} \end{cases}$$

It is better than the UMVUE, $\bar{X}$ of $p \in \left[\frac{1}{4}, \frac{3}{4}\right]$, in terms of variability

Properties

(13) MLE, if exists is a function of (minimal) sufficient statistic

(14) Under the regularity conditions of CR inequality MVBUE exists, then that is the MLE

(15) **Invariance property:** $\hat{\theta}$ is the MLE of $\theta \implies h(\hat{\theta})$ is the MLE of $h(\theta)$ for any function $h(\cdot)$

(16) For large n, the bias of MLE become insignificant

(17) Under <u>normality</u>, LSE $\equiv$ MLE.

(18) **Asymptotic property**

(a) Under certain regularity conditions, the MLE $\hat{\theta}$ of θ is consistent and also

$$\hat{\theta} \overset{a}{\sim} \mathcal{N}\left(\theta, \frac{1}{nI_1(\theta)} = \frac{1}{I_n(\theta)}\right) \qquad \underline{\text{or}} \qquad \sqrt{n}\left(\hat{\theta} - \theta\right) \overset{a}{\sim} \mathcal{N}\left(0, \frac{1}{I_1(\theta)}\right)$$

(b) In OPEF, let $\hat{\theta}$ is the MLE of θ then -

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \overset{a}{\sim} \mathcal{N}\left(0, \frac{1}{I_1(\theta)}\right) \implies \sqrt{n}\left\{\psi(\hat{\theta}) - \psi(\theta)\right\} \overset{a}{\sim} \mathcal{N}\left(0, \frac{\{\psi'(\theta)\}^2}{I_1(\theta)}\right)$$

2.2 Testing of Hypothesis

2.2.1 Tests of Significance

Univariate Normal

(1) $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$

(a) $\sigma = \sigma_0$ (known) $\rightarrow \omega = \left\{\underset{\sim}{x} : \left|\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma_0}\right| > \tau_{\frac{\alpha}{2}}\right\}$ $\left(\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma_0} > \tau_\alpha \text{ if } H_1 : \mu > \mu_0\right)$

(b) σ unknown $\rightarrow \omega = \left\{\underset{\sim}{x} : \left|\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\right| > t_{\frac{\alpha}{2};n-1}\right\}$, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

(2) $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma \neq \sigma_0$

(a) $\mu = \mu_0$ (known), $Z = \frac{ns_0^2}{\sigma_0^2} \rightarrow \omega = \left\{Z_{obs} > \chi^2_{\frac{\alpha}{2};n} \text{ or } Z_{obs} < \chi^2_{1-\frac{\alpha}{2};n}\right\}$

(b) μ unknown, $Z = \frac{(n-1)s^2}{\sigma_0^2} \rightarrow \omega = \left\{Z_{obs} > \chi^2_{\frac{\alpha}{2};n-1} \text{ or } Z_{obs} < \chi^2_{1-\frac{\alpha}{2};n-1}\right\}$

Two Independent Univariate Normal

(1) $H_0 : \mu_1 - \mu_2 = \xi_0$ (known) against $H_0 : \mu_1 - \mu_2 \neq \xi_0$

(a) σ_1, σ_2 are known $Z = \frac{\bar{X}_1 - \bar{X}_2 - \xi_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \to \omega = \left\{ |Z_{obs}| > \tau_{\frac{\alpha}{2}} \right\}$

(b) $\sigma_1 = \sigma_2 = \sigma$ (unknown), $Z = \frac{\bar{X}_1 - \bar{X}_2 - \xi_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \to \omega = \left\{ |Z_{obs}| > t_{\frac{\alpha}{2}; n_1+n_2-2} \right\}$, $s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

(2) $H_0 : \frac{\sigma_1}{\sigma_2} = \xi_0$ (known) against $H_1 : \frac{\sigma_1}{\sigma_2} \neq \xi_0$

(a) μ_1, μ_2 are known, $F = \frac{s_{10}^2}{s_{20}^2} \cdot \frac{1}{\xi_0^2} \to \omega = \left\{ F_{obs} > F_{\frac{\alpha}{2}; n_1, n_2} \text{ or } \frac{1}{F_{obs}} > F_{\frac{\alpha}{2}; n_2, n_1} \right\}$

(b) μ_1, μ_2 are unknown $F = \frac{s_1^2}{s_2^2} \cdot \frac{1}{\xi_0^2} \to \omega = \left\{ F_{obs} > F_{\frac{\alpha}{2}; n_1-1, n_2-1} \text{ or } \frac{1}{F_{obs}} > F_{\frac{\alpha}{2}; n_2-1, n_1-1} \right\}$

Bivariate Normal (Correlated Case)

(1) $H_0 : \mu_1 - \mu_2 = \xi_0$ (known) $\to \omega = \left\{ (\underset{\sim}{x}, \underset{\sim}{y}) : \left| \frac{\sqrt{n}(\bar{x} - \bar{y} - \xi_0)}{s_{xy}} \right| > t_{\frac{\alpha}{2}; n-1} \right\}$, $s_{xy}^2 = s_x^2 + s_y^2 + 2r s_x s_y$

(2) $H_0 : \rho = 0 \to \omega = \left\{ \left| \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \right| > t_{\frac{\alpha}{2}; n-2} \right\}$ [r: sample correlation coefficient of $(\underset{\sim}{x}, \underset{\sim}{y})$]

(3) $H_0 : \frac{\sigma_1}{\sigma_2} = \xi_0 \to \omega = \left\{ \left| \frac{r_{uv}\sqrt{n-2}}{\sqrt{1-r_{uv}^2}} \right| > t_{\frac{\alpha}{2}; n-2} \right\}$ $\quad U = X + \xi_0 Y$, $V = X - \xi_0 Y$

Binomial Proportion

(I) Single Proportion - $H_0 : p = p_0$, observed value: x_0

(a) $H_1 : p > p_0$, p-value $= P_1 = P_{H_0}(X \geq x_0)$

(b) $H_1 : p < p_0$, p-value $= P_2 = P_{H_0}(X \leq x_0)$

(c) $H_1 : p \neq p_0$, p-value $= P_3 = 2 \cdot \min\{P_1, P_2\}$ (Reject H_0 if p-value $\leq \alpha$)

(II) Two Proportions - $H_0 : p_1 = p_2 = p$, observed value of $X_1 : x_{10}$ and $X_1 + X_2 : x_0$

(a) $H_1 : p_1 > p_2$, p-value $= P_1 = P_{H_0}(X_1 \geq x_{10} \mid X_1 + X_2 = x_0)$

(b) $H_1 : p_1 < p_2$, p-value $= P_2 = P_{H_0}(X_1 \leq x_{10} \mid X_1 + X_2 = x_0)$

(c) $H_1 : p_1 \neq p_2$, p-value $= P_3 = 2 \cdot \min\{P_1, P_2\}$ (Reject H_0 if p-value $\leq \alpha$)

Poisson Mean

(I) Single Population - $H_0 : \lambda = \lambda_0$, observed value of $S = \sum_{i=1}^{n} X_i : s_0$

(a) $H_1 : \lambda > \lambda_0$, p-value $= P_1 = P_{H_0}(S \geq s_0)$

(b) $H_1 : \lambda < \lambda_0$, p-value $= P_2 = P_{H_0}(S \leq s_0)$

(c) $H_1 : \lambda \neq \lambda_0$, p-value $= P_3 = 2 \cdot \min\{P_1, P_2\}$ (Reject H_0 if p-value $\leq \alpha$)

(II) Two Populations - $H_0 : \lambda_1 = \lambda_2 = \lambda$, observed value of $S_1 = \sum_{i=1}^{n_1} X_{1i} : s_{10}$ and $S_1 + S_2 : s_0$

(a) $H_1 : \lambda > \lambda_0$, p-value $= P_1 = P_{H_0}(S_1 \geq s_{10} \mid S_1 + S_2 = s_0)$

(b) $H_1 : \lambda < \lambda_0$, p-value $= P_2 = P_{H_0}(S_1 \leq s_{10} \mid S_1 + S_2 = s_0)$

(c) $H_1 : \lambda \neq \lambda_0$, p-value $= P_3 = 2 \cdot \min\{P_1, P_2\}$ (Reject H_0 if p-value $\leq \alpha$)

2.3 Interval Estimation

T : sufficient statistic and $\big(\theta_1(T), \theta_2(T)\big)$ is a confidence interval with confidence coefficient $(1 - \alpha)$

$$\implies P\left[\big(\theta_1(T), \theta_2(T)\big) \ni \psi(\theta)\right] = 1 - \alpha \quad \forall \theta \in \Omega$$

2.3.1 Methods of finding C.I.

Find a function $\phi(T, \theta)$, whose sampling distribution is completely specified. This is the pivot. Then find c_1, c_2 based on the restriction: $P_\theta\left[c_1 < \phi(T, \theta) < c_2\right] = 1 - \alpha$

Note

For a parameter θ, the method of guessing θ is known as estimation and an interval estimate of θ is known as confidence interval for θ.

For a R.V. Y, a method of guessing Y is known as prediction and an interval prediction of Y is known as prediction limits.

$$X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2),\ i = 1(1)n \implies \text{Prediction limits for } X_{n+1} : \left(\bar{X} \mp t_{\frac{\alpha}{2}; n-1}\sqrt{\tfrac{n+1}{n}}s\right)$$

2.3.2 Wilk's Optimum Criteria

Definition: *A* $(1-\alpha)$ *level confidence interval* $\big(\underline{\theta}^*(T), \overline{\theta}^*(T)\big)$ *of* $\theta \in \Omega$*, is said to be shortest length confidence interval, in the class of all level* $(1-\alpha)$ *confidence intervals based on a pivot* $\psi(T, \theta)$*, if*

$$E_\theta\big[\underline{\theta}^*(T) - \overline{\theta}^*(T)\big] \leq E_\theta\big[\underline{\theta}(T) - \overline{\theta}(T)\big],\ \forall \theta \in \Omega$$

whatever the other $(1-\alpha)$ *level confidence interval* $\big(\underline{\theta}(T), \overline{\theta}(T)\big)$ *based on* $\psi(T, \theta)$*.*

2.3.3 Test Inversion Method

Let $A(\theta_0)$ be the "Acceptance Region" of a size 'α' test of $H_0 : \theta = \theta_0$. Define,

$$I(\underset{\sim}{x}) = \left\{\theta \in \Omega : A(\theta) \ni x\right\},\ \underset{\sim}{x} \in \mathfrak{X}$$

then $I(\underset{\sim}{x})$ is a confidence interval for θ at confidence coefficient $(1 - \alpha)$.

List of Confidence Intervals

(**1**) $X_i \overset{iid}{\sim} \mathrm{U}(0,\theta) : Pivot = \frac{X_{(n)}}{\theta} \implies \left(X_{(n)}, \frac{X_{(n)}}{\sqrt[n]{\alpha}} \right)$

(**2**) $X_i \overset{iid}{\sim} \text{Shifted Exp}\,(\mu,\sigma_0) : Pivot = \frac{n}{\sigma_0}[X_{(1)} - \mu] \implies \left[X_{(1)} + \frac{\sigma_0}{n}\ln\alpha, X_{(1)} \right]$ (finite length)

Infinite Length: $\left[-\infty, X_{(1)} + \frac{\sigma_0}{n}\ln(1-\alpha) \right]$ (σ_0 known)

(**3**) $X_i \overset{iid}{\sim} \mathcal{N}(\mu,\sigma^2) : Pivot = \sqrt{n}\left(\frac{\bar{X}-\mu}{S} \right) \implies \left(\bar{X} \mp t_{\frac{\alpha}{2};n-1} \frac{S}{\sqrt{n}} \right)$

(**4**) $X_i \overset{iid}{\sim} \text{Exp}\,(\theta) : Pivot = \frac{2}{\theta}\sum_{i=1}^{n} X_i \implies \left(\dfrac{2T}{\chi^2_{\frac{\alpha}{2};2n}}, \dfrac{2T}{\chi^2_{1-\frac{\alpha}{2};2n}} \right)$

Based on $X_{(1)}$, $Pivot = \frac{2n}{\theta}X_{(1)} \implies \left(\dfrac{2nX_{(1)}}{\chi^2_{\frac{\alpha}{2};2}}, \dfrac{2nX_{(1)}}{\chi^2_{1-\frac{\alpha}{2};2}} \right)$

2.4 Large Sample Theory

2.4.1 Modes of Convergence

(I) Convergence in Distribution

Definition: *A sequence $\{X_n\}$ of random variables with the corresponding sequence $F_n(x)$ of D.F.s is said to converge to a random variable X with D.F. $F(x)$, if*

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \textit{at every continuity point of F(x)}$$

Results

(**1**) $X_n \overset{iid}{\sim} \mathrm{U}(0,\theta) \implies n\left(\theta - X_{(n)}\right) \xrightarrow{D} \text{Exp}\,(\theta) \xleftarrow{D} nX_{(1)}$

(**2**) $X_n \overset{iid}{\sim} \text{Shifted Exp}\,(0,\theta) \implies n\left(X_{(n)} - \theta\right) \xrightarrow{D} \text{Exp}\,(1)$

(**3**) $X_n \overset{iid}{\sim} \mathcal{N}(\mu,\sigma^2) \implies \bar{X} \xrightarrow{D} \mu$

(**4**) $X_n \sim \text{Geo}\,(p_n = \frac{\theta}{n}) \implies \frac{X_n}{n} \xrightarrow{D} \text{Exp}\,(\frac{1}{\theta})$

(**5**) $X \sim \text{NB}\,(n,p) \implies 2pX \xrightarrow{D} \chi^2_{2n}$ as $p \to 0$

(**6**) $X_n \sim \text{Gamma}\,(n,\beta) \implies \frac{X_n}{n} \xrightarrow{D} \beta$

Limiting MGF

(**7**) MGF $\to X_n : M_n(t)$, $X : M(t)$, $E(X_n)$ exists $\forall\, n$ and $X_n \xrightarrow{D} X$

If $\lim_{n\to\infty} M_n(t)$, $\lim_{n\to\infty} E(X_n)$ is finite then $M_N(t) \to M(t)$, $E(X_n) \to E(X)$ as $n \to \infty$

(**8**) **Theorem:** *Let, $\{F_n\}$ be a sequence of D.F.s with corresponding M.G.F.s $\{M_n\}$ and suppose that $M_n(t)$ exists for $|t| \le t_0$, $\forall n$. If there exists a D.F. F with corresponding M.G.F. M, which exists for $|t| \le t_1 < t_0$, such that $M_n(t) \to M(t)$ as $n \to \infty$ for every $t \in [-t_1, t_1]$, then $F_n \xrightarrow{W} F$*

(II) Convergence in Probability

Definition: *Let, $\{X_n\}$ be a sequence of R.V.s defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$. Then we say that $\{X_n\}$ converges in probability to a R.V. X, defined on $(\Omega, \mathcal{A}, \mathcal{P})$, if for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P[|X_n - X| < \epsilon] = 1 \qquad or \qquad \lim_{n\to\infty} P[|X_n - X| > \epsilon] = 0$$

Sufficient Condition: *If $\{X_n\}$ is a sequence of R.V.s such that $E(X_n) \to C$ and $Var(X_n) \to 0$ as $n \to \infty$ or $E(X_n - C)^2 \to 0$ as $n \to \infty$, then $X_n \xrightarrow{P} C$.*

Counter example:

$$P(X_n = x) = \begin{cases} 1 - \dfrac{1}{n} & , x = k \\ \dfrac{1}{n} & , x = k + n \end{cases} \implies E(X_n - k)^2 \not\to 0 \text{ as } n \to \infty \text{ but } X_n \xrightarrow{P} k$$

Results

(**1**) $X_i \overset{iid}{\sim} \mathrm{U}(0,1) \implies \left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}} \xrightarrow{P} \frac{1}{e}$

(**2**) $X_n \xrightarrow{P} X,\ \lim_{n\to\infty} a_n = a \in \mathbb{R} \implies a_n X_n \xrightarrow{P} aX$

(**3**) $X_n \xrightarrow{D} X,\ \lim_{n\to\infty} a_n = \infty,\ a_n > 0\, \forall n \implies a_n^{-1} X_n \xrightarrow{P} 0$

(**4**) **Limit Theorems:** If $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, then,

(a) $X_n \pm Y_n \xrightarrow{P} X \pm Y$

(b) $X_n Y_n \xrightarrow{P} XY$

(c) $g(X_n) \xrightarrow{P} g(X)$, if $g(\cdot)$ is continuous (Invariance Property)

(d) $\dfrac{X_n}{Y_n} \xrightarrow{P} \dfrac{X}{Y}$, provided $P(Y_n = 0) = 0 = P(Y = 0)$

Chapter 3

Mathematics

3.1 Basics

3.1.1 Combinatorial Analysis

(**1**) For a population with n elements, the number of samples of size r is -

$$\text{ordered sample} = \begin{cases} n^r & \text{, WR} \\ {}^nP_r \text{ or } (n)_r & \text{, WOR} \end{cases}$$

$$\text{unordered sample} = {}^nC_r \text{ or } \binom{n}{r}\text{, WOR}$$

(**2**) **Partition of population -** The number of ways in which a population of n elements can be divided into k ordered parts of which i^{th} part consists of r_i members, $i = 1, 2, \ldots, k$ is -

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\cdots\binom{n-r_1-r_2-\cdots-r_{k-1}}{r_k}\text{ , } r_1+r_2+\cdots+r_k = n$$

$$= \frac{n!}{r_1!r_2!\cdots r_k!} = \binom{n}{r_1\ r_2\cdots r_k}$$

(a) The number of different distributions of r identical balls into n cells i.e. the number of different solutions $(r_1, r_2, \ldots, r_n)$ of the equation:

$$r_1+r_2+\cdots+r_n = r \quad \text{where, } r_i \geq 0\text{, are integers, is } \binom{n+r-1}{r} = \binom{n+r-1}{n-1}$$

(b) The number of different distributions of r indistinguishable balls into n cells in which no cell remains empty i.e. the number of different solutions $(r_1, r_2, ..., r_n)$ of the equation:

$$r_1+r_2+\cdots+r_n = r \quad \text{where, } r_i \geq 1\text{, are integers, is } \binom{r-1}{n-1}$$

(**3**) (a) $\binom{n}{0}^2 + \binom{n}{1}^2 + \cdots + \binom{n}{n}^2 = \binom{2n}{n}$

(b) $\binom{n}{0}\binom{n}{m} + \binom{n}{1}\binom{n}{m+1} + \cdots + \binom{n}{n-m}\binom{n}{n} = \binom{2n}{n-m}$

3.1.2 Difference Equation

$\{x_n\}$ is a sequence, $x_n = f(x_{n-1}, \ldots, x_2, x_1)$ is difference equation
A.P.: $x_n = x_{n-1} + d$
G.P.: $x_n = r\, x_{n-1}$
$x_n = a_1 x_{n-1} + \cdots + a_p x_{n-p}$ is a linear difference equation of order p.

First Order Linear Difference Equation

$x_n = a x_{n-1} + b,\ n \geq 1$
$\implies x_n - c = (x_1 - c) a^{n-1},\ b = c(1-a)$

Second Order Linear Difference Equation

$x_n = a x_{n-1} + b x_{n-2},\ n \geq 2$
Characteristic Equation: $u^2 - au - b = 0$ with roots u_1, u_2
Case I: $u_1 \neq u_2 \implies x_n = A u_1^n + B u_2^n$
Case II: $u_1 = u_2 = u \implies x_n = (A + Bn) u^n$

3.2 Linear Algebra

3.2.1 Vectors & Vector Spaces

(1) Length of $\underset{\sim}{a} = \sqrt{\sum_{i=1}^{n} a_i^2}$ and $\sum_{i=1}^{n} a_i = \underset{\sim}{1}.\underset{\sim}{a}$ where, $\underset{\sim}{1} = \{1, 1, \ldots, 1\}$

(2) Distance between $\underset{\sim}{a}$ and $\underset{\sim}{b} = |\underset{\sim}{b} - \underset{\sim}{a}| = \sqrt{(\underset{\sim}{b} - \underset{\sim}{a}).(\underset{\sim}{b} - \underset{\sim}{a})} = \sqrt{\sum_{i=1}^{n}(b_i - a_i)^2}$

(3) Angle (θ) between two non-null vectors $\underset{\sim}{a}$ and $\underset{\sim}{b}$ is given by, $\cos\theta = \dfrac{\underset{\sim}{a}.\underset{\sim}{b}}{|\underset{\sim}{a}|\,|\underset{\sim}{b}|}$

(4) **Cauchy-Schwarz:** $(\underset{\sim}{a}.\underset{\sim}{b})^2 \le |\underset{\sim}{a}|^2\,|\underset{\sim}{b}|^2$ $\quad$ ('=' holds iff $\underset{\sim}{b} = \lambda\underset{\sim}{a}$ for some λ)

(5) **Triangular Inequality:** $|\underset{\sim}{a} - \underset{\sim}{b}| + |\underset{\sim}{b} - \underset{\sim}{c}| \ge |\underset{\sim}{a} - \underset{\sim}{c}|$

Vector Spaces

A vector space V over a field $\mathcal{F}$ is a non-empty collection of elements, satisfying the following axioms:

(a) $a, b \in V \implies a + b \in V$ $\quad$ [closed under vector addition]

(b) $a \in V \implies \alpha a \in V,\ \forall\, \alpha \in \mathcal{F}$ $\quad$ [closed under scalar multiplication]

Note: Every vector space must include the <u>null vector</u> ($\underset{\sim}{0}$).

Useful Results

(1) "$\sum_{i=1}^{r} \lambda_i \underset{\sim}{a}_i = \underset{\sim}{0} \implies \underset{\sim}{\lambda} = \underset{\sim}{0}$" $\iff \{\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_r\}$ are Linearly Independent.

The set is L.D. iff $\exists\, \underset{\sim}{\lambda} \ne \underset{\sim}{0}$ for which $\sum_{i=1}^{r} \lambda_i \underset{\sim}{a}_i = \underset{\sim}{0}$

(2) A set of vectors is LIN $\implies$ any subset is LIN

A set of vectors is LD $\implies$ any superset is LD

(3) **Basis Vector:** (i) Spans V $\quad$ (ii) LIN

(a) $\{1, t, t^2, \ldots, t^n\}$: polynomial of degree $\le n$

(b) $\{\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_n\}$: n dimensional vector space

(4) The representation of **every** vector in terms of basis is **unique**.

(5) **Replacement theorem:** $\{\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_r\}$ is a basis vector and $\underset{\sim}{b} = \sum_{i=1}^{r} \lambda_i \underset{\sim}{a}_i$

If $\lambda_i \ne 0$ then replace $\underset{\sim}{a}_i$ by $\underset{\sim}{b} \implies \{\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_{i-1}, \underset{\sim}{b}, \underset{\sim}{a}_{i+1}, \ldots, \underset{\sim}{a}_r\}$ is also a basis vector.

(6) (a) Any set of basis of basis vector for V_n space contains exactly n vectors

(b) Any n LIN vectors from V_n form a basis for V_n

(c) Any set of $(n+1)$ vectors from V_n is LD

(7) **Extension theorem:** A set of $m(< n)$ LIN vectors from V_n can be extended to a basis of V_n

(8) **Dimension:** Number of vectors in a basis **or** maximum number of LIN vectors in the space
Dimension of a subspace: {Total no. of vectors} - {No. of LIN restrictions}

(9) $V = \{\underset{\sim}{0}\}$ has no basis $\implies$ $\dim(V)$ is undefined [We assume $\dim(V) = 0$]

(10) Consider two subspaces S, T of a vector space V over the field $\mathcal{F}$ -

(a) $S \cap T$ is also a subspace and $\dim(S \cap T) \leq \min\{\dim(S), \dim(T)\} \leq \sqrt{\dim(S)\dim(T)}$
(b) $S + T = \{a + b : a \in S, b \in T\} \implies \dim(S + T) = \dim(S) + \dim(T) - \dim(S \cap T)$
(c) $S \subseteq T \implies \dim(S) \leq \dim(T)$ and $\dim(S) = \dim(T) \iff S = T$
In general, $\dim(S) = \dim(T) \not\Rightarrow S = T$

Orthogonal Vectors

(11) $\underset{\sim}{a}, \underset{\sim}{b} \in E^n$ are orthogonal ($\perp$) if $\underset{\sim}{a}.\underset{\sim}{b} = 0$ [$\underset{\sim}{0}$ is orthogonal to every vector]

(12) The set of vectors $\{\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_n\}$ is mutually orthogonal if $\underset{\sim}{a}_i.\underset{\sim}{a}_j = 0, \ \forall\, i \neq j$

(13) If a mutually orthogonal set includes the null vector then it becomes LD, else LIN

(14) $\underset{\sim}{E}^n \ni \underset{\sim}{a} \perp S_n \subseteq E^n \iff \underset{\sim}{a}$ is orthogonal to a basis of S_n

(15) <u>Ortho complement</u> of $S_n = \mathcal{O}(S_n)$: Collection of all vectors in E^n which are orthogonal to S_n

(16) (a) $S_n \cap \mathcal{O}(S_n) = \{\underset{\sim}{0}\}$ (b) $S_n + \mathcal{O}(S_n) = E^n$ (c) $\mathcal{O}\{\mathcal{O}(S_n)\} = S_n$
where, S_n is a subspace of E^n

(17) $S \oplus T$ if, $S + T = \{\underset{\sim}{x} + \underset{\sim}{y} : \underset{\sim}{x} \in S, \underset{\sim}{y} \in T\}$

(18) $S \oplus T \iff$ "$S \cap T = \{\underset{\sim}{0}\}$" $\iff$ "If $\underset{\sim}{x} \in S,\ \underset{\sim}{y} \in T,\ \underset{\sim}{x}, \underset{\sim}{y} \neq \underset{\sim}{0}$ then $\{\underset{\sim}{x}, \underset{\sim}{y}\}$ is LIN"
$\iff$ "$\underset{\sim}{x} + \underset{\sim}{y} = \underset{\sim}{0} \implies \underset{\sim}{x} = \underset{\sim}{y} = \underset{\sim}{0}$ for $\underset{\sim}{x} \in S,\ \underset{\sim}{y} \in T$" $\iff$ "$\dim(S + T) = \dim(S) + \dim(T)$"

(19) $S, T \in V$ is said to be complement if $S \oplus T = V$

(20) $M_{n\times n}(R)$: Vector space of all $(n \times n)$ real matrices
$S : (n \times n)$ Symmetric matrices, $T : (n \times n)$ Skew-symmetric matrices
$\implies S, T$ are subspaces of $M_{n\times n}$ and $S \oplus T = M_{n\times n},\ \dim(S) = \dfrac{n(n+1)}{2},\ \dim(T) = \dfrac{n(n-1)}{2}$

3.2.2 Matrices

(1) Consider a matrix $A^{m\times n}$

(a) **Row Space:** $\mathcal{R}(A) = \{\underset{\sim}{x}'A : \underset{\sim}{x} \in \mathbb{R}^m\}$
(b) **Column Space:** $\mathcal{C}(A) = \{A\underset{\sim}{x} : \underset{\sim}{x} \in \mathbb{R}^n\}$
(c) **Null Space:** $\mathcal{N}(A) = \{\underset{\sim}{x} \in \mathbb{R}^n : A\underset{\sim}{x} = \underset{\sim}{0}\}$
(d) **Left Null Space:** $\mathcal{N}'(A) = \{\underset{\sim}{x} \in \mathbb{R}^m : \underset{\sim}{x}'A = \underset{\sim}{0}'\}$

(2) $\mathcal{N}(A) = \mathcal{O}\{\mathcal{R}(A)\} \implies \dim\mathcal{N}(A) = n - \dim\mathcal{R}(A) =$ nullity of A

(3) $A^{m\times n}, B^{n\times p} \ni AB = O \implies \mathcal{C}(B) \subseteq \mathcal{N}(A) \implies \mathrm{r}(A) + \mathrm{r}(B) \leq n$

(4) $A^{n\times n} \ni A^2 = A \implies \mathcal{C}(I_n - A) = \mathcal{N}(A)$

Rank

(5) $\mathrm{r}\left(A^{m\times n}\right) \leq \min\{m, n\}$

(6) $\mathrm{r}(AB) \leq \min\{\mathrm{r}(A), \mathrm{r}(B)\}$ if AB is defined

(7) $\mathrm{r}(AB) = \mathrm{r}(A)$, if $\det B \neq 0$

(8) $A^2 = A \iff \mathrm{r}(A) + \mathrm{r}(I_n - A) = n$

(9) $\mathrm{r}(A + B) \leq \mathrm{r}(A) + \mathrm{r}(B)$

(10) $\mathrm{r}(A) = \mathrm{r}(A') = \mathrm{r}(A'A) = \mathrm{r}(AA')$

(11) $\mathrm{r}(A) = r \implies A = \sum\limits_{k=1}^{r} M_k$ with $\mathrm{r}(M_k) = 1,\ k = 1, 2, \ldots, r$

(12) $\mathrm{r}(AB - I) \leq \mathrm{r}(A - I) + \mathrm{r}(B - I)$

(13) $A^{m\times n}, B^{s\times n} \ni AB' = O \implies \mathrm{r}(A'A + B'B) = \mathrm{r}(A) + \mathrm{r}(B)$

(14) $A^2 = A, B^2 = B, \det(I - A - B) \neq 0 \implies \mathrm{r}(A) = \mathrm{r}(B)$

(15) $\mathrm{r}\begin{pmatrix} A & B \\ O & C \end{pmatrix} \geq \mathrm{r}(A) + \mathrm{r}(C)$

(16) **Sylvester Inequality:** $\mathrm{r}(AB) \geq \mathrm{r}(A) + \mathrm{r}(B) - n$
$\implies \mathrm{r}(A) + \mathrm{r}(B) - n \leq \mathrm{r}(AB) \leq \min\{\mathrm{r}(A), \mathrm{r}(B)\}$

(17) $\mathrm{r}(A + B) \leq \mathrm{r}(A) + \mathrm{r}(B) \leq \mathrm{r}(AB) + n$, provided $AB, (A + B)$ is defined

(18) $\mathrm{r}(AB) = \mathrm{r}(B) - \dim\{\mathcal{N}(A) \cap \mathcal{C}(B)\}$

(19) $\mathrm{r}(\underset{\sim}{x}'\underset{\sim}{x}) = \mathrm{r}(\underset{\sim}{x}'\underset{\sim}{y}) = \mathrm{r}(\underset{\sim}{x}\underset{\sim}{y}') = 1$, where $\underset{\sim}{x}, \underset{\sim}{y} \neq \underset{\sim}{0} \in \mathbb{R}^n$

Other Results

(20) Sum of all entries in a matrix $A : \underset{\sim}{1}'A\underset{\sim}{1}$

(21) $A^2 = A \implies (I_n - A)$ is also idempotent

(22) $C^{m\times r} = A^{m\times n}B^{n\times r} \implies C = (c_{ij}) = \sum\limits_{k=1}^{n} a_{ik}b_{kj}$

(23) $tr(A + B) = tr(A) + tr(B)$
$tr(AB) = tr(BA)$

3.2.3 Determinants

(1) $\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} = (a-b)(b-c)(c-a)$

(2) $\begin{vmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \cdots & a \end{vmatrix}_{n\times n} = \left(a + \overline{n-1}\, b\right)(a-b)^{n-1}$

(3) $\begin{vmatrix} a+b & b+c & c+a \\ b+c & c+a & a+b \\ c+a & a+b & b+c \end{vmatrix} = \begin{vmatrix} a & b & c \\ b & c & a \\ c & a & b \end{vmatrix} \times \begin{vmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{vmatrix} = 2 \begin{vmatrix} a & b & c \\ b & c & a \\ c & a & b \end{vmatrix}$

(4) **Tridiagonal matrix**

$$A_n = \begin{vmatrix} a & b & 0 & 0 & \cdots & 0 & 0 \\ c & a & b & 0 & \cdots & 0 & 0 \\ 0 & c & a & b & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & c & a \end{vmatrix}_{n\times n} \implies A_n = aA_{n-1} - bcA_{n-2} = 1 + bc + (bc)^2 + \cdots + (bc)^n$$

(5) $\begin{vmatrix} x_1^2 + y_1^2 & x_1x_2 + y_1y_2 & \cdots & x_1x_n + y_1y_n \\ x_2x_1 + y_2y_1 & x_2^2 + y_2^2 & \cdots & x_2x_n + y_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 + y_ny_1 & x_nx_2 + y_ny_2 & \cdots & x_n^2 + y_n^2 \end{vmatrix} = |A|\,|A'| = 0,\ A = \begin{pmatrix} x_1 & y_1 & 0 & \cdots & 0 \\ x_2 & y_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 0 & \cdots & 0 \end{pmatrix}$

(6) $A^{n\times n}$ and $B^{n\times n}$ differ only by a single row (or column) $\implies |A+B| = 2^{n-1}\Big(|A| + |B|\Big)$

(7) $A = (a_{ij}), B = (b_{ij}) = r^{i-j}a_{ij} \implies |B| = |A|$

(8) $\begin{vmatrix} I & O \\ O & A \end{vmatrix} = |A|,\ \begin{vmatrix} I & B \\ O & A \end{vmatrix} = |A| \implies \begin{vmatrix} A & B \\ O & C \end{vmatrix} = \begin{vmatrix} I & O \\ O & C \end{vmatrix} \times \begin{vmatrix} A & B \\ O & I \end{vmatrix} = |C||A| = |A||C|$

(9) $A^{n\times n}, A(\text{adj}\, A) = (\text{adj}\, A)A = |A|\, I_n$

$$\sum_{i=1}^{n} a_{ri}A_{si} = \begin{cases} |A|, & r = s \\ 0, & r \neq s \end{cases} \qquad r, s = 1, 2, \ldots, n$$

(10) (a) $|\text{adj}\, A| = |A|^{n-1}$

(b) $(\text{adj}\, A)^{-1} = \text{adj}\,(A^{-1}) = \frac{A}{|A|}$

(c) $\text{adj}(\text{adj}\, A) = |A|^{n-2}A$

(d) $|\text{adj}(\text{adj}\, A)| = |A|^{(n-1)^2}$

(11) $|kA| = k^n|A|,\ \text{adj}(kA) = k^{n-1}\text{adj}\, A$

(12) $\text{adj}(AB) = \text{adj}(B)\,\text{adj}(A) \qquad [\text{if } |A|, |B| \neq 0]$

(13) Adjoint of a symmetric matrix is symmetric
Adjoint of a skew-symmetric matrix is symmetric for even order, skew-symmetric if odd order
Adjoint of a diagonal matrix is diagonal

(14) $\mathrm{r}(\mathrm{adj}\, A) = \begin{cases} 0, & \text{if } \mathrm{r}(A) \le n-2 \\ 1, & \text{if } \mathrm{r}(A) = n-1 \\ n, & \text{if } \mathrm{r}(A) = n \end{cases}$

Inverse of a Matrix

(15) (a) $(AB)^{-1} = B^{-1}A^{-1}$
(b) $(A^{-1})' = (A')^{-1}$
(c) $|A+B| \neq 0 \implies |A^{-1}+B^{-1}| \neq 0 \implies (B^{-1}+A^{-1})^{-1} = A(A+B)^{-1}B$

(16) $A^{n\times n} = \begin{pmatrix} A_{11}^{k\times k} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \; |A| \neq 0$

$$|A| = \begin{cases} |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}|, & \text{if } |A_{11}| \neq 0 \\ |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}|, & \text{if } |A_{22}| \neq 0 \end{cases}$$

(17) $M = \begin{pmatrix} A & -\underset{\sim}{u} \\ \underset{\sim}{v} & 1 \end{pmatrix}, |A| \neq 0 \implies |M| = |A|(1+\underset{\sim}{v}'A^{-1}\underset{\sim}{u}) = |A+\underset{\sim}{u}\underset{\sim}{v}'|$

(18) $|A| \neq 0,\ |A+\underset{\sim}{u}\underset{\sim}{v}'| \neq 0 \iff (1+\underset{\sim}{v}'A^{-1}\underset{\sim}{u}) \neq 0$

$$(A+\underset{\sim}{u}\underset{\sim}{v}')^{-1} = A^{-1} - \frac{A^{-1}\underset{\sim}{u}\underset{\sim}{v}'A^{-1}}{1+\underset{\sim}{v}'A^{-1}\underset{\sim}{u}}$$

(19) $A_{a,b}^{n\times n} = (a-b)I_n + b\,\underset{\sim}{1}\underset{\sim}{1}' \implies A_{a,b}^{-1} = A_{c,d}$ iff $\Delta = (a-b)(a+\overline{n-1}\,b) \neq 0$
where, $c = \frac{a+(n-2)b}{\Delta}$, $d = -\frac{b}{\Delta}$

(20) $A^{n\times n} = (a_{ij}), |A| \neq 0,\ \sum_{j=1}^{n} a_{ij} = k,\ \forall\, i \implies \sum_{j=1}^{n} b_{ij} = \frac{1}{k},\ \forall\, i$ where $A^{-1} = (b_{ij})$

Orthogonal Matrix

(21) $A^{n\times n} = \begin{pmatrix} \underset{\sim}{a}_1' \\ \underset{\sim}{a}_2' \\ \vdots \\ \underset{\sim}{a}_n' \end{pmatrix}$ where, $\underset{\sim}{a}_i' = \begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{pmatrix},\ i = 1, 2, \ldots, n$

$$AA' = I_n \implies \underset{\sim}{a}_i'\underset{\sim}{a}_j = \begin{cases} 1, & \text{when } i = j \\ 0, & \text{when } i \neq j \end{cases} \qquad |A| = \pm 1$$

(22) $AA' = A'A = I_n,\ (I_n + A) \neq 0 \implies (I_n + A)^{-1}(I_n - A)$ is skew-symmetric

(23) A, B real orthogonal $\ni |A| + |B| = 0 \implies |A+B| = 0$

(24) $AA' = kI_n \implies A'A = kI_n$

Rank & Determinant

(25) **Theorem:** *For a matrix $A^{m\times n}$ the rank of A is the order of the "highest order non-vanishing minor" of A.*

(26) $A^{n\times n},\ \mathrm{r}(A) = n \iff |A| \neq 0$

(27) **Elementary Matrices:** An elementary matrix is a matrix which differs from the Identity matrix by single row (or column) operation.

(28) Elementary Row Operation $\equiv$ Pre-multiplying by corresponding elementary row matrix
Elementary Column Operation $\equiv$ Post-multiplying by corresponding elementary column matrix

(29) $\mathrm{r}(A^{m\times n}) = k, \exists\ P^{m\times m}, Q^{n\times n}, |P|, |Q| \neq 0 \ni PAQ = \begin{pmatrix} I_k & O \\ O & O \end{pmatrix}$ (Normal Form)

(30) Any non-singular matrix can be written as a product of elementary matrices

(31) **Rank Factorization:** Let, $\mathrm{r}(A^{m\times n}) = k$, then a pair $(P^{m\times k}, Q^{k\times n})$ of matrices is said to be a rank factorization of A if, $A^{m\times n} = P^{m\times k}\, Q^{k\times n}$ (non-unique way)

(32) $A^{m\times n} = P^{m\times k}\, Q^{k\times n}, \mathrm{r}(A) \leq k$, the following statements are equivalent -

(a) $\mathrm{r}(A) = k$ i.e. (P, Q) is a rank factorization of A

(b) $\mathrm{r}(P^{m\times k}) = \mathrm{r}(Q^{k\times n}) = k$

(c) The columns of P forms a basis of $\mathcal{C}(A)$ and the rows of Q forms a basis of $\mathcal{R}(A)$

(33) $A^2 = A,\ A^{m\times m} = P^{m\times k}\, Q^{k\times m}$ where $k = \mathrm{r}(A) \implies (i)\ QP = I_n \qquad (ii)\ \mathrm{r}(A) = tr(A)$

(34) $\mathrm{r}(A \mid B) = \mathrm{r}(A) \iff B = AC$, for some C

3.2.4 System of Linear Equation

System of Linear Equations: $A^{m\times n}\underset{\sim}{x}^{n\times 1} = \underset{\sim}{b}^{m\times 1}$

(1) **Homogeneous System: $\mathbf{A}\underset{\sim}{\mathbf{x}} = \underset{\sim}{\mathbf{0}}$**

(a) $\underset{\sim}{x} = \underset{\sim}{0}$ is always consistent as $\underset{\sim}{x} = \underset{\sim}{0}$ is a trivial solution

(b) $A^{m\times n}\underset{\sim}{x} = \underset{\sim}{0}$ has a non-trivial solution iff $\mathrm{r}(A) < n$

(c) The no. of LIN solutions of $A\underset{\sim}{x} = \underset{\sim}{0} = \dim \mathcal{N}(A) = n - \mathrm{r}(A)$

(d) Elementary row operation on a matrix A doesn't alter the $\mathcal{N}(A)$

(2) **General System: $\mathbf{A}\underset{\sim}{\mathbf{x}} = \underset{\sim}{\mathbf{b}},\ \underset{\sim}{\mathbf{b}} \neq \underset{\sim}{\mathbf{0}}$**

(a) $\mathrm{r}(A \mid \underset{\sim}{b})$ is either $\mathrm{r}(A)$ or $\mathrm{r}(A) + 1$
$\mathcal{C}(A \mid \underset{\sim}{b}) \supseteq \mathcal{C}(A)$

(b) $A\underset{\sim}{x} = \underset{\sim}{b} \to \begin{cases} \text{Consistent} & \iff \mathrm{r}(A \mid \underset{\sim}{b}) = \mathrm{r}(A) \\ \text{Inconsistent} & \iff \mathrm{r}(A \mid \underset{\sim}{b}) > \mathrm{r}(A) \end{cases}$

(c) $A^{m\times n}\underset{\sim}{x} = \underset{\sim}{b}$ is consistent $\to \begin{cases} \text{Unique solution} & \iff \mathrm{r}(A) = n \\ \text{At least two solutions} & \iff \mathrm{r}(A) < n \end{cases}$

(d) $A\underset{\sim}{x}_1 = A\underset{\sim}{x}_2 = \underset{\sim}{b} \implies \alpha\underset{\sim}{x}_1 + (1-\alpha)\underset{\sim}{x}_2$ is also a solution
$\implies$ If a system has two distinct solutions then it has infinitely many solutions

(3) Theorem: *$A\underset{\sim}{x} = \underset{\sim}{b}$ be a consistent system with $\underset{\sim}{x}_0$ as a particular solution. Then the set of all possible solutions of $A\underset{\sim}{x} = \underset{\sim}{b}$ is given by $\underset{\sim}{x}_0 + \mathcal{N}(A) = \{\underset{\sim}{x}_0 + \underset{\sim}{u} : \underset{\sim}{u} \in \mathcal{N}(A)\}$*

• Point, lines, planes not necessarily passing through the origin are called 'flats'. If W is non-empty flat and $\underset{\sim}{x}_0$ is a fixed vector, then the translation of W by $\underset{\sim}{x}_0$ is, $\underset{\sim}{x}_0 + W = \{\underset{\sim}{x}_0 + \underset{\sim}{w} : \underset{\sim}{w} \in W\}$ and is a flat parallel to W.