Random experiment

A random experiment is an experiment such that

(i) all possible outcome of the exp are known in advance

(ii) outcome of a particular trial is not known or cannot be predicted in advance

(iii) the exp can be repeated under identical conditions

Sample space: Set of all possible outcomes of a random experiment (~~usually~~ usually denoted by Ω)

Event: Suppose Ω is the sample space of a random experiment. If the outcome of the random experiment is a member of a set E, we say that the event E has happened; $E \subset \Omega$, thus is a collection of possible outcomes.

Examples

(1) Random experiment : tossing a coin until a head is observed

Sample space : $\Omega = \{H, TH, TTH, \ldots\ldots\}$

Event : number of tails reqd is odd

$E = \{TH, TTTH, \ldots\ldots\} \subset \Omega$

(2) 3 White, 4 red balls are numbered 1, 2, 3, 4, 5, 6, 7

Random exp : drawing a ball

Sample space : $\Omega = \{1, 2, 3, 4, 5, 6, 7\}$

Event : drawing a white ball

$$E = \{1, 2, 3\} \subset \Omega$$

Mutually exclusive event : 2 event A and B are mutually exclusive if occurrence of one signifies non-occurrence of the others, i.e. the 2 events can not occur simultaneously.

(i.e. $A_i \cap A_j = \phi \ \forall i \neq j$ for n events)

Exhaustive events : $A_1, \ldots A_n$ events are said to be exhaustive if one of them must necessarily occur.

i.e. $\bigcup_{i=1}^{n} A_i = \Omega$

Classical definition of probability

Setup : random experiment ~~results~~ has finite number of equally likely possible outcomes

$\Omega = \{w_1, w_2, \ldots, w_n\}$ say.

An outcome $w \in \Omega$ is said to be favorable to event E if $w \in E$.

$$P(E) = \frac{\text{no. of outcomes favorable to } E}{\text{total no. of outcomes}} = \frac{\text{no. of elements in } E}{\text{no. of elements in } \Omega}$$

Note: under the above def^n

(i) $\forall$ event $E \subset \Omega$; $P(E) \geq 0$

(ii) $P(\Omega) = 1$

(iii) If $E_1, \ldots E_n$ are mutually exclusive ($E_i \cap E_j = \phi$ $\forall i \neq j$)

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \frac{\text{no. of elements in } \bigcup E_i}{\text{no of elements in } \Omega}$$

$$= \frac{\sum_{i=1}^{n} (\text{no of elements in } E_i)}{\text{no. of elements in } \Omega} = \sum_{i=1}^{n} P(E_i)$$

Note: The def^n depends on 2 crucial restrictive assumptions; (i) finite number of possible outcomes & (ii) equally likely outcomes.

Relative frequency def^n of probability

Ω : sample space

E : $E \subset \Omega$ is an event

Suppose the random experiment is repeated N times

$f_N(E)$: number of times event E occurs out of N

Relative frequency of event $E = \frac{f_N(E)}{N}$

Relative freq def^n of prob

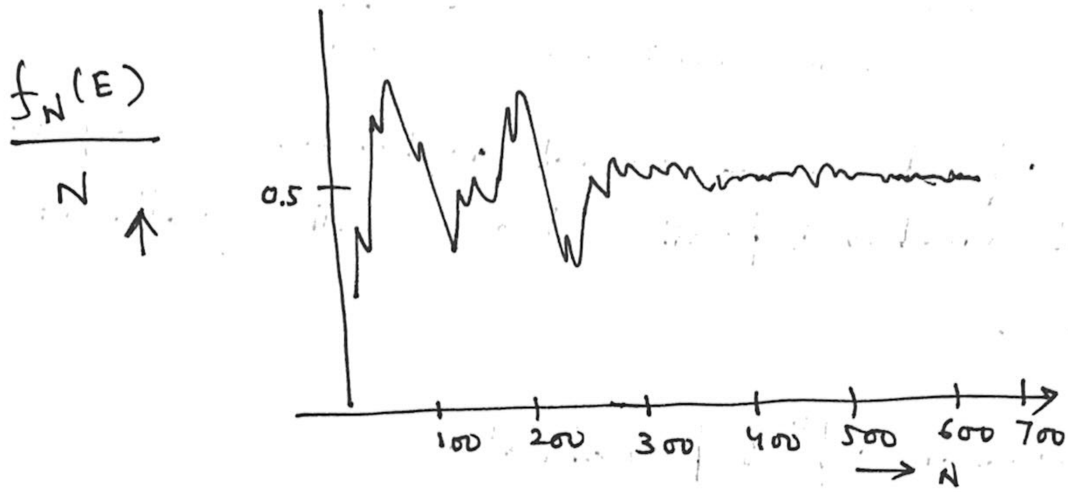
$$p_E = \lim_{N \to \infty} \frac{f_N(E)}{N}$$

prob of E

Note: p_E depends on "large" number of replications of the random exp which may not be feasible.

Note: Behavior of $\frac{f_N(E)}{N}$ can be quite erratic for small N and would eventually stabilize for "reasonably" large N

Example: To determine prob of head of a coin it is tossed N times; coin is fair



Note: Relative freq def^n of prob also satisfies

(i) $\forall E \subset \Omega \quad P(E) \geq 0$

(ii) $P(\Omega) = 1$

(iii) If $E_1, \ldots E_n$ are mutually exclusive

then $P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$

Note! relative freq def^n is based on the approximation

$\frac{f_N(E)}{N} \approx p_E$ for large N. Hence is likely to be assigned a different value by different experiment

σ- field

Let Ω be a sample space

Defn : A σ-field of subsets of Ω, $\mathcal{F}$, is a class of subsets of Ω having the following properties

(i) $\Omega \in \mathcal{F}$

(ii) $\forall A \in \mathcal{F}$, $A^c \in \mathcal{F}$ (i.e closed under complementation)

(iii) If $A_1, A_2, \ldots \in \mathcal{F}$ be a countable collection of sets, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

(i.e. closed under countable union).

Note: The above defn of $\mathcal{F}$ implies

(i) $\phi \in \mathcal{F}$ $\quad (\Omega \in \mathcal{F} \Rightarrow \Omega^c = \phi \in \mathcal{F})$

(ii) $A_1, A_2, \ldots \in \mathcal{F}$

$\Rightarrow A_1^c, A_2^c, \ldots \in \mathcal{F}$

$\Rightarrow \bigcup_1^{\infty} A_i^c \in \mathcal{F}$

$\Rightarrow \left(\bigcup_1^{\infty} A_i^c\right)^c \in \mathcal{F}$

i.e. $\bigcap_1^{\infty} A_i \in \mathcal{F}$

(iii) $A, B \in \mathcal{F}$

$\Rightarrow A - B = A \cap B^c \in \mathcal{F}$

Sly $B - A \in \mathcal{F}$

$\Rightarrow A \Delta B \stackrel{\text{def}}{=} (A-B) \cup (B-A) \in \mathcal{F}$

(iv) $A_1, \ldots A_n \in \mathcal{F} \Rightarrow \bigcup_1^n A_i \in \mathcal{F}$ & $\bigcap_1^n A_i \in \mathcal{F}$

Take $A_{n+1} = A_{n+2} = \cdots = \phi$

$\Rightarrow \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i$ or $A_{n+1} = \cdots = \Omega$ for $\bigcap A_i \in \mathcal{F}$

Remark: Although the power set of Ω, say $P(\Omega)$, is a σ-field of subsets of Ω, in general a σ-field of Ω may not contain all subsets of Ω.

Examples

(i) $\mathcal{F} = \{\phi, \Omega\}$ – a trivial sigma field

(ii) $\forall A \in \Omega$, $\mathcal{F} = \{A, A^c, \phi, \Omega\}$ – σ-field of subsets of Ω

smallest σ-field containing A

(iii) power set of Ω

$\Omega = \{H, T\}$

$\mathcal{F} = \{\{H\}, \{T\}, \{H, T\}, \phi\}$ $\quad 2^2$

$\Omega = \{HH, HT, TH, TT\}$

$\mathcal{F} = 2^4$ elements

(iv) Let $\mathcal{C}$ be a class of subsets of Ω and suppose

$\Phi = \{\phi_\alpha : \alpha \in \Delta\}$ be the collection of all σ-fields that contain $\mathcal{C}$. Then

$$\mathcal{F} = \bigcap_{\alpha \in \Delta} \phi_\alpha$$ is also a σ-field and it is the σ-field generated by, $\sigma(\mathcal{C})$ ($\sigma(\mathcal{C})$ is the smallest ~~sigma~~ σ-field that contains $\mathcal{C}$)

If we take $\Omega = \mathbb{R}$ and $\mathcal{C}$ to be the class of all open intervals in $\mathbb{R}$, then $\sigma(\mathcal{C}) = \mathcal{B}$, say, is called the Borel σ-field on $\mathbb{R}$.

Axiomatic definition of probability

Ω : sample space

$\mathcal{F}$: σ-field of subsets of Ω

Def^n : A probability f^n (or a probability measure) is a real valued set f^n defined on $\mathcal{F}$ which satisfies

(i) $\forall A \in \mathcal{F}$, $P(A) \geq 0$ — non-negativity

(ii) $P(\Omega) = 1$ — normed measure

(iii) If $A_1, A_2, \ldots \in \mathcal{F}$ such that $A_i \cap A_j = \phi \quad \forall i \neq j$

then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ — σ-additivity

$(\Omega, \mathcal{F}, P)$: probability space

Some important properties of probability f^n

(i) $P(\phi) = 0$

Let $A_1 = \Omega$ and $A_i = \phi$ for $i = 2, 3, \ldots$

Then $P(A_1) = 1$

Note that $A_1 = \bigcup_{i=1}^{\infty} A_i$ and $A_i \cap A_j = \phi \ \forall i \neq j$

$$\Rightarrow 1 = P(A_1) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

$$\text{i.e. } 1 = 1 + \sum_{i=2}^{\infty} P(\phi)$$

$$\text{i.e. } 0 = \sum_{i=2}^{\infty} P(\phi)$$

$$\Rightarrow P(\phi) = 0$$

(ii) $A_1, \ldots, A_n \in \mathcal{F}$ and $A_i \cap A_j = \phi \ \forall i \neq j$

$$\Rightarrow P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Take $A_i = \phi$ for $i = n+1, n+2, \ldots$ then

$A_1, \ldots \ldots \ni A_i \cap A_j = \phi \ \forall i \neq j$

$\& \ P(A_i) = 0 \ \forall i \geq n+1$

$$\Rightarrow P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right)$$

$$= \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i)$$

(iii) $\forall A \in \mathcal{F}$, $P(A^c) = 1 - P(A)$

Note that $\Omega = A \cup A^c$; $A \cap A^c = \phi$

$$\Rightarrow 1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

$$\Rightarrow P(A^c) = 1 - P(A)$$

(iv) $\forall A_1, A_2 \in \mathcal{F} \ni A_1 \subseteq A_2$

$A_2 = A_1 \cup (A_1^c A_2)$; A_1 & $A_1^c A_2$ are disjoint

$\Rightarrow P(A_2) = P(A_1) + P(A_1^c A_2)$

Now $A_1^c A_2 \in \mathcal{F} \Rightarrow P(A_1^c A_2) \geq 0$

$\Rightarrow P(A_2) \geq P(A_1)$ — monotonicity property

(v) $\forall A \in \mathcal{F}$, $0 \leq P(A) \leq 1$

Note that $\phi \subseteq A \subseteq \Omega$

$\Rightarrow P(\phi) \leq P(A) \leq P(\Omega)$

i.e. $0 \leq P(A) \leq 1$

(vi) $\forall A_1, A_2 \in \mathcal{F}$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 A_2)$$

Note that $A_1 \cup A_2 = A_1 \cup (A_1^c A_2)$

$$P(A_1 \cup A_2) = P(A_1) + P(A_1^c A_2)$$

Also $A_2 = \underbrace{A_1 A_2 \cup A_1^c A_2}_{\text{disjoint}}$

$\Rightarrow P(A_2) = P(A_1 A_2) + P(A_1^c A_2)$

$\Rightarrow P(A_1^c A_2) = P(A_2) - P(A_1 A_2)$

$\Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 A_2)$

<u>Note</u>: $A_1, A_2, A_3 \in \mathcal{F}$

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 A_2) - P(A_1 A_3) - P(A_2 A_3) + P(A_1 A_2 A_3)$$

Inclusion - Exclusion formula

$(\Omega, \mathcal{F}, P)$ be a probability space and let $A_1, A_2, \dots A_n \in \mathcal{F}$

$n \in \mathbb{N}$, $n \geq 2$

Let
$$p_{1,n} = \sum_{i=1}^{n} P(A_i)$$

$$p_{2,n} = \sum_{1 \leq i < j < n} P(A_i A_j)$$

$$p_{3,n} = \sum_{i<j<k} P(A_i A_j A_k)$$

$$p_{l,n} = \sum_{j_1 < j_2 < \dots < j_l} P(A_{j_1} \dots A_{j_l})$$

$$\vdots$$

$$p_{n,n} = P(A_1 \dots A_n)$$

Then $P\left(\bigcup_{i=1}^{n} A_i\right) = p_{1,n} - p_{2,n} + p_{3,n} - \dots + (-1)^{n-1} p_{n,n}$

Pf: Note that for $n=2$

$$p_1 = P(A_1) + P(A_2)$$

$$p_2 = P(A_1 A_2)$$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 A_2)$$
$$= p_{1,2} - p_{2,2}$$

results holds true for $n=2$ and also for $n=3$

Suppose it holds for $n = 2, 3, \dots m$, i.e.

$$P\left(\bigcup_{i=1}^{m} A_i\right) = p_{1,m} - p_{2,m} + p_{3,m} - \dots + (-1)^{m-1} p_{m,m}$$

Then $P\left(\bigcup_{i=1}^{m+1} A_i\right) = P\left(\left(\bigcup_{i=1}^{m} A_i\right) \cup A_{m+1}\right)$

$$= P\left(\bigcup_{i=1}^{m} A_i\right) + P(A_{m+1}) - P\left(\left(\bigcup_{i=1}^{m} A_i\right) \cap A_{m+1}\right)$$

(using $n=2$ result)

$$= \sum_{j=1}^{m} (-1)^{j-1} p_{j,m} + P(A_{m+1}) - P\left(\bigcup_{i=1}^{m} A_i A_{m+1}\right)$$

Let $B_i = A_i A_{m+1}$

then $P\left(\bigcup_{i=1}^{m} B_i\right) = p_{1,m}^{(B)} - p_{2,m}^{(B)} + \cdots + (-1)^{m-1} p_{m,m}^{(B)}$

$$= \sum_{j=1}^{m} (-1)^{j-1} p_{j,m}^{(B)} .$$

where $p_{1,m}^{(B)} = \sum_{i=1}^{m} P(A_i A_{m+1})$

$$p_{2,m}^{(B)} = \sum_{i<j} P(A_i A_j A_{m+1})$$

$$p_{m,m}^{(B)} = P(A_1 A_2 \cdots A_m A_{m+1}) = p_{m+1,m+1}$$

Thus

$$P\left(\bigcup_{i=1}^{m+1} A_i\right) = \sum_{j=1}^{m} (-1)^{j-1} p_{j,m} + P(A_{m+1}) - \sum_{j=1}^{m} (-1)^{j-1} p_{j,m}^{(B)}$$

$$= \left(p_{1,m} + P(A_{m+1})\right)$$

$$- \left(p_{2,m} + p_{1,m}^{(B)}\right)$$

$$+ \left(p_{3,m} + p_{2,m}^{(B)}\right)$$

$$- \cdots$$

$$+ (-1)^{m} p_{m,m}^{(B)}$$

Note that

$$p_{1,m} + P(A_{m+1})$$

$$= \sum_{i=1}^{m} P(A_i) + P(A_{m+1}) = \sum_{i=1}^{m+1} P(A_i) = p_{1,m+1}.$$

$$p_{2,m} + p_{1,m}^{(B)} = \sum_{1 \le i < j \le m} P(A_i A_j) + \sum_{i=1}^{m} P(A_i A_{m+1})$$

$$= \sum_{1 \le i < j \le m+1} P(A_i A_j) = p_{2,m+1}$$

and so on hence

$$P\left(\bigcup_{i=1}^{m+1} A_i\right) = p_{1,m+1} - p_{2,m+1} + \cdots + (-1)^m p_{m+1,m+1}$$

Result follows by induction

<u>Boole's inequality</u>

$$P\left(\bigcup_1^n A_i\right) \le \sum P(A_i)$$

For $n = 2$ $\quad P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 A_2)$

$$\le P(A_1) + P(A_2)$$

result holds for $n = 2$

Suppose the result holds for $n = m$, then

$$P\left(\bigcup_{i=1}^{m+1} A_i\right) = P\left(\left(\bigcup_1^m A_i\right) \cup A_{m+1}\right)$$

$$\le P\left(\bigcup_1^m A_i\right) + P(A_{m+1})$$

$$\le \sum_{i=1}^{m} P(A_i) + P(A_{m+1})$$

$$= \sum_{i=1}^{m+1} P(A_i)$$

Bonferroni's inequality

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

Realize that

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

$$= 1 - P\left(\bigcup_{i=1}^{n} A_i^c\right)$$

$$\geq 1 - \sum_{i=1}^{n} P(A_i^c)$$

$$= 1 - \sum_{i=1}^{n} \left(1 - P(A_i)\right)$$

$$= 1 - n + \sum_{i=1}^{n} P(A_i)$$

$$= \sum_{i=1}^{n} P(A_i) - (n-1)$$

Also $P\left(\bigcap_{i=1}^{n} A_i\right) \geq 0$

$\Rightarrow P\left(\bigcap_{1}^{n} A_i\right) \geq \max\left(0, \sum_{1}^{n} P(A_i) - (n-1)\right)$

Inequalities from Inclusion-exclusion formula

Boole's inequality

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i) = p_{1,n}$$

By induction, one can show that

$$p_{1,n} - p_{2,n} \leq P\left(\bigcup_{i=1}^{n} A_i\right) \leq p_{1,n}$$

$$p_{1,n} - p_{2,n} + p_{3,n} - p_{4,n} \leq P\left(\bigcup_{i=1}^{n} A_i\right) \leq p_{1,n} - p_{2,n} + p_{3,n}$$

Conditional Probability

Conditional prob of A given B : $P(A|B) = \dfrac{P(AB)}{P(B)}$; $A, B \in \mathcal{F}$

Intuitive interpretation thro' rel freq : $\dfrac{f_N(AB)}{f_N(B)}$

Def^n : Let $(\Omega, \mathcal{F}, P)$ be a probability space and $B \in \mathcal{F}$ be such that $P(B) > 0$. For any arbitrary $A \in \mathcal{F}$

$$Q(A) = P(A|B) = \frac{P(AB)}{P(B)}$$ is the conditional prob of A given B.

Result : $Q(\cdot)$ is a probability measure

Pf : (i) $\forall \; A \in \mathcal{F}$

$$Q(A) = P(A|B) = \frac{P(AB)}{P(B)} \geq 0$$

(ii) $$Q(\Omega) = P(\Omega|B) = \frac{P(\Omega B)}{P(B)} = 1$$

(iii) If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint, then

$$Q\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_i^{\infty} A_i \,|\, B\right) = \frac{P\left(\left(\bigcup_i^{\infty} A_i\right) \cap B\right)}{P(B)}$$

$$= \frac{P\left(\bigcup_i^{\infty} A_i B\right)}{P(B)} = \sum_i^{\infty} \frac{P(A_i B)}{P(B)} = \sum_i^{\infty} P(A_i|B)$$

$$= \sum_i^{\infty} Q(A_i)$$

($A_i B$ s are disjoint)

Note : $P(A \cap B) = P(A)\,P(B|A) = P(B)\,P(A|B)$ if $P(A), P(B) > 0$

Note : We can interpret cond prob by restricting the sample space to B

$\mathcal{F}_B = \{ A \cap B : A \in \mathcal{F} \}$ - σ-field of subsets of B

$(B, \mathcal{F}_B, Q)$ as prob space

Note: If $P(B) > 0$ and $B \subseteq A$, then $P(A|B) = 1$

If $P(B) > 0$ and $A \cap B = \phi$, then $P(A|B) = 0$

Example: 5 cards drawn at random (WOR) from a pack of 52 cards

B: all are spades

A: at least 4 are spades

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B)}{P(A)} \qquad B \subseteq A$$

$$= \frac{\binom{13}{5} / \binom{52}{5}}{\left(\binom{13}{4}\binom{39}{1} + \binom{13}{5}\right) / \binom{52}{5}} = \dots$$

Multiplication law

(i) $P(AB) = P(A)\,P(B|A)$ If $P(A) > 0$

$= P(B)\,P(A|B)$ If $P(B) > 0$

(ii) $P(ABC) = P(AB)\,P(C|AB)$

$= P(A)\,P(B|A)\,P(C|AB)$

provided $P(AB) > 0$ (ensures that $P(A) > 0$)

Chain in any order is possible

(iii) $P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1 A_2 \dots A_n)$

$= \cancel{P(A_n A_2 \dots A_{n-1} | A_n)}$

$= P(A_1, \dots A_{n-1})\,P(A_n | A_1 \dots A_{n-1})$

$= P(A_1, \dots A_{n-2})\,P(A_{n-1} | A_1 \dots A_{n-2})\,P(A_n | A_1, \dots A_{n-1})$

$\vdots$

$= P(A_1)\,P(A_2|A_1)\,P(A_3|A_1 A_2) \dots P(A_n | A_1 \dots A_{n-1})$

provided $P(A_1 \dots A_{n-1}) > 0$ (ensures $P(A_1 \dots A_i) > 0$, $i = 1, \dots n-2$)

Theorem of total probability

Let $A_1, A_2 \ldots$ be mutually exclusive and exhaustive events $\in \mathcal{F}$ (i.e. $A_i \cap A_j = \phi \; \forall \, i \neq j$ & $\cup A_i = \Omega$).

Suppose B is any other event ($B \in \mathcal{F}$)

Then $B = B \cap \Omega = B \cap (\bigcup_i A_i) = \bigcup_1^{\infty} (A_i \cap B)$

$P(B) = P(\bigcup_1^{\infty} A_i B) = \sum_1^{\infty} P(A_i B)$ ($A_i B$ s are m.e)

$= \sum_{i=1}^{\infty} P(A_i) P(B|A_i)$ (provided $P(A_i) > 0$)

Bayes theorem

Suppose that $A_1, A_2, \ldots$ are mutually exclusive and exhaustive and B be any other event $\ni P(B) > 0$,

then

$$P(A_K | B) = \frac{P(A_K B)}{P(B)} = \frac{P(A_K) P(B|A_K)}{\sum_1^{\infty} P(A_K) P(B|A_K)}$$

$P(A_K)$: prior prob

$P(A_K|B)$: posterior prob.

$\rightarrow$ Thm of total probability

Independence of events

Def^n: Let $(\Omega, \mathcal{F}, P)$ be a prob space. $A, B \in \mathcal{F}$ are independent if

$$P(AB) = P(A) P(B)$$

Remark: Intuitively

$P(A|B) = P(A)$ $\quad$ $P(B|A) = P(B)$ $\quad$ with $P(A), P(B) > 0$

Multiplication rule: $P(AB) = P(A) P(B|A) = P(B) P(A|B)$

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

The above def^n however does not require the $P(A), P(B) > 0$

Result: If A & B are indep, then

(i) A^c & B are indep

(ii) A & B^c are indep

(iii) A^c & B^c are indep

Pf: (i)

$$B = AB \cup A^c B$$

$$P(B) = P(AB) + P(A^c B)$$

i.e. $P(A^c B) = P(B) - P(AB)$

$$= P(B) - P(A)P(B) = P(B)P(A^c)$$

(ii) Sly $A = AB \cup AB^c$

$$P(AB^c) = P(A) - P(AB)$$
$$= P(A)P(B^c)$$

(iii)

$$P(A^c B^c) = P(A \cup B)^c$$
$$= \cancel{B}\, 1 - P(A \cup B)$$
$$= 1 - P(A) - P(B) + P(AB)$$
$$= 1 - P(A) - P(B) + P(A)P(B)$$
$$= P(A^c)P(B^c)$$

Def^n: Events $A_1, \ldots A_n$ are pairwise indep if

$$P(A_i A_j) = P(A_i)P(A_j) \quad \forall i \neq j$$

Def^n: Events $A_1, A_2, \ldots A_n$ are mutually indep if

(i) $P(A_i A_j) = P(A_i)P(A_j) \quad \forall i \neq j$

(ii) $P(A_i A_j A_k) = P(A_i)P(A_j)P(A_k) \quad \forall i \neq j \neq k$

$\vdots$

(n-1) $P(A_1 \ldots A_n) = P(A_1) \ldots P(A_n)$.

Note: For a countable collection $\{A_1, A_2, \dots\}$; we say that this class is of indep events if every finite subclass of these events is mutually indep.

Note: Mutual indep $\Rightarrow$ pairwise indep

Converse is not true

Counter example

$\Omega = \{1, 2, 3, 4\}$; $\mathcal{F} = \mathcal{P}(\Omega) \rightarrow$ powerset of Ω

$(\Omega, \mathcal{F}, P)$: prob space

$P(\{i\}) = \frac{1}{4}$ $\quad i = 1, 2, 3, 4$

$A = \{1, 4\}$, $B = \{2, 4\}$, $C = \{3, 4\}$

$P(A) = P(B) = P(C) = \frac{1}{2}$

$P(AB) = \frac{1}{4} = P(AC) = P(BC) = P(\{4\})$

$P(ABC) = P(\{4\}) = \frac{1}{4}$

$P(AB) = P(A)P(B)$; $P(AC) = P(A)P(C)$; $P(BC) = P(B)P(C)$

$\Rightarrow$ A, B, C are pairwise indep

But $\quad P(ABC) = \frac{1}{4} \neq P(A)P(B)P(C) = \frac{1}{8}$

$\Rightarrow$ A, B, C are not indep.

Note: if ~~the~~ collection is not p.i then it can't be m.i.

$A = \{1, 2\}$, $B = \{2, 3\}$, $C = \{3, 4\}$

$P(AC) = 0 \neq P(A)P(C)$

A, B, C can't be m.i.

Note : If $A_1, \ldots A_n$ are m.i. then collection of events and complementary events set would be indep

i.e for any $k \in \{1, 2, \ldots n-1\}$ and $(\alpha_1, \ldots \alpha_n)$ of $(1, \ldots n)$ events $A_{\alpha_1}, \ldots A_{\alpha_k}, A^c_{\alpha_{k+1}} \ldots A^c_{\alpha_n}$ are indep.

Continuity of probability measure

Def^n : $(\Omega, \mathcal{F}, P)$: prob space

$A_1, \ldots$ events $\{A_n : n = 1, 2, \ldots\}$ seq of events in $\mathcal{F}$

(i) $A_n \uparrow$ if $A_n \subseteq A_{n+1}$, $n = 1, 2, \ldots$

(ii) $A_n \downarrow$ if $A_{n+1} \subseteq A_n$, $n = 1, 2, \ldots$

(iii) A_n is monotone if either $A_n \downarrow$ or $A_n \uparrow$

(iv) If $A_n \uparrow$, we define

$$\lim_{n \to \infty} A_n = \bigcup_{n=1}^{\infty} A_n$$

(v) If $A_n \downarrow$, we define

$$\lim_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} A_n$$

Continuity of probability measure

Let $\{A_n : n = 1, 2, \ldots\}$ be a sequence of monotone events in $(\Omega, \mathcal{F}, P)$, then

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

Random variable

$(\Omega, \mathcal{F}, P)$: prob space

In many situations, we may not be directly interested in the sample space Ω or the $\mathcal{F}$; rather we may be interested in some numerical aspect of Ω, i.e. assignment of numbers to elements of Ω.

e.g : Interested to know prop of defective items in a lot

Sample of size n is drawn

Sample space : 2^n elements of the form

$(a_1, a_2, \dots, a_n)$; $a_i = D$ if item is def

$= N$ if item is non-def

$$\Omega \xrightarrow{X} \{0, 1, \dots, n\}$$

$X(a_1, \dots a_n) = r$ if r of a_is are D

e.g : fair coin tossed 2 times

$$\Omega = \{HH, HT, TH, TT\}$$

$$P(\{\omega\}) = \frac{1}{4} \quad \forall \omega \in \Omega$$

$X(\omega)$: # of heads in ω

$X : \Omega \to \mathbb{R}$

$$X(\omega) = \begin{cases} 0, & \text{if } \omega = TT \\ 1, & \text{if } \omega = TH \text{ or } HT \\ 2, & \text{if } \omega = HH \end{cases}$$

$\omega : X(\omega) = 0$

$$P(X=0) = P(TT) = \frac{1}{4}$$

$$P(X=1) = P(TH \text{ or } HT) = \frac{1}{2}$$

$$P(X=2) = P(HH) = \frac{1}{4}$$

$$P(X \in \{0, 1, 2\}) = 1$$

<u>Def^n</u>: $(\Omega, \mathcal{F}, P)$ be prob space

A real valued f^n $X : \Omega \to \mathbb{R}$ defined on the sample space Ω is called a random variable.

<u>Remark</u>: A more advanced textbook on prob would define r.v. as.

A real valued f^n $X : \Omega \to \mathbb{R}$ is called a r.v. If the inverse images under X of all Borel sets in $\mathbb{R}$ are events, i.e. If

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F} \quad \forall\, B \in \mathcal{B}. \quad -(*)$$

Def^n: $(\Omega, \mathcal{F}, P)$ be prob space

A real valued f^n $X : \Omega \to \mathbb{R}$ defined on the sample space Ω is called a random variable.

Remark: A more advanced textbook on prob would define r.v. as.

A real valued f^n $X : \Omega \to \mathbb{R}$ is called a r.v. if the inverse images under X of all Borel sets in $\mathbb{R}$ are events, i.e. if

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{B}. \quad -(*)$$

Further, to check whether a real valued f^n on $(\Omega, \mathcal{F})$ is a r.v., it is not necessary to check $(*)$ $\forall$ Borel sets $B \in \mathcal{B}$. It suffices to verify $(*)$ for any class of subsets of $\mathbb{R}$ that generates $\mathcal{B}$; e.g. we can take the class of subsets as semiclosed intervals $(-\infty, x]$, $x \in \mathbb{R}$ or $(-\infty, x)$, $x \in \mathbb{R}$. In such a case, we would say

X is a r.v. iff $\forall x \in \mathbb{R}$

$$X^{-1}(-\infty, x] = \{\omega : X(\omega) \le x\} \in \mathcal{F}.$$

Ex: $\Omega = \{HH, TH, HT, TT\}$; $\mathcal{F}$: power set of Ω

$X(\omega)$: # of Hs in $\{\omega\}$ $\quad X(\omega) = \begin{cases} 0, & TT \\ 1, & TH, HT \\ 2, & HH \end{cases}$

To show that X is r.v., we look at

$$X^{-1}(-\infty, x] = \{\omega : X(\omega) \le x\} = \begin{cases} \phi, & x < 0 \\ \{TT\}, & 0 \le x < 1 \\ \{TT, HT, TH\} & 1 \le x < 2 \\ \Omega & x \ge 2 \end{cases}$$

$\Rightarrow \downarrow \in \mathcal{F} \quad \forall x \in \mathbb{R}$

$\Rightarrow X$ is a r.v.

Induced probability space

$(\Omega, \mathcal{F}, P)$: prob space

$X : \Omega \to \mathbb{R}$ a r.v.

$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{B}$

Define a set f^n $P_X : \mathcal{B} \to [0,1]$

$$P_X(B) = P(\omega \in \Omega : X(\omega) \in B) = P(X^{-1}(B))$$

$$(\Omega, \mathcal{F}, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}, P_X)$$

This is a prob space with $P_X(.)$ as a probability measure, referred to as the induced prob measure

$(\mathbb{R}, \mathcal{B}, P_X)$ is the induced prob space, induced by X.

Distribution function of a random variable

<u>Def^n</u>: Let X be a r.v. defined on a prob space $(\Omega, \mathcal{F}, P)$ and let $(\mathbb{R}, \mathcal{B}, P_X)$ be the prob space induced by X. Define $F_X : \mathbb{R} \to \mathbb{R}$ as

$$F_X(x) = P(\omega : X(\omega) \leq x) = P_X(-\infty, x]$$

$F_X(.)$ is called the cumulative dist^n f^n or just dist^n f^n of r.v. X

<u>Remark</u>: As ~~intervals~~ class of intervals of the type $(-\infty, x]$ generates $\mathcal{B}$, c.d.f $F_X(.)$ determine the $P_X(.)$ uniquely.

Thus to study the random behavior of r.v. X it suffices to study it's c.d.f F.

<u>Examples</u>

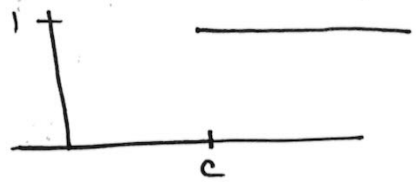
(1) $(\Omega, \mathcal{F}, P)$

$$X(\omega) = c \quad \forall\ \omega \in \Omega$$

$$P(X = c) = P(\omega : X(\omega) = c) = P(\Omega) = 1$$

$$F(x) = P(X \le x) = P(\omega : X(\omega) \le x)$$

$$= \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}$$



Note that

$F(-\infty) = 0$; $F(\infty) = 1$
F is non-decreasing
F is right continuous — (*)

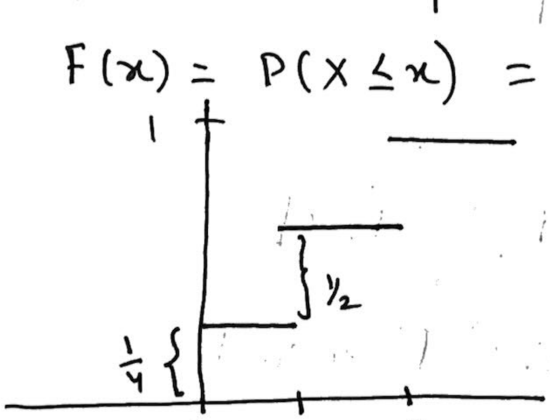
F(.) have 1 pt of jump discontinuity

(2) $\Omega = \{HH, HT, TH, TT\}$ example

$X(\omega)$: # of heads in ω

$$P(X=0) = \frac{1}{4} \ ; \ P(X=1) = \frac{1}{2} \ ; \ P(X=2) = \frac{1}{4}$$

$$F(x) = P(X \le x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & 0 \le x < 1 \\ \frac{1}{4} + \frac{1}{2}, & 1 \le x < 2 \\ \frac{1}{4} + \frac{1}{2} + \frac{1}{4}, & x \geq 2 \end{cases} = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & 0 \le x < 1 \\ 3/4, & 1 \le x < 2 \\ 1, & x \geq 2 \end{cases}$$



Once again (*) is satisfied by the above F(.)

F(.) has 3 pts of jump discontinuity

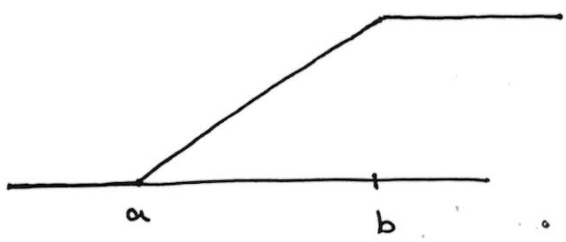
Example 3:

$$\Omega = [a, b]$$

For every $I \subset \Omega$; $P(I) = \dfrac{\text{length of } I}{b-a}$

Define. $X(\omega) = \omega$; $\omega \in \Omega$

$$F_X(x) = P(X \le x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x < b \\ \dfrac{b-a}{b-a} = 1, & x \ge b \end{cases}$$



$F(.)$ satisfies (*)

$F(.)$ is continuous everywhere

Result: Let $F(.)$ be the d.f. of a r.v. X defined on a prob space $(\Omega, \mathcal{F}, P)$. Then

(i) F is non-decreasing

(ii) F is right continuous

(iii) $F(-\infty) = \lim_{n \uparrow \infty} F(-n) = 0$ and

$F(\infty) = \lim_{n \uparrow \infty} F(n) = 1$

Pf:

(a) Let $-\infty < x < y < \infty$, then

$$(-\infty, x] \subseteq (-\infty, y]$$

$$\Rightarrow P_X(-\infty, x] \le P_X(-\infty, y]$$

i.e $P(\omega : X(\omega) \le x) \le P(\omega : X(\omega) \le y)$

$$\Rightarrow F(x) \le F(y)$$

$\Rightarrow F(.)$ is non-decreasing

(b) $$F(x+) = \lim_{h \downarrow 0} F(x+h)$$

$$= \lim_{n \to \infty} F\left(x + \frac{1}{n}\right)$$

$$= \lim_{n \to \infty} P_X\left(\left(-\infty, x + \frac{1}{n}\right]\right)$$

Realize that $A_n = \left(-\infty, x + \frac{1}{n}\right]$, $n = 1, 2, \ldots$ is $\ni A_n \downarrow$

and $\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right] = (-\infty, x]$

$$\Rightarrow \lim_{n \to \infty} P_X\left(\left(-\infty, x + \frac{1}{n}\right]\right) = P_X\left(\lim_{n \to \infty} A_n\right)$$

$$= P_X\left(\bigcap_{1}^{\infty} A_n\right)$$

$$= P_X\left(\bigcap_{1}^{\infty} \left(-\infty, x + \frac{1}{n}\right]\right)$$

$$= P_X\left((-\infty, x]\right) = F(x)$$

$\Rightarrow F(x+) = F(x)$; i.e. $F(.)$ is right continuous

(b) $$F(x+) = \lim_{h \downarrow 0} F(x+h)$$

$$= \lim_{n \to \infty} F\left(x + \frac{1}{n}\right)$$

$$= \lim_{n \to \infty} P_X\left(\left(-\infty, x + \frac{1}{n}\right]\right)$$

Realize that $A_n = \left(-\infty, x + \frac{1}{n}\right]$, $n = 1, 2, \ldots$ is $\ni A_n \downarrow$

and $\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right] = (-\infty, x]$

$$\Rightarrow \lim_{n \to \infty} P_X\left(\left(-\infty, x + \frac{1}{n}\right]\right) = P_X\left(\lim_{n \to \infty} A_n\right)$$

$$= P_X\left(\bigcap_{1}^{\infty} A_n\right)$$

$$= P_X\left(\bigcap_{1}^{\infty} \left(-\infty, x + \frac{1}{n}\right]\right)$$

$$= P_X\left((-\infty, x]\right) = F(x)$$

$\Rightarrow F(x+) = F(x)$; i.e. $F(.)$ is right continuous

(c) $$F(-\infty) = \lim_{n \to \infty} F(-n)$$

$$= \lim_{n \to \infty} P_X\left((-\infty, -n]\right)$$

$$= P_X\left(\bigcap_{1}^{\infty} (-\infty, -n]\right) \quad \left((-\infty, -n] \downarrow\right)$$

$$= P_X(\phi) = 0 \quad \bigcap_{1}^{\infty} (-\infty, -n] = \phi$$

$$F(+\infty) = \lim_{n \to \infty} F(n) \quad \left((-\infty, n] \uparrow\right)$$

$$= \lim_{n \to \infty} P_X\left((-\infty, n]\right)$$

$$= P_X\left(\lim_{n \to \infty} (-\infty, n]\right) = P_X\left(\bigcup_{1}^{\infty} (-\infty, n]\right)$$

$= P_X(\mathbb{R})$ $\quad\quad\quad\quad \left(\bigcup_{n=1}^{\infty}(-\infty, n] = \mathbb{R}\right)$

$= 1$

Remark: Converse of the prev result is true, i.e. If $F(.)$ be a function $F: \mathbb{R} \to [0, 1] \ni$

(i) $F(.)$ is non-decreasing

(ii) $F(.)$ is right continuous

(iii) $F(-\infty) = 0$ and $F(\infty) = 1$

then $F(.)$ is d.f. of some appropriate random variable

Remark: For a d.f. $F(.)$, both $F(x+)$ and $F(x-)$ exist $\forall x \in \mathbb{R}$ as $F(.)$ is monotone, bdd below and bdd above.

Remark: A d.f. is continuous at $a \in \mathbb{R}$ iff $F(a) = F(a-)$

Remark: For any $a \in \mathbb{R}$

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

If d.f. is continuous at a, then $P(X = a) = 0$

Remark: For $-\infty < a < b < \infty$

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

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

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

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

If $F(.)$ is continuous at a and b, then all the above are equal and equal to $F(b) - F(a)$

Remark: $F(.)$ can only have points of jump discontinuities and can have atmost countable number of such jump discontinuities

Discrete random variable

$(\Omega, \mathcal{F}, P)$: prob space

$X : \Omega \to \mathbb{R}$ be a r.v.

$(\mathbb{R}, \mathcal{B}, P_X)$: induced prob space (induced by X)

$F(.)$: d.f. of X

Def^n: Random variable X is said to be a discrete r.v.

If $\exists$ a countable set $D \subset \mathbb{R}$ $\ni$

$$P(X = x) = F(x) - F(x-) > 0 \quad \forall\, x \in D$$

and $P(X \in D) = 1$

$\because$ D is called the support of the r.v. X

D is the set of all discontinuity pts of $F(.)$

Def^n: Let $D = \{x_1, x_2, \ldots\}$ $\leftarrow$ countable (finite or infinite)

$$P(X = x_i) = p_i \text{, say}, \quad p_i > 0 \;\forall\, i$$

$$P(X \in D) = \sum_i p_i = 1$$

The collection $\{p_1, p_2, \ldots\}$ is called the probability mass f^n of r.v. X.

i.e. $f_X(x) = P(X = x) \quad x \in S$ is the p.m.f. of X

$= F(x) - F(x-)$; $f(x) > 0 \;\forall\, x \in D$

$\sum_{x \in D} f(x) = 1$

Remark: (i) d.f. of a discrete r.v. increases only by jumps

(ii) number of jump discontinuities are atmost countable

(iii) d.f. determines the p.m.f uniquely and vice-versa

Example :

(1) X : r.v.

$F(.)$: d.f. of X

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & 0 \le x < 1 \\ \frac{1}{3}, & 1 \le x < 2 \\ \frac{3}{4}, & 2 \le x < 3 \\ 1, & x \geq 3 \end{cases}$$

pts of jump discontinuities $\{0, 1, 2, 3\} = D$
(finite collection)

$P(X \in D) = 1$;

X is a discrete r.v. with support D

p.m.f.

x	$P(X = x)$
0	$F(0) - F(0-) = \frac{1}{4}$
1	$F(1) - F(1-) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$
2	$F(2) - F(2-) = \frac{3}{4} - \frac{1}{3} = \frac{5}{12}$
3	$F(3) - F(3-) = \frac{1}{4}$

i.e. p.m.f is

$$f(x) = \begin{cases} \frac{1}{4}, & \text{If } x = 0, 3 \\ \frac{1}{12}, & \text{If } x = 1 \\ \frac{5}{12}, & \text{If } x = 2 \\ 0, & \text{o/w}. \end{cases}$$

Example :

(2) Random exp : tossing coin until head appears

$\Omega = \{H, TH, TTH, \ldots\}$

X : r.v. which counts number of tosses reqd to get 1^{st} H

i.e. $X(\omega) =$ no. of T in ω + 1

possible values of X : 1, 2, 3, ...

$$P(X = i) = \frac{1}{2^i} \; ; \; i = 1, 2, \ldots$$

d.f. of X

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{2}, & 1 \le x < 2 \\ \frac{3}{4}, & 2 \le x < 3 \\ \vdots & \end{cases}$$

magnitude of jump at i $= \frac{1}{2^i}$ for $i = 1, 2, \ldots$.

$D = \{1, 2, 3, \ldots\}$

countably infinite

Support

p.m.f. $\quad f(x) = \frac{1}{2^x} \; ; \; x = 1, 2, \ldots$.

Continuous random variable

Def^n: A random variable X is said to be a continuous r.v. if $\exists$ a non-negative, integrable function $f : \mathbb{R} \to [0, \infty)$ such that for any $x \in \mathbb{R}$

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

$f(.)$ is called the probability density function (p.d.f.) of X.

Remark: Support of a continuous r.v is the set

$$S = \{ x \in \mathbb{R} : F(x+h) - F(x-h) > 0, \ \forall h > 0 \}$$

Remark: For a cont r.v. ($F(.)$ is continuous everywhere)

$$P(X = x) = F(x) - F(x-) = 0 \quad \forall x \in \mathbb{R}$$

In general, suppose $A \subset \mathbb{R}$ is any countable subset, then

$$P(X \in A) = \sum_{x \in A} P(X = x) = 0$$

Remark: p.d.f. $f(x)$, then

(i) $f(x) \geq 0 \quad \forall x \in \mathbb{R}$

and (ii) $\int_{-\infty}^{\infty} f(t)\, dt = 1$

Converse is also true

Remark: F d.f. of r.v. X, if F is differentiable, then

$$f(x) = \frac{d}{dx} F(x)$$

Remark: For a Continuous r.v.

$$P(X < x) = P(X \le x) = F(x) \quad \forall\, x \in \mathbb{R}$$

$$P(X \ge x) = 1 - P(X < x) = 1 - F(x) \quad \forall\, x \in \mathbb{R}$$

$\forall \quad -\infty < a < b < \infty$

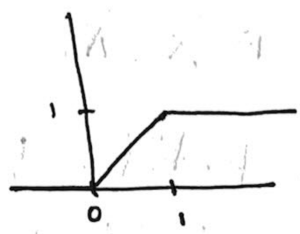
$$P(a < X < b) = P(a \le X < b) = P(a < X \le b)$$
$$= P(a \le X \le b)$$
$$= F(b) - F(a)$$

$$\& \quad F(b) - F(a) = \int_{-\infty}^{b} f(t)\,dt - \int_{-\infty}^{a} f(t)\,dt$$
$$= \int_{a}^{b} f(t)\,dt \,.$$

Remark: p.d.f. determines the d.f. uniquely. Converse is not true

Examples

(1)
$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$



F(.) is cont everywhere

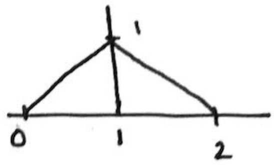
p.d.f. of X is
$$\frac{d}{dx} F(x) \rightarrow f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & o/w \end{cases}$$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}$$

(2)

p.d.f. of a r.v. X is

$$f(x) = \begin{cases} x, & 0 \le x \le 1 \\ 2-x, & 1 \le x \le 2 \\ 0, & o/w \end{cases}$$



$f(x) \geq 0 \;\forall x$

$$\int_{-\infty}^{\infty} f(t)\,dt = \int_0^1 x\,dx + \int_1^2 (2-x)\,dx$$

$$= \frac{x^2}{2}\Big|_0^1 + \left(2x - \frac{x^2}{2}\right)\Big|_1^2 = \frac{1}{2} + \left(2 - \frac{3}{2}\right) = 1$$

$\Rightarrow f(\cdot)$ is a p.d.f.

dist^n f^n: $F(x) = \int_{-\infty}^{x} f(t)\,dt$

$$= \begin{cases} 0, & x < 0 \\ x^2/2, & 0 \le x \le 1 \\ \frac{1}{2} + \left(\left(2x - \frac{x^2}{2}\right) - \frac{3}{2}\right), & 1 \le x \le 2 \\ 1, & x \geq 2 \end{cases}$$

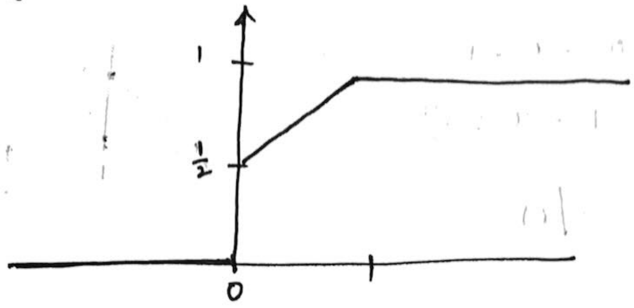
$$P\left(\frac{1}{4} \le X \le \frac{3}{2}\right) = F(3/2) - F(1/4)$$

<u>Remark</u>: There are random variables that are neither discrete nor continuous - random variables of mixed type

<u>Example</u>:

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x+1}{2}, & 0 \le x < 1 \\ 1, & x \geq 1 \end{cases}$$

It's easy to check that $F(.)$ is a d.f.



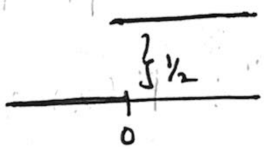
$F(.)$ has jump discontinuity at 0 - jump size $\frac{1}{2}$

$F(.)$ is continuous everywhere, except at 0

$$P(X=0) = F(0) - F(0-) = \frac{1}{2}$$

Discrete part:
(f^n which increases by jump only)

$$F_d = \begin{cases} 0, & x<0 \\ \frac{1}{2}, & x \geq 0 \end{cases}$$



Continuous part:
(increasing continuously part)

$$F_c = \begin{cases} 0, & x<0 \\ x/2, & 0 \leq x < 1 \\ \frac{1}{2}, & x \geq 1 \end{cases}$$

$$F_c = F - F_d$$

Note: F_d & F_c are not d.f.s.

Realise that

$$F(x) = \frac{1}{2} F_1(x) + \frac{1}{2} F_2(x)$$

$$F_1(x) = \begin{cases} 0, & x<0 \\ 1, & x \geq 0 \end{cases} \qquad F_2(x) = \begin{cases} 0, & x<0 \\ x, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

$F_1(x)$ & $F_2(x)$ are proper d.f.s.

d.f. of a discrete r.v.

d.f. of a continuous r.v.

Remark:

$\alpha F_1(x) + (1-\alpha) F_2(x)$ will be d.f. of mixed type

$\forall \alpha \ni 0 < \alpha < 1$

if $\alpha = 0$; X is continuous r.v.

If $\alpha = 1$. X is discrete r.v.

Remark: Any dist^n f^n $F(.)$ can be expressed as.

$$F(x) = \alpha F_d(x) + (1-\alpha) F_c(x)$$

d.f. of discrete r.v.   d.f. of cont r.v.

Example: Let X be r.v. with d.f.

$$F(x) = \begin{cases} 0, & x < 0 \\ x/4, & 0 \le x < 1 \\ x/3, & 1 \le x < 2 \\ 1, & x \geq 2 \end{cases}$$

jump discontinuities at $x = 1, 2$

$D = \{1, 2\}$; $D \neq \phi$

$\Rightarrow X$ is not cont r.v.

$$P(X \in D) = \sum_{x \in D} P(X = x)$$

$$= \sum_{x \in D} (F(x) - F(x-))$$

$$= \sum_{x = 1,2} (F(x) - F(x-))$$

$$= \left(\frac{1}{3} - \frac{1}{4}\right) + \left(1 - \frac{2}{3}\right) = \frac{1}{12} + \frac{1}{3} = \frac{5}{12}$$

$P(X \in D) = \frac{5}{12} < 1$

$\Rightarrow X$ is not discrete r.v.

$\Rightarrow X$ is neither discrete or cont

Discrete part of $F(.)$:

$$F_1(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{12}, & 1 \leq x < 2 \\ \frac{5}{12} & x \geq 2 \end{cases}$$

Take $\alpha = \frac{5}{12}$; $\alpha F_d(x) = F_1(x)$

$$F_d(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{5}, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

$F_d(x)$ is d.f. of a discrete r.v. $\rightarrow$ p.m.f. $\left\{ \begin{array}{l} P(X_d = 1) = \frac{1}{5} \\ P(X_d = 2) = \frac{4}{5} \end{array} \right\}$.

$F_2(x)$: continuous part of $F(.)$

$F_2(x) = F(x) - F_1(x)$

$$F_2(x) = \begin{cases} 0, & x < 0 \\ x/4, & 0 \leq x < 1 \\ x/3 - \frac{1}{12}, & 1 \leq x < 2 \\ \frac{7}{12} & x \geq 2 \end{cases}$$

$F_2(x) = (1-\alpha) F_c(x)$; $1 - \alpha = \frac{7}{12}$

$$F_c(x) = \begin{cases} 0, & x < 0 \\ 3x/7, & 0 \leq x < 1 \\ \frac{12}{7}\left(\frac{x}{3} - \frac{1}{12}\right) = \frac{4}{7}x - \frac{1}{7}, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

$F_c(x)$ is continuous everywhere

$F_c(x)$ is the d.f. of a cont r.v.

p.d.f. of X_c

$$f_{X_c}(x) = \begin{cases} 3/7, & 0 \leq x < 1 \\ 4/7, & 1 \leq x < 2 \\ 0, & o/w \end{cases}$$

$$f_{X_c}(x) \geq 0 \quad \forall x$$

$$\int_{-\infty}^{\infty} f_{X_c}(x)\, dx = \int_0^1 \frac{3}{7}\, dx + \int_1^2 \frac{4}{7}\, dx = 1$$

$$F(x) = \alpha F_d(x) + (1-\alpha) F_c(x)$$

Mathematical Expectation

$$(\Omega, \mathcal{F}, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}, P_X)$$

$$g : \mathbb{R} \xrightarrow{g} \mathbb{R}$$

Expected value of $g(X)$: mathematical expectation of $g(X)$

$E(g(X))$ exists if $E|g(X)| < \infty$

Suppose, X is a discrete r.v. with p.m.f.

$X : x_1, x_2, - - -$

$P(X = x) : p_1, p_2, - - -$

$E\, g(X)$ is said to exist and equals $\sum_{i=1}^{\infty} g(x_i)\, p_i$ provided $\sum_i |g(x_i)|\, p_i < \infty$

If X is continuous with p.d.f. $f_X(x)$, then $E(g(X))$ exists and equals $\int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$ provided $\int_{-\infty}^{\infty} |g(x)|\, f_X(x)\, dx < \infty$

Special cases

(i) $g(X) = X$

$E\, g(X) = E\, X = \mu_1'$: mean of the distn of X

(ii) $g(X) = X^n$: n is a positive integer

$E\, g(X) = E\, X^n = \mu_n'$

n^{th} moment about origin of r.v. X

(iii) $g(X) = (X-a)^n$

$E\,g(X) = E(X-a)^n$: nth moment of X about the pt a

If $a = E(X)$, then

$E(X - E(X))^n = \mu_n$: nth order central moment of X

$n = 2$; $\mu_2 = E(X - E(X))^2 \rightarrow$ variance of X

$= \sigma^2$

$\mu_2^{1/2} = \sigma$: standard deviation of X

(iii) $g(X) = (X-a)^n$

$E\,g(X) = E(X-a)^n$: n^{th} moment of X about the pt a

If $a = E(X)$, then

$E(X - E(X))^n = \mu_n$: n^{th} order central moment of X

$n=2$; $\mu_2 = E(X - E(X))^2 \rightarrow$ variance of X

$= \sigma^2$

$\mu_2^{1/2} = \sigma$: standard deviation of X

Remark: Measure of skewness

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \left(\frac{\mu_3^2}{\mu_2^3}\right)^{1/2} = \frac{\mu_3}{(\sigma^2)^{3/2}} = \frac{\mu_3}{\sigma^3}$$

$= 0$ for distn sym about $E(X)$

> 0 for positively skewed distn

< 0 for negatively skewed distn

Remark: Measure of peakedness.

Kurtosis measure $\gamma_2 = \dfrac{\mu_4}{\sigma^4}$

Remark: If $g_1, g_2, \ldots g_r$ be r real valued f^ns on $\mathbb{R}$ and let X be a r.v. $\ni E\,g_i(X)$ exists for $i = 1, \ldots r$, then $E\left(\sum_{i=1}^{r} g_i(X)\right)$ exists and is equal to $\sum_{i=1}^{r} E\,g_i(X)$

e.g.: $\mu_n = E(X - E(X))^n$

$$= E\left(X^n + \binom{n}{1} X^{n-1}(-E(X)) + \binom{n}{2} X^{n-2}(-E(X))^2 + \cdots + (-1)^n E X^n\right)$$

i.e. $\mu_n = \mu_n' - \binom{n}{1}\mu_{n-1}'\mu_1' + \binom{n}{2}\mu_{n-2}'(\mu_1')^2 - \cdots + (-1)^n(\mu_1')^n$

$\underline{n=2}$

$$\sigma^2 = \mu_2 = E(X - E(X))^2$$
$$= \mu_2' - 2(\mu_1')^2 + (\mu_1')^2$$
$$= \mu_2' - (\mu_1')^2$$
$$= EX^2 - (E(X))^2$$

<u>Remark</u>: Suppose EX^m exists for a positive int m then $\forall$ positive int $K \ni K \leq m$, EX^K exists

$$\int_{-\infty}^{\infty} |x|^K f(x)\,dx = \int_{|x|\leq 1} |x|^K f(x)\,dx + \int_{|x|\geq 1} |x|^K f(x)\,dx$$

(for cont case)

$$\leq \int_{|x|\leq 1} f(x)\,dx + \int_{|x|>1} |x|^K f(x)\,dx$$

$$\leq \int_{|x|\leq 1} f(x)\,dx + \int_{|x|>1} |x|^m f(x)\,dx$$

$$\leq \int_{-\infty}^{\infty} f(x)\,dx + \int_{-\infty}^{\infty} |x|^m f(x)\,dx$$

$$= 1 + E|X|^m < \infty$$

$\Rightarrow$ EX^K exists.

Moment Generating function (m.g.f)

Defⁿ: Let X be r.v. The function

$$M_X(t) = E(e^{tX})$$ is known as the m.g.f of the r.v. X if the expectation exists in some neighborhood of origin

Note: If m.g.f. exists, it determines the d.f. uniquely.

Note: Suppose all derivates of $M_X(t)$ exists at $t=0$ and can be obtained by differentiating under the expectation, then

$$\frac{\partial^k}{\partial t^k} M_X(t)\Big|_{t=0} = \frac{\partial}{\partial t^k} E(e^{tX})\Big|_{t=0}$$

$$= E\left(\frac{\partial^k}{\partial t^k} e^{tX}\Big|_{t=0}\right)$$

$$= E\left(X^k e^{tX}\Big|_{t=0}\right) = \mu_k'$$

i.e. $M_X(t)$ generates the moments of X.

Note: Taylor series expansion of $M_X(t)$ about 0 gives

$$M_X(t) = M_X(0) + \frac{M_X'(0)}{1!} t + \frac{M_X''(0)}{2!} t^2 + \cdots$$

Coeff of $\frac{t^k}{k!} = \mu_k' = EX^k$

<u>Note</u>: Although m.g.f. exists for most of the common distributions, there are cases when it does not exist e.g. a Cauchy distⁿ

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} ; \quad x \in \mathbb{R}$$

Some standard inequalities

(1) Chebyshev's inequality

X is a r.v. with $E(X) = \mu$ and $V(X) = \mu_2 = \sigma^2$

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \quad \forall \epsilon > 0$$

(2) Generalization of Chebyshev's inequality

$h : \mathbb{R} \to \mathbb{R}$ be non-negative f^n $E\,h(X)$ is finite.

then $\forall c > 0$

$$P(h(X) \geq c) \leq \frac{E\,h(X)}{c} ;$$

$h(X) = |X|^r$ gives Markov's inequality

(3) Let g be a non-negative and strictly increasing function $g : [0, \infty) \to \mathbb{R}$ such that $E\,g(X)$ is finite, then for any $c > 0$ $\ni$ $g(c) > 0$

$$P(|X| \geq c) \leq \frac{E(g(|X|))}{g(c)}$$

(4) Jensen's inequality

Let $\psi : (a, b) \to \mathbb{R}$ be a convex function and let X be r.v. with d.f. F having support $S \subseteq (a, b)$.

Then

$$E\,\psi(X) \geq \psi(E(X))$$

provided the expectations exist.

Note : From Jensen's inequality it follows that, for any r.v. X

$$E X^2 \geq (E X)^2$$

$$E|X| \geq |E(X)|$$

$$E e^X \geq e^{EX}$$

- For r.v. $X \ni P(X > 0) = 1$

$$E(\ln X) \leq \ln(E X)$$

Quantile & percentile

Def^n : Let $0 < p < 1$. The quantile of order p of the dist^n of a r.v. X, say ζ_p, is a point $\ni$

$$P(X < \zeta_p) \leq p \quad \text{and} \quad P(X \leq \zeta_p) \geq p$$

ζ_p is also called the $(100 \times p)$th percentile

Note : ζ_p is $\ni$

$$P(X \leq \zeta_p) \geq p \quad \text{and} \quad P(X \geq \zeta_p) \geq 1-p$$

$$\left(\begin{array}{l} 1 - P(X < \zeta_p) \geq 1-p \\ \text{i.e } P(X < \zeta_p) \leq p \end{array} \right).$$

Further

$$\text{i.e. } P(X \leq \zeta_p) - P(X = \zeta_p) \leq p$$

$$\cancel{P(X \leq \zeta_p) \geq p}$$

$$p \leq P(X \leq \zeta_p) \leq p + P(X = \zeta_p)$$

Note: For a continuous dist^n ζ_p is thus solution of

$$F(\zeta_p) = p$$

If $F(.)$ is strictly increasing then sol^n is unique, o/w there can be many solutions.

Note: For $p = \frac{1}{2}$, we get median of a distⁿ

i.e. If $z_{1/2} = \text{Med}$, then

$$P(X < \text{Med}) \leq \frac{1}{2} \text{ and } P(X \leq \text{Med}) \geq \frac{1}{2}$$

Like mean, $E(X)$, Med is a measure of central tendency.

Some Standard discrete $dist^n$s

(I) 1-point / degenerate $dist^n$

$$P(X=c)=1$$

d.f. $F(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}$

$E(X) = c \quad V(X) = 0$

m.g.f. $M_X(t) = E(e^{tX}) = e^{tc}$

(II) 2-pt $dist^n$

$X=x$	x_1	x_2	$(x_1 < x_2$, say$)$
$P(X=x)$	p	$1-p$	

d.f. $F(x) = \begin{cases} 0, & x < x_1 \\ p, & x_1 \leq x < x_2 \\ 1, & x_2 \leq x \end{cases}$

m.g.f. $M_X(t) = e^{tx_1} p + e^{tx_2}(1-p)$

$$E(X) = x_1 p + (1-p)x_2$$

$$EX^2 = x_1^2 p + x_2^2 (1-p)$$

$$VX = EX^2 - (E(X))^2 = \ldots$$

Sp. case: Bernoulli r.v. $\quad x_1 = 1, \; x_2 = 0$

success/failure

occurrence/non-occurrence

$X = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}$

$$M_X(t) = pe^t + (1-p)$$

$$EX = p \; ; \; V(X) = p(1-p) < EX$$

(III) Binomial

Suppose we perform n indep Bernoulli trials (outcome 0 or 1) with prob of 1 (success) in each trial p

X : r.v. counting the number of successes in n trials

Possible values of X : 0, 1, ... n

$$P(X=x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \ldots n \\ 0, & o/w \end{cases}$$

$X \sim Bin(n, p) \quad 0 < p < 1$

m.g.f : $M_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}$$

$$= ((1-p) + pe^t)^n$$

$$EX = \frac{\partial}{\partial t} M_X(t)\Big|_{t=0} = np$$

$$EX^2 = \frac{\partial^2 M_X(t)}{\partial t^2}\Big|_{t=0} = \frac{\partial}{\partial t}\left(n(q+pe^t)^{n-1} pe^t\right)\Big|_{t=0}$$

$$= npq + n^2p^2 \qquad (q = 1-p)$$

$$VX = EX^2 - (EX)^2 = npq$$

(IV) Negative Binomial

Repeat independent Bernoulli trials until r successes

$$\Omega = \{ \underbrace{S \ldots S}_{r\,S}, \underbrace{\ldots \ldots S}_{1F,\, rS}, \underbrace{\ldots \ldots \ldots S}_{2F,\, rS}$$

X : number of failures preceeding the r^{th} success

$$P(X=x) = \begin{cases} \binom{x+r-1}{x} q^x p^{r-1} \cdot p, & x = 0, 1, 2, \ldots \\ 0, & o/w \end{cases}$$

$q = 1-p$

$X \sim NB(r, p)$

Note : $\binom{x+r-1}{x} = (-1)^x \binom{-r}{x}$ → hence the name "negative binomial".

Note: Negative Binomial series

$$(x+a)^{-n} = \sum_{k=0}^{\infty} \binom{-n}{k} x^k a^{-n-k}$$

Note that

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$$

$$= \frac{(n+k-1)(n+k-2)\cdots n}{k!}$$

$$= (-1)^k \frac{\left((-n)(-n-1)\cdots\cdots(-n-k+1)\right)}{k!}$$

$$= (-1)^k \binom{-n}{k}$$

i.e. $$(x+a)^{-n} = \sum_{k=0}^{\infty} \binom{-n}{k} x^k a^{-n-k}$$

$$= \sum_{k=0}^{\infty} (-1)^k \binom{n+k-1}{k} x^k a^{-n-k}$$

$$\sum_0^\infty P(X=x) = \sum_0^\infty \binom{x+r-1}{x} q^x p^r$$
$$= p^r \sum_0^\infty \binom{-r}{x} (-q)^x$$
$$= p^r (1-q)^{-r} = 1$$

m.g.f. : $M_X(t) = E(e^{tX}) = \sum_0^\infty e^{tx} \binom{x+r-1}{x} q^x p^r$
$$= p^r \sum_0^\infty \binom{-r}{x} (-qe^t)^x$$
$$= p^r (1-qe^t)^{-r}$$

$$EX = \left.\frac{\partial M_X(t)}{\partial t}\right|_{t=0} = \left. p^r r (1-qe^t)^{-r-1} qe^t \right|_{t=0} = \frac{rq}{p}$$

$$VX = \frac{rq}{p^2}$$

<u>Note</u>: Sp. case of NB(r, p) - r = 1, Geometric distn

p.m.f. $P(X=x) = q^x p$; $x = 0, 1, 2, \ldots$

<u>Note</u>: $P(X \geq m) = \sum_m^\infty p q^x = p q^m \frac{1}{1-q} = q^m$

For $m, n \geq 0$
$$P(X \geq m+n \mid X \geq m) = \frac{P(X \geq m+n \text{ and } X \geq m)}{P(X \geq m)}$$
$$= \frac{P(X \geq m+n)}{P(X \geq m)} = \frac{q^{m+n}}{q^m} = q^n = P(X \geq n)$$

also $P(X = m+n \mid X \geq m) = \frac{q^{m+n} p}{q^m} = q^n p = P(X=n)$

The above is called "lack of memory property"

(v) Discrete uniform

uniform distⁿ on n pts

$$P(X = x_i) = \frac{1}{n} \; ; \quad i = 1, 2, \ldots n$$

$$M_X(t) = \frac{1}{n} \sum_{1}^{n} e^{t x_i}$$

$$E X = \frac{1}{n} \sum_{1}^{n} x_i$$

$$V X = \frac{1}{n} \sum_{1}^{n} (x_i - \bar{x})^2$$

vi Hypergeometric

Urn containing M white & $N-M$ balls of some other color

n balls drawn at a time or one after another without replacement

X : number of white balls in the sample

minimum value of X : $\max(0, n-(N-M))$

maximum value of X : $\min(n, M)$

$$P(X = x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & x = \max(0, n-(N-M)), \ldots, \min(n, M) \\ 0, & \text{o/w.} \end{cases}$$

vi Hypergeometric

urn containing M white & $N-M$ balls of some other color

n balls drawn at a time or one after another without replacement

X : number of white balls in the sample

minimum value of X : $\max(0, n-(N-M))$

maximum value of X : $\min(n, M)$

$$P(X=x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & x = \max(0, n-(N-M)), \ldots, \min(n,M) \\ 0, & \text{o/w.} \end{cases}$$

Note: Same setup with replacement sampling — $\text{Bin}\left(n, \frac{M}{N}\right)$

m.g.f. $$E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

(for $n > M$, $n > N-M$

$$E(X) = n\frac{M}{N}$$

$$V(X) = n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

use $$\sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k} = \binom{a+b}{n}$$ to derive $E(X)$, $E(X(X-1))$ to get $V(X)$

VII Poisson

$$X \sim P(\lambda) \qquad \lambda > 0$$

$$P(X=x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} ; & x = 0, 1, 2, \ldots \\ 0 ; & o/w \end{cases}$$

m.g.f : $$E(e^{tX}) = \sum_{x=0}^{\infty} \frac{e^{tx}\lambda^x}{x!} e^{-\lambda}$$

$$= e^{\lambda e^t} e^{-\lambda} = e^{-\lambda(1-e^t)}$$

$$E(X) = V(X) = \lambda$$

Note: Poisson distⁿ is applicable to model count of events, change of states, failures etc.

Note: Poisson approximation to Binomial

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

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \frac{n(n-1)\cdots(n-x+1)}{x!} p^x (1-p)^{n-x}$$

$$= \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{(n-x+1)}{n} \frac{(np)^x}{x!} (1-p)^n (1-p)^{-x}$$

As $n \to \infty$, $p \to 0 \ni np = \lambda$ (fixed)

$$P(X=x) \longrightarrow \frac{\lambda^x}{x!} e^{-\lambda}$$

Some standard continuous distributions

(I) Uniform distⁿ

$X \sim U[a, b]$

p.d.f. $f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & o/w. \end{cases}$

d.f. $F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$

m.g.f. $M_X(t) = \frac{1}{b-a} \int_a^b e^{tx}\, dx = \begin{cases} \frac{e^{tb} - e^{ta}}{t(b-a)} & t \neq 0 \\ 1 & t = 0 \end{cases}$

II Normal distⁿ

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

$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \; ; \; x \in Q$

$\mu \in Q, \; \sigma > 0$

Distⁿ is symmetric around μ $\forall (\mu, \sigma)$

$P(X \le \mu - c) = P(X \ge \mu + c)$

$f(\mu - c) = f(\mu + c) \quad \forall c$

m.g.f: $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx$

$$\frac{x-\mu}{\sigma} = z; \quad dx = \sigma\, dz$$

$$M_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\mu+\sigma z)}\, e^{-z^2/2}\, dz$$

$$= \frac{e^{t\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2 - 2t\sigma z + t^2\sigma^2 - t^2\sigma^2)}\, dz$$

$$= \frac{e^{t\mu}}{\sqrt{2\pi}}\, e^{t^2\sigma^2/2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t\sigma)^2}\, dz$$

$$= e^{t\mu + t^2\sigma^2/2} \qquad t \in (-\infty, \infty)$$

$$E(X) = \frac{\partial}{\partial t} M_X(t)\Big|_{t=0} = e^{t\mu + t^2\sigma^2/2}(\mu + t\sigma^2)\Big|_{t=0}$$

$$= \mu$$

$$E X^2 = \frac{\partial^2}{\partial t^2} M_X(t)\Big|_{t=0} = e^{t\mu + t^2\sigma^2/2}(\sigma^2) + (\mu + t\sigma^2)\, e^{t\mu + t^2\sigma^2/2}(\mu + t\sigma^2)\Big|_{t=0}$$

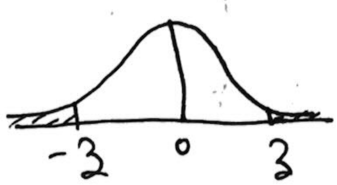
$$= \sigma^2 + \mu^2$$

$$E X^2 = \mu^2 + \sigma^2$$

$$V(X) = \sigma^2$$

<u>Standard normal distn</u> $N(0,1)$

$$Z \sim N(0,1) \quad \phi(z) = f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad z \in (-\infty, \infty)$$



$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$ (values are tabulated)

$$\Phi(-z) = P(Z \le -z) = P(Z \ge z)$$

$$= 1 - \Phi(z)$$

$$\Phi(-z) + \Phi(z) = 1 \quad ; \quad \Phi(0) = \frac{1}{2}$$

Note: If $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$

Let $Z = \frac{X-\mu}{\sigma}$

d.f. of Z: $P(Z \le z) = P\left(\frac{X-\mu}{\sigma} \le z\right)$

$$= P(X \le \mu + \sigma z)$$

$$= \int_{-\infty}^{\mu+\sigma z} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$\frac{x-\mu}{\sigma} = y$; $\quad = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-y^2/2} dy = \Phi(z)$

i.e. $Z \sim N(0,1)$

Alt: use m.g.f.

<u>Note</u>: If $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$

Let $Z = \frac{X-\mu}{\sigma}$

d.f. of Z: $P(Z \le z) = P\left(\frac{X-\mu}{\sigma} \le z\right)$

$= P(X \le \mu + \sigma z)$

$$= \int_{-\infty}^{\mu+\sigma z} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$\frac{x-\mu}{\sigma} = y$; $\quad = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-y^2/2}\, dy = \Phi(z)$

i.e. $Z \sim N(0,1)$

<u>Alt</u>: use m.g.f.

<u>Note</u>: Normal approximation to Binomial

$X \sim B(n, p)$ $\qquad EX = np$; $V(X) = np(1-p)$

To compute $P(\alpha \le X \le \beta) = \sum_{x=\alpha}^{\beta} \binom{n}{x} p^x (1-p)^{n-x}$

$\alpha, \beta \to$ points in support of X

for large n $\approx \Phi\left(\frac{\beta + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - \frac{1}{2} - np}{\sqrt{npq}}\right)$

Note that the above approximation is based on

- For large n, $\frac{X - np}{\sqrt{npq}}$ has a limiting distⁿ $N(0,1)$
- $P(\alpha \le X \le \beta) = P(\alpha - \frac{1}{2} < X < \beta + \frac{1}{2})$

Continuity correction

(iii) Gamma dist^n

$X \sim \text{Gamma}(\alpha, \beta) \qquad \alpha > 0,\ \beta > 0$

p.d.f. $f(x) = \begin{cases} \dfrac{1}{\Gamma\alpha\, \beta^\alpha} \cdot e^{-x/\beta}\, x^{\alpha-1}, & x > 0 \\ 0, & \text{o/w.} \end{cases}$

m.g.f.

$$M_X(t) = E(e^{tX}) = \frac{1}{\beta^\alpha \Gamma\alpha} \int_0^\infty e^{tx} e^{-x/\beta} x^{\alpha-1}\, dx$$

$$= \frac{1}{\beta^\alpha \Gamma\alpha} \int_0^\infty e^{-x\left(\frac{1}{\beta} - t\right)} x^{\alpha-1}\, dx$$

$$= \frac{\Gamma\alpha}{\left(\frac{1}{\beta} - t\right)^\alpha} \cdot \frac{1}{\beta^\alpha \Gamma\alpha} = \frac{1}{(1-\beta t)^\alpha}$$

Note that $M_X(t)$ exists if $t < \frac{1}{\beta}$

$$E X^k = \mu_k' = \frac{1}{\Gamma\alpha\, \beta^\alpha} \int_0^\infty x^k e^{-x/\beta} x^{\alpha-1}\, dx$$

$$= \frac{\Gamma(\alpha+k)}{\Gamma\alpha} \cdot \beta^k$$

$$\left.\begin{aligned} E(X) &= \frac{\Gamma(\alpha+1)}{\Gamma\alpha}\beta = \alpha\beta \\ E X^2 &= \frac{\Gamma(\alpha+2)}{\Gamma\alpha}\beta^2 = \alpha(\alpha+1)\beta^2 \end{aligned}\right\} \Rightarrow V(X) = \alpha\beta^2$$

Sp. case: $\alpha = 1 \rightarrow$ exponential dist^n with scale parameter β

$X \sim \exp(\beta)$ or $\exp(0, \beta)$

location, scale parameter

$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0 \\ 0, & \text{o/w.} \end{cases}$

m.g.f. $M_X(t) = \dfrac{1}{1-\beta t}$ exists for $t < \dfrac{1}{\beta}$

$$E(X) = \beta, \quad V(X) = \beta^2$$

Note: Lack of memory property of $\exp(\beta)$

Note that

$$\underset{x>0}{P(X \geq x)} = \frac{1}{\beta} \int_x^\infty e^{-t/\beta}\, dt = \frac{1}{\beta} \cdot \left. \frac{e^{-t/\beta}}{-\frac{1}{\beta}} \right|_x^\infty$$

$$= e^{-x/\beta}$$

$\Rightarrow$

$$\underset{r,s>0}{P(X \geq r+s \mid X \geq r)} = \frac{P(X \geq r+s,\ X \geq r)}{P(X \geq r)}$$

$$= \frac{P(X \geq r+s)}{P(X \geq r)} = \frac{e^{-(r+s)/\beta}}{e^{-r/\beta}} = e^{-s/\beta}$$

$$= P(X \geq s)$$

Note: Alternate def^n of Gamma(α, β)

p.d.f. $f(x) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma\alpha} e^{-\beta x} x^{\alpha-1}, & x > 0 \\ 0, & o/w \end{cases}$

Note: Chi-square distⁿ as a special case of Gamma distⁿ

Consider $G(\alpha, \beta)$ $\quad \alpha = p/2 \; , \; \beta = 2$

i.e. $G(p/2, 2)$ $\quad p = 1, 2, \ldots$

$$f(x) = \begin{cases} \dfrac{1}{2^{p/2}\, \Gamma p/2}\, e^{-x/2}\, x^{p/2 - 1} \; , & x > 0 \\ 0 \; , & o/w \end{cases}$$

$E(X) = p \; ; \; V(X) = 2p$

$\widehat{\underline{IV}}$ Exponential distⁿ

As the sp case of Gamma (α, β) with $\alpha = 1$, we have $\exp(\beta)$

$\uparrow$ scale

In general, 2-parameter exp distⁿ.

$X \sim exp(\alpha, \beta)$

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x-\alpha}{\beta}}, & x > \alpha \\ 0, & o/w \end{cases}$$

$\underline{\hat{V}}$ Beta distⁿ

$X \sim Beta(\alpha, \beta)$

p.d.f. $$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 < x < 1 \\ 0, & o/w \end{cases}$$

$\alpha, \beta > 0$; $B(\alpha, \beta) = \frac{\Gamma\alpha \, \Gamma\beta}{\Gamma\alpha+\beta}$.

i.e. $B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$

Sp. case : $\alpha = 1, \beta = 1$

i.e. $X \sim B(1,1) \leftarrow U[0,1]$, a cont uniform distⁿ

Moments

$$E(X^K) = \frac{1}{B(\alpha,\beta)} \int_0^1 x^K \, x^{\alpha-1} (1-x)^{\beta-1} dx$$

$$= \frac{B(\alpha+K, \beta)}{B(\alpha,\beta)} = \frac{\Gamma\alpha+K \, \Gamma\beta / \Gamma\alpha+\beta+K}{\Gamma\alpha \, \Gamma\beta / \Gamma\alpha+\beta}$$

$$= \frac{\Gamma\alpha+K \, \Gamma\alpha+\beta}{\Gamma\alpha+\beta+K \, \Gamma\alpha}$$

$E(X) = \frac{\alpha}{\alpha+\beta}$ $\qquad EX^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$

$V(X) = EX^2 - (E(X))^2 = \text{--} \; \text{--}$

m.g.f.

$$M_X(t) = \sum_{j=0}^{\infty} \frac{t^j}{j!} \frac{\Gamma(\alpha+j)\,\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+j)\,\Gamma(\alpha)}$$

$\hat{\text{VI}}$ Double exponential or a Laplace distⁿ

$$f(x) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}, \quad x \in \mathbb{R}$$

$\mu \in \mathbb{R}, \sigma > 0$

$X \sim DE(\mu, \sigma)$ m.g.f : $\dfrac{e^{t\mu}}{1-(\sigma t)^2}$; $|t| < \dfrac{1}{\sigma}$

$EX = \mu$; $V(X) = 2\sigma^2$

$\hat{\text{VII}}$ Cauchy distⁿ

$X \sim C(a, b)$

$$f(x) = \frac{1}{\pi b} \frac{1}{1+\left(\frac{x-a}{b}\right)^2}, \quad x \in \mathbb{R}$$

$a \in \mathbb{R}, b > 0$

None of the moments exist

m.g.f does not exist

Functions of random variables

Discrete r.v. Case

X : discrete r.v. $\mathscr{X}$: range space of X

$Y = g(X)$; $g(\cdot)$ a real valued f^n

$\mathscr{Y}$: range space of Y

$\mathscr{Y} = \{g(x) : x \in \mathscr{X}\}$

Problem is to find p.m.f. of Y given the p.m.f. of X

Note that $g(x) : \mathscr{X} \to \mathscr{Y}$

Inverse mapping : g^{-1} $\ni$

$$g^{-1}(y) = \{x \in \mathscr{X} : g(x) = y\}$$

In general, for $A \subset \mathscr{Y}$

$$g^{-1}(A) = \{x \in \mathscr{X} : g(x) \in A\}$$

* If there is only one x for which $g(x) = y$, then $g^{-1}(y) = x$

p.m.f. of Y :

for $y \in \mathscr{Y}$; $P(Y = y) = \sum\limits_{x \in g^{-1}(y)} P(X = x) = \sum\limits_{x : g(x) = y} P(X = x)$

for $y \notin \mathscr{Y}$: $P(Y = y) = 0$

Example: $X \sim Bin(n, p)$

$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots n$

$$\mathscr{X} = \{0, 1, \ldots n\}$$

$X \to Y = n - X$

$\Rightarrow \mathscr{Y} = \{0, 1, \ldots n\}$

For any $y \in \mathscr{Y}$, $y = g(x) = n - x$ iff $x = n - y$ and $g^{-1}(y)$ is a single pt p.m.f of Y :

$$f_Y(y) = P(Y=y) = \sum_{x:\, g(x)=y} f_X(x) = f_X(n-y)$$

$$= \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)}$$

$$= \binom{n}{y} (1-p)^{y} p^{n-y} \quad ; \quad y = 0, 1, \ldots n$$

$$= 0 \qquad o/w$$

$$\Rightarrow Y \sim B(n, 1-p)$$

Example: $\mathcal{X} = \{-1, 0, 1, 2\}$

p.m.f. of X : $P(X=-1) = 0.2$; $P(X=0) = 0.3$; $P(X=1) = .4$

$P(X=2) = 0.1$

$X \to Y = X^2$; $\mathcal{Y} = \{0, 1, 4\}$

p.m.f. of Y: $P(Y=0) = P(X=0) = 0.3$

$P(Y=1) = P(X=1) + P(X=-1) = 0.6$

$P(Y=4) = P(X=2) = 0.1$

Example: $X \sim P(\lambda)$; $\mathcal{X} = \{0, 1, 2, \ldots\}$

$X \to Y = X^2 + 3$; $\mathcal{Y} = \{3, 4, 7, \ldots\}$

p.m.f. of Y : $P(Y=y) = P(X^2 + 3 = y)$

$$= P(X = \sqrt{y-3})$$

$$= \begin{cases} \dfrac{e^{-\lambda} \lambda^{\sqrt{y-3}}}{(\sqrt{y-3})!}, & y = 3, 4, 7, \ldots \\ 0, & o/w \end{cases}$$

Transformation : Continuous r.v.

Distribution f^n method:

$$X \to Y = g(X) \qquad \mathcal{X} = \{x : f_X(x) > 0\}$$

d.f. of Y: $\qquad \mathcal{Y} = \{y : g(x) = y\}_{x \in \mathcal{X}}$

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

$$= P(\{x : g(x) \leq y\})$$

$$= \int_{x : g(x) \leq y} f_X(x)\, dx$$

Note: The d.f. approach is straightforward if $g(.)$ is strictly monotone (increasing or decreasing). In such a case

$$y = g(x) \Rightarrow x = g^{-1}(y)$$

If $g(x)$ is increasing, then

$$F_Y(y) = \int_{x : x \leq g^{-1}(y)} f_X(x)\, dx = \int_{-\infty}^{g^{-1}(y)} f_X(x)\, dx = F_X(g^{-1}(y))$$

If $g(x)$ is decreasing, then

$$F_Y(y) = \int_{x : x \geq g^{-1}(y)} f_X(x)\, dx = \int_{g^{-1}(y)}^{\infty} f_X(x)\, dx = 1 - F_X(g^{-1}(y))$$

Example:

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{o/w} \end{cases} \qquad X \sim U(0,1)$$

$$F(x) = \begin{cases} x, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

$Y = -\ln X \quad \downarrow$ in x

$y = -\ln x \Rightarrow x = e^{-y}$ i.e. $g^{-1}(y) = e^{-y}$

$\mathcal{Y} = (0, \infty)$ from $\mathcal{X} = (0, 1)$

Thus for $y > 0$

$$F_Y(y) = 1 - F_X(g^{-1}(y))$$
$$= 1 - F_X(e^{-y})$$
$$= 1 - e^{-y}$$

& $F_Y(y) = 0$ for $y \leq 0$

p.d.f. of Y: $f_Y(y) = \begin{cases} e^{-y}, & y > 0 \\ 0, & o/w \end{cases}$

i.e. $Y \sim \exp(1)$

__Note__: If $g(.)$ is not monotone, the above cannot be applied.

e.g. $Y = X^2$

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y})$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

p.d.f. of Y:

$$f_Y(y) = \frac{\partial}{\partial y} F_Y(y)$$
$$= \frac{\partial}{\partial y}\left(F_X(\sqrt{y}) - F_X(-\sqrt{y})\right)$$
$$= f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}}$$

Note: If $g(\cdot)$ is not monotone, the above cannot be applied.

e.g. $Y = X^2$ $\quad \mathcal{X} = (-\infty, \infty)$; $\mathcal{Y} = (0, \infty)$

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y})$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

p.d.f. of Y:

$$f_Y(y) = \frac{\partial}{\partial y} F_Y(y)$$

$$= \frac{\partial}{\partial y}\left(F_X(\sqrt{y}) - F_X(-\sqrt{y})\right)$$

$$= f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}}$$

Example: $X \sim N(0,1)$

$Y = X^2$ $\quad \mathcal{X} = (-\infty, \infty)$; $\mathcal{Y} = (0, \infty)$

$$F_Y(y) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \qquad y > 0$$

$$= \Phi(\sqrt{y}) - (1 - \Phi(\sqrt{y}))$$

$$= 2\,\Phi(\sqrt{y}) - 1$$

p.d.f. of Y: $f_Y(y) = \cancel{2}\,\phi(\sqrt{y}) \dfrac{1}{\cancel{2}\sqrt{y}} = \dfrac{\phi(\sqrt{y})}{\sqrt{y}}$

i.e. $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y/2} y^{-1/2}$ $\quad y > 0$

Realise that the above is density of a χ^2 r.v. with $n = 1$

<u>Example</u>: $X \sim U(-1, 2)$

$$f_X(x) = \begin{cases} 1/3, & -1 < x < 2 \\ 0, & o/w \end{cases}$$

$\mathcal{X} = (-1, 2)$

$Y = X^2 \quad \mathcal{Y} = (0, 4)$

$F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y})$

If $0 \leq y < 1$, then $F_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{3} = \frac{2\sqrt{y}}{3}$

if $1 \leq y < 4$, then

$$F_Y(y) = \int_{-1}^{\sqrt{y}} \frac{1}{3}\,dx = \frac{1}{3}(1 + \sqrt{y})$$

If $y \geq 4$ $\quad F_Y(y) = 1$

& if $y < 0$ $\quad F_Y(y) = 0$

i.e. $$F_Y(y) = \begin{cases} 0, & y < 0 \\ \frac{2\sqrt{y}}{3}, & 0 \leq y < 1 \\ \frac{1}{3}(1 + \sqrt{y}), & 1 \leq y < 4 \\ 1, & y \geq 4 \end{cases}$$

p.d.f. of Y:

$$f_Y(y) = \begin{cases} \frac{1}{3\sqrt{y}}, & 0 < y < 1 \\ \frac{1}{6\sqrt{y}}, & 1 < y < 4 \\ 0, & o/w \end{cases}$$

Jacobian based method

Note that from d.f. based method we have for strictly monotone $g(\cdot)$ $\left(y = g(x) \,;\ x = g^{-1}(y) \right)$

$$F_Y(y) = \begin{cases} F_X(g^{-1}(y)), & \text{if } g \text{ is increasing} \\ 1 - F_X(g^{-1}(y)), & \text{if } g \text{ is decreasing} \end{cases}$$

$$\Rightarrow \quad f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \dfrac{\partial}{\partial y} g^{-1}(y), & \text{if } g \text{ is increasing} \\ - f_X(g^{-1}(y)) \dfrac{\partial}{\partial y} g^{-1}(y), & \text{if } g \text{ is decreasing.} \end{cases}$$

i.e. $f_Y(y) = f_X(g^{-1}(y)) \left| \dfrac{\partial}{\partial y} g^{-1}(y) \right| \quad \forall\, y \in \mathcal{Y}$

$\qquad\quad = 0 \qquad\qquad$ o/w

$\dfrac{\partial}{\partial y} g^{-1}(y)$ is called the Jacobian of transformation.

<u>Example</u>: $\quad X \sim U(0,1)$

$\qquad Y = -2 \ln X$

$y = g(x) = -2 \ln x$

$\qquad x = e^{-y/2} = g^{-1}(y)$

$J = \dfrac{dx}{dy} = e^{-y/2}\left(-\frac{1}{2}\right) = \dfrac{d\, g^{-1}(y)}{dy} \quad \leftarrow$ Jacobian

$f_Y(y) = f_X(g^{-1}(y)) \, |J| \qquad\qquad 0 < y < \infty$

$\qquad\quad = 0, \qquad\qquad$ o/w

i.e. $f_Y(y) = \begin{cases} \frac{1}{2} e^{-y/2}, & y > 0 \\ 0, & o/w \end{cases}$

Note that $Y \sim \chi^2_2$

Example : $X \sim \text{Gamma}(n, \beta)$

$$f_X(x) = \frac{1}{\Gamma n \; \beta^n} e^{-x/\beta} x^{n-1}, \quad 0 < x < \infty$$

$\beta > 0$, n is a positive integer

$\mathcal{X} = (0, \infty)$

$X \to Y = \frac{1}{X}$ $\qquad \mathcal{Y} = (0, \infty)$

$y = g(x) = \frac{1}{x}$; $\quad x = \frac{1}{y} = g^{-1}(y)$

$J = \frac{\partial x}{\partial y} = \frac{\partial g^{-1}(y)}{\partial y} = -\frac{1}{y^2}$ ← Jacobian

$$\Rightarrow f_Y(y) = f_X(g^{-1}(y)) \, |J|$$

$$= \left(\frac{1}{\Gamma n \; \beta^n} e^{-\frac{1}{y\beta}} \left(\frac{1}{y}\right)^{n-1} \right) \left(\frac{1}{y^2} \right)$$

$$= \frac{1}{\Gamma n \; \beta^n} \left(\frac{1}{y}\right)^{n+1} e^{-\frac{1}{y\beta}} \qquad y > 0$$

Note: The above works if $g(.)$ is strictly monotone in $\mathcal{X}$.

The following extension to non-monotone setup is useful:

Suppose $\exists$ a partition $A_0, A_1, A_2, \ldots A_K$ of $\mathcal{X}$ such that

$P(X \in A_0) = 0$

$f_X(x)$ is continuous on each A_i

Suppose further that there exist functions $g_1(x), \ldots g_K(x)$ defined on $A_1, \ldots, A_K$ satisfying

(a) $g(x) = g_i(x)$ for $x \in A_i$

(b) $g_i(x)$ is strictly monotone on A_i

(c) $\mathcal{Y} = \{y : y = g_i(x) \text{ for } x \in A_i\}$ is same for each $i = 1(1)K$

(d) $g_i^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$ $\forall i$

Then

$$f_Y(y) = \begin{cases} \displaystyle\sum_{i=1}^{K} f_X\left(g_i^{-1}(y)\right) \left|\frac{\partial g_i^{-1}(y)}{\partial y}\right|, & y \in \mathcal{Y} \\ \vdots & \\ 0, & \text{o/w} \end{cases}$$

Example: $X \sim N(0,1)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty$$

$X \to Y = X^2$

Note that $g(x) = x^2$ is monotone on $(-\infty, 0)$ and on $(0, \infty)$

$\mathcal{Y} = (0, \infty)$

Apply the above result with

$A_0 = \{0\}$; $A_1 = (-\infty, 0)$; $A_2 = (0, \infty)$

For the region $(-\infty, 0) = A_1$

$$g_1(x) = x^2 = y \;;\; x = -\sqrt{y} = g_1^{-1}(y)$$

& for $A_2 = (0, \infty)$; $g_2(x) = x^2 = y$; $x = \sqrt{y}$

$$g_2^{-1}(y) = \sqrt{y}$$

$\therefore$ p.d.f. of $Y = X^2$

$$f_Y(y) = \overset{(-\infty,0)}{\overset{\downarrow}{f_X(g_1^{-1}(y))}} \left| \frac{\partial g_1^{-1}(y)}{\partial y} \right| + \overset{(0,\infty)}{\overset{\downarrow}{f_X(g_2^{-1}(y))}} \left| \frac{\partial g_2^{-1}(y)}{\partial y} \right|$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-\sqrt{y})^2} \left| -\frac{1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2} \left| \frac{1}{2\sqrt{y}} \right|$$

i.e. $f_Y(y) = \dfrac{1}{\sqrt{2\pi}} e^{-y/2} \, y^{-1/2}$

i.e. $Y \sim \chi^2_1$

<u>Note</u>: For the example of $X \sim U(-1, 2)$; $X \to Y = X^2$

for $y \in (0,1)$. $f_Y(y) = \overset{(-1,0)\in\mathcal{X}}{\overset{\downarrow}{f_X(g_1^{-1}(y))}} \left| \dfrac{\partial g_1^{-1}(y)}{\partial y} \right| + \overset{(0,1)\in\mathcal{X}}{\overset{\downarrow}{f_X(g_2^{-1}(y))}} \left| \dfrac{\partial g_2^{-1}(y)}{\partial y} \right|$

$g(x) = x^2$ is monotone strictly in $(-1, 0)$ and $(0, 1)$ $= \dfrac{1}{3}\dfrac{1}{2\sqrt{y}} + \dfrac{1}{3}\dfrac{1}{2\sqrt{y}}$

$(-1, 0) \longrightarrow (0, 1)$

& $(0, 1) \longrightarrow (0, 1)$ $\qquad = \dfrac{1}{3\sqrt{y}}$

for $y \in (1, 4)$

$g(x) = x^2$ is strictly monotone

$(1, 2) \to (1, 4)$ $\qquad f_Y(y) = f_X(g^{-1}(y)) \left| \dfrac{\partial g^{-1}(y)}{\partial y} \right|$

$$= \frac{1}{3}\frac{1}{2\sqrt{y}} = \frac{1}{6\sqrt{y}}$$

Multivariate random vectors

$\underset{\sim}{X} = (X_1, \ldots, X_p)'$ $\quad$ X_i is a r.v. on $(\Omega, \mathcal{F}, P)$ say

$\forall \omega \in \Omega$; assign $\underset{\sim}{X}(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_p(\omega) \end{pmatrix}$

Distⁿ fⁿ or joint distⁿ fⁿ

$\underset{\sim}{X}_{p \times 1}$ is a random vector

$$F_{\underset{\sim}{X}}(x_1, \ldots, x_p) = P(\omega : X_1(\omega) \le x_1, X_2(\omega) \le x_2, \ldots X_p(\omega) \le x_p)$$

$$= P(X_1 \le x_1, X_2 \le x_2 \ldots, X_p \le x_p)$$

Note that $F_{\underset{\sim}{X}}(\cdot)$ as defined above satisfies

(i) $\lim\limits_{\min_i x_i \to \infty} F_X(\underset{\sim}{x}) = 1$

(ii) $\lim\limits_{\text{any } x_i \to -\infty} F_{\underset{\sim}{X}}(\underset{\sim}{x}) = 0$ for $i = 1, \ldots, p$

(iii) $F_{\underset{\sim}{X}}(\underset{\sim}{x})$ is non decreasing in each argument

(iv) $F_X(\underset{\sim}{x})$ is right continuous in each argument

Remark: $\quad p = 2$

$$\lim_{x_2 \to \infty} F_{\underset{\sim}{X}}(x_1, x_2) = F_{\underset{\sim}{X}}(x_1, \infty) = F_{X_1}(x_1)$$

marginal distⁿ fⁿ of X_1

In general for any $K = 1, \ldots, p$

$$\lim_{x_K \to \infty} F_{\underset{\sim}{X}}(x_1, \ldots \underset{K\text{th}}{\uparrow} \ldots x_p) = F_{X_1, \ldots X_{K-1}, X_{K+1} \ldots, X_p}(x_1, \ldots x_{K-1}, x_{K+1}, \ldots x_p)$$

marginal joint d.f. of $(X_1, \ldots X_{K-1}, X_{K+1}, \ldots X_p)$

Remark :

$P(\underset{\sim}{X} \in B_p)$ can be expressed through joint d.f.

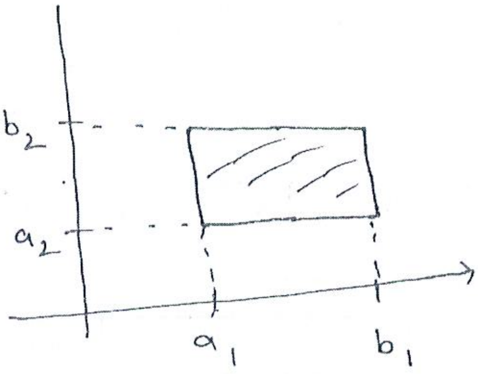
p-dimensional semi closed rectangle of the form

$(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_k, b_k]$

$a_i < b_i$ for $i = 1, \ldots, k$

Consider $p = 2$, to have a feel

$a_1 < b_1$; $a_2 < b_2$



$$P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2)$$

$$= P(X_1 \leq b_1, X_2 \leq b_2)$$

$$- P(X_1 \leq a_1, X_2 \leq b_2)$$

$$- P(X_1 \leq b_1, X_2 \leq a_2)$$

$$+ P(X_1 \leq a_1, X_2 \leq a_2)$$

$$= F_{X_1,X_2}(b_1, b_2) - F_{X_1,X_2}(a_1, b_2) - F_{X_1,X_2}(b_1, a_2) + F_{X_1,X_2}(a_1, a_2)$$

Remark : The four conditions that $F_{\underset{\sim}{X}}(\cdot)$ satisfies (i)-(iv) are not n.s.c. for a f^n to be d.f. of random vector.

We additionally need condition (for $p=2$) that

$$P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) \geq 0 \quad \forall \; a_1 < b_1, \; a_2 < b_2$$

Discrete random vector

Def^n: A random vector $\underset{\sim}{X} = (X_1, \ldots, X_p)'$ is said to be discrete

If there exist a countable set $E \subset \mathbb{R}^p$ $\ni$
(finite or countably infinite)

$$P(\underset{\sim}{X} \in E) = 1$$

p.m.f. or joint p.m.f of $(X_1, \ldots X_p)$

Let $E = \{ \underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots \}$; $e_i \in \mathbb{R}^p$

$$p_X(\underline{x}) = P(\underline{X} = \underset{\sim}{x}) = \begin{cases} P(X = e_i), & \text{If } \underset{\sim}{x} = \underset{\sim}{e}_i, i = 1, 2, \ldots \\ 0, & \text{If } \underset{\sim}{x} \notin E \end{cases}$$

i.e. $p_X(x) = P(X_1 = x_1, \ldots, X_p = x_p)$.

Consider a bivariate setup $p = 2$ case

(i) marginal p.m.f. of X_i can be obtained by summing over possible values of the other variable

e.g $P(X_1 = x_i) = \sum\limits_y P(X_1 = x_i, X_2 = y)$; $i = 1, 2, \ldots$

Sly $P(X_2 = y_j) = \sum\limits_x P(X_1 = x, X_2 = y_j)$; $j = 1, 2, \ldots$

(ii) Conditional dist^n of X_1 given X_2 or X_2 given X_1

conditional p.m.f of X_2 given X_1

$$p_{X_2 | X_1 = x_i}(y | x_i) = \frac{P(X_2 = y, X_1 = x_i)}{P(X_1 = x_i)}$$

For each level of fixed x_i, we get a conditional dist^n

Sly $$p_{X_1 | X_2 = y_j}(x | y_j) = \frac{P(X_1 = x, X_2 = y_j)}{P(X_2 = y_j)}$$

For a p-dimensional random vector $\underset{\sim}{X} = (X_1, \ldots X_p)$, we can obtain p ~~...~~ 1-variate marginals for $X_1, X_2, \ldots X_p$

$\binom{p}{2}$ 2-variate joint marginals for (X_i, X_j)

and so on

One can obtain conditional p.m.f. of $(X_{i_1}, X_{i_2}, \ldots, X_{i_r})$ given the remaining variables ~~...~~ or any ~~...~~ subset of the remaining variables.

<u>Independence</u> : Discrete random variables ~~...~~ $X_1, \ldots X_p$ are independent iff the joint p.m.f can be expressed as

$$P(X_1 = x_1, \ldots, X_p = x_p) = \prod_{i=1}^{p} P(X_i = x_i)$$

$$\forall (x_1, \ldots x_p) \in E$$

Note that in such a case conditional p.m.f s will be identical to unconditional p.m.f.s.

<u>Example of a multivariate discrete distn</u>

Consider a random experiment with 3 mutually exclusive and exhaustive outcomes, A_1, A_2 & A_3 with probabilities $\theta_1, \theta_2, \theta_3$, respectively. Repeat the trials n times

Define

X_1 : number of times A_1 occurs out of n trials

X_2 : - - - - - - A_2 occurs out of n trials

X_3 : - - - - - - A_3 occurs out of n trials let (x_1, x_2, x_3) denote the observed count in n trials

$$x_i \geq 0, \quad x_i \leq n \quad \forall i = 1, 2, 3 \ \& \ \sum_1^3 x_i = n$$

$$E = \left\{ (x_1, x_2, x_3) : 0 \leq x_i \leq n, \ \sum_{i=1}^3 x_i = n \right\}$$ - finite number of points

jt p.m.f.

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1! \, x_2! \, x_3!} \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3}; \quad 0 \leq x_i \leq n, \ \sum x_i = n$$

Note that $x_3 = n - x_1 - x_2$ and

$$\theta_3 = 1 - \theta_1 - \theta_2$$

$$P(X_1 = x_1, X_2 = x_2) = \frac{n!}{x_1! \, x_2! \, (n - x_1 - x_2)!} \theta_1^{x_1} \theta_2^{x_2} (1 - \theta_1 - \theta_2)^{n - x_1 - x_2}$$

$$x_1, x_2 \geq 0$$
$$x_1 + x_2 \leq n$$

$$= 0, \quad \text{o/w}$$

(X_1, X_2) is said to follow a trinomial distn (n, θ_1, θ_2)

Marginal dist^{n}s:

Marginal p.m.f. of X_1:

$$P(X_1 = x_1) = \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1! (n - x_1)!} \theta_1^{x_1} \frac{(n - x_1)!}{x_2! (n - x_1 - x_2)!} \theta_2^{x_2} (1 - \theta_1 - \theta_2)^{n - x_1 - x_2}$$

$$= \binom{n}{x_1} \theta_1^{x_1} (1 - \theta_1 - \theta_2 + \theta_2)^{n - x_1}$$

$$= \binom{n}{x_1} \theta_1^{x_1} (1 - \theta_1)^{n - x_1}.$$

i.e. $X_1 \sim Bin(n, \theta_1)$

Sly $X_2 \sim Bin(n, \theta_2)$

Example of a multivariate discrete distⁿ

Consider a random experiment with 3 mutually exclusive and exhaustive outcomes, A_1, A_2 & A_3 with probabilities $\theta_1, \theta_2, \theta_3$, respectively. Repeat the trials n times

Define

X_1 : number of times A_1 occurs out of n trials

X_2 : - - - - - A_2 occurs out of n trials

X_3 : - - - - - A_3 occurs out of n trials

Let (x_1, x_2, x_3) denote the observed count in n trials

$$x_i \geq 0 , \quad x_i \leq n \quad \forall i = 1,2,3 \quad \& \quad \sum_1^3 x_i = n$$

$$E = \left\{ (x_1, x_2, x_3) : 0 \leq x_i \leq n , \ \sum_{i=1}^{3} x_i = n \right\} \text{ - finite number of points}$$

jt p.m.f.

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1!\, x_2!\, x_3!} \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3} ; \quad 0 \leq x_i \leq n, \ \sum x_i = n$$

Note that $x_3 = n - x_1 - x_2$ and

$$\theta_3 = 1 - \theta_1 - \theta_2$$

$$P(X_1 = x_1, X_2 = x_2) = \frac{n!}{x_1!\, x_2!\, (n - x_1 - x_2)!} \theta_1^{x_1} \theta_2^{x_2} (1 - \theta_1 - \theta_2)^{n - x_1 - x_2} \quad x_1, x_2 \geq 0, \ x_1 + x_2 \leq n$$

$$= 0, \quad o/w$$

(X_1, X_2) is said to follow a trinomial distn (n, θ_1, θ_2)

<u>Marginal distn's</u>:

Marginal p.m.f. of X_1:

$$P(X_1 = x_1) = \sum_{x_2 = 0}^{n - x_1} \frac{n!}{x_1!\,(n - x_1)!} \theta_1^{x_1} \frac{(n - x_1)!}{x_2!\,(n - x_1 - x_2)!} \theta_2^{x_2} (1 - \theta_1 - \theta_2)^{n - x_1 - x_2}$$

$$= \binom{n}{x_1} \theta_1^{x_1} (1 - \theta_1 - \theta_2 + \theta_2)^{n - x_1}$$

$$= \binom{n}{x_1} \theta_1^{x_1} (1 - \theta_1)^{n - x_1} .$$

i.e. $X_1 \sim Bin(n, \theta_1)$

Sly $X_2 \sim Bin(n, \theta_2)$

Conditional dist^n of $X_1 | X_2$:

$$P(X_1 = x_1 | X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)}$$

$$= \frac{\frac{\cancel{n!}}{x_1!\, \cancel{x_2!}\, (n - x_1 - x_2)!} \theta_1^{x_1} \cancel{\theta_2^{x_2}} (1 - \theta_1 - \theta_2)^{n - x_1 - x_2}}{\left(\frac{\cancel{n!}}{\cancel{x_2!}\, (n - x_2)!}\right) \cancel{\theta_2^{x_2}} (1 - \theta_2)^{n - x_2}}$$

$$= \frac{(n - x_2)!}{x_1!\, (n - x_2 - x_1)!} \left(\frac{\theta_1}{1 - \theta_2}\right)^{x_1} \left(1 - \frac{\theta_1}{1 - \theta_2}\right)^{n - x_2 - x_1}$$

$$x_1 = 0, 1, \ldots, n - x_2$$

i.e. $X_1 | X_2 \sim \text{Bin}\left(n - x_2, \frac{\theta_1}{1 - \theta_2}\right)$

Sly $X_2 | X_1 \sim \text{Bin}\left(n - x_1, \frac{\theta_2}{1 - \theta_1}\right)$

<u>Note</u>: Extension to $p > 3$ case – multinomial dist^n

$p > 3$ outcomes $A_1, \ldots, A_p$ – mutually exclusive & exhaustive

w.p. $\theta_1, \ldots, \theta_p$ $\quad \sum_1^p \theta_i = 1$; $\theta_i \geq 0$

n repeated trials

X_1: number of times A_1 occurs

X_2: - - - - - A_2 - - -

$\vdots$

X_p: - - - - - A_p - - -

Particular realisation $(x_1, \ldots x_p)$; $0 \leq x_i \leq n$ $\quad i = 1, \ldots p$

$$\sum_{i=1}^p x_i = n$$

$E = \{(x_1, \ldots, x_p) : 0 \leq x_i \leq n, \sum_1^p x_i = n\}$ ← finite number of points

↗ possible values of random vector $\underset{\sim}{X} = (X_1, \ldots, X_p)'$

jt p.m.f.

$$P(X_1=x_1,\ldots,X_p=x_p)=\begin{cases}\dfrac{n!}{x_1!\cdots x_p!}\,\theta_1^{x_1}\theta_2^{x_2}\cdots\theta_p^{x_p}, & x_i\geq 0,\ \sum_1^p x_i=n\\[2ex] 0, & o/w\end{cases}$$

Note that $x_p = n-x_1-\cdots-x_{p-1}$

$\theta_p = 1-\theta_1-\cdots-\theta_{p-1}$

jt p.m.f. of $X_1,\ldots X_{p-1}$

$$P(X_1=x_1,\ldots X_p=x_p)=\begin{cases}\dfrac{n!}{x_1!\,x_2!\cdots(n-x_1-\cdots-x_{p-1})!}\,\theta_1^{x_1}\theta_2^{x_2}\cdots(1-\theta_1-\cdots-\theta_{p-1})^{n-x_1-\cdots-x_{p-1}}, & x_i\geq 0;\ \sum_1^{p-1}x_i\leq n\\[2ex] 0, & o/w.\end{cases}$$

Marginal dist^{n}s:

$$X_i\sim B(n,\theta_i)\quad i=1,\ldots,p$$

jt marginal $(X_i,X_j)\sim \text{trinomial}(n,\theta_i,\theta_j)$

jt marginal $(X_i,X_j,X_K)\sim \text{multinomial}(n,\theta_i,\theta_j,\theta_K)$

Conditional dist^{n}s:

$$X_i\,|\,X_j=x_j\sim Bin\left(n-x_j,\ \frac{\theta_i}{1-\theta_j}\right)$$

$$(X_i,X_j)\,|\,X_K=x_K\sim \text{trinomial}\left(n-x_K,\ \frac{\theta_i}{1-\theta_K},\ \frac{\theta_j}{1-\theta_K}\right)$$

$$\left(n-x_K,\ \frac{\theta_i}{1-\theta_K},\ \frac{\theta_j}{1-\theta_K}\right)$$

Continuous multivariate distributions

A p-dimensional random vector $\underset{\sim}{X} = (X_1, \ldots, X_p)'$ is said to be (absolutely) continuous if $\exists$ a f^n $f_{X_1, \ldots X_p}(x_1, \ldots, x_p) \geq 0$ $\ni$ the jt d.f. of $(X_1, \ldots X_p)$ is expressed as

$$F_{X_1, \ldots X_p}(x_1, \ldots x_p) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f_{X_1, \ldots X_p}(t_1, \ldots, t_p)\, dt_1 \ldots dt_p$$

$f_{X_1, \ldots X_p}(x_1, \ldots, x_p)$ is called the jt p.d.f. of $(X_1, \ldots X_p)$

$$f_{X_1, \ldots X_p}(x_1, \ldots x_p) = \frac{\partial^p F_{X_1, \ldots X_p}(x_1, \ldots x_p)}{\partial x_1 \partial x_2 \ldots \partial x_p}$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 \ldots X_p}(x_1, \ldots x_p)\, dx_1 \ldots dx_p = 1$$

Note that all marginal p.d.f.s can be obtained from the joint p.d.f. (or jt d.f.)

Marginal dist^{n}s

Marginal ~~dist~~ p.d.f. of X_1:

$$f_{X_1}(x_1) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{p-1 \text{ fold};\ x_2 \to x_p} f_{\underset{\sim}{X}}(x_1, \ldots, x_p)\, dx_2 \ldots dx_p$$

marginal p.d.f. of any X_i

$$f_{X_i}(x_i) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{p-1 \text{ fold; all except } x_i} f_{\underset{\sim}{X}}(x_1, \ldots, x_p)\, dx_1 \ldots dx_{i-1}\, dx_{i+1} \ldots dx_p.$$

jt marginal p.d.f. of $X_{i_1}, \ldots X_{i_q}$

$$f_{X_{i_1}, \ldots, X_{i_q}}(x_{i_1}, \ldots, x_{i_q}) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{p-q \text{ fold; all except } x_{i_1}, \ldots x_{i_q}} f_{\underset{\sim}{X}}(x_1, \ldots, x_p) \, d(\quad)$$

Conditional dist^{n}s

Conditional p.d.f. of X_i given X_j :

$$f_{X_i | X_j}(x_i) = \frac{f_{X_i, X_j}(x_i, x_j)}{f_{X_j}(x_j)} \qquad \left(\text{for } f_{X_j}(x_j) > 0\right)$$

joint conditional p.d.f. of $(X_1, \ldots X_q)$ given $(X_{q+1}, \ldots, X_p)$

$$f_{X_1, \ldots, X_q | X_{q+1}, \ldots, X_p}(x_1, \ldots x_q) = \frac{f_{\underset{\sim}{X}}(\underset{\sim}{x})}{f_{X_{q+1}, \ldots, X_p}(x_{q+1}, \ldots, x_p)}$$

$\left(\text{for } f_{X_{q+1}, \ldots X_p}(x_{q+1}, \ldots, x_p) > 0\right)$

Sly conditional dist^n for any subset given any other subset can be obtained.

Independence

Def^n : $(X_1, \ldots, X_p)$ are pairwise indep iff

$\forall \, i \neq j \qquad f_{X_i, X_j}(x_i, x_j) = f_{X_i}(x_i) \, f_{X_j}(x_j) \quad \forall (x_i, x_j)$

Def^n : $(X_1, \ldots, X_p)$ are indep iff

$$f_{X_1, \ldots X_p}(x_1, \ldots, x_p) = \prod_{i=1}^{p} f_{X_i}(x_i) \qquad \forall \, \underset{\sim}{x}$$

Note : Independence $\Rightarrow$ pair wise indep.

Converse is not true

Expectation vector, covariance matrix

$\underset{\sim}{X}_{p\times 1}$: random vector

$$E\,\underset{\sim}{X} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_p \end{pmatrix} = \underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$$

Covariance:

$$\text{cov}(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j) \qquad i \neq j$$

Note that $\text{cov}(X_i, X_i) = E(X_i - \mu_i)^2 = V(X_i)$

$$\text{cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j).$$

In general, for $k_1, \ldots, k_p$ non-negative integers, we can define joint moment as

$$E\left(X_1^{k_1} X_2^{k_2} \cdots X_p^{k_p}\right) = \mu'_{k_1, \ldots, k_p}.$$

joint moment of order $k_1 + \cdots + k_p$.

Correlation:

$$\rho = \rho_{X_i, X_j} = \frac{\text{cov}(X_i, X_j)}{\left[V(X_i)\, V(X_j)\right]^{1/2}}$$

Note that $\text{cov}(X_i, X_j) = 0$ is referred to as uncorrelated. If X_i & X_j are indep then $\text{cov}(X_i, X_j) = 0$; but the converse is not true

Cauchy-Schwarz inequality

For any 2 r.v.s X & Y

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

(provided X & Y have finite 2^{nd} moment)

[Pf: Let $h(t) = E(tX - Y)^2 \geq 0 \quad \forall t$

$$E(tX - Y)^2 = t^2 E(X^2) + E(Y^2) - 2t\,E(XY)$$

If $h(t) > 0 \;\; \forall t$, then roots of $h(t)$ are not real

i.e. $\quad 4\left(E(XY)^2 - E(X^2)\,E(Y^2)\right) < 0.$

i.e $\left(E(XY)\right)^2 < E(X^2)\,E(Y^2)$

If $h(t) = 0$ for some t, say t^*, then

$$E(t^* X - Y)^2 = 0 \Rightarrow P(t^* X = Y) = 1$$

and we have equality in C-S inequality]

<u>Remark</u>:

Take $X = (X_i - EX_i)$ & $Y = (X_j - EX_j)$ in C-S inequality

$$\Rightarrow \mathrm{Cov}(X_i, X_j)^2 \le V(X_i)\,V(X_j)$$

i.e. $|\rho| \le 1$

<u>Covariance matrix</u>

$$\mathrm{Cov}(\underset{\sim}{X}) = \underset{p\times p}{\Sigma} = E(\underset{\sim}{X} - \underline{\mu})(\underline{X} - \underline{\mu})'$$

a symmetric matrix

$$= \begin{pmatrix} V(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_p) \\ & V(X_2) & \cdots & \mathrm{Cov}(X_2, X_p) \\ \vdots & & \ddots & \\ & & & V(X_p) \end{pmatrix}$$

$\underset{\sim}{X}_{p\times 1}$, $\underset{\sim}{Y}_{q\times 1}$

$$\mathrm{Cov}(\underset{\sim}{X}, \underset{\sim}{Y}) = E(\underset{\sim}{X} - \underset{\sim}{\mu}_X)(\underset{\sim}{Y} - \underset{\sim}{\mu}_Y)' = \underset{p\times q}{\Sigma_{XY}}$$

$$= \begin{pmatrix} \mathrm{Cov}(X_1, Y_1) & \mathrm{Cov}(X_1, Y_2) & \cdots & \mathrm{Cov}(X_1, Y_q) \\ \mathrm{Cov}(X_2, Y_1) & \mathrm{Cov}(X_2, Y_2) & \cdots & \mathrm{Cov}(X_2, Y_q) \\ - & - & - & - \\ \mathrm{Cov}(X_p, Y_1) & \mathrm{Cov}(X_p, Y_2) & \cdots & \mathrm{Cov}(X_p, Y_q) \end{pmatrix}$$

Correlation matrix

$$\mathrm{Cov}(\underset{\sim}{X}) = \Sigma = ((\sigma_{ij}))$$

$$R = \mathrm{Corr}^n(\underset{\sim}{X}) = \begin{pmatrix} 1 & \mathrm{Corr}(X_1, X_2) & \cdots & \mathrm{Corr}(X_1, X_p) \\ & 1 & \cdots & \mathrm{Corr}(X_2, X_p) \\ & & \ddots & \\ & & & 1 \end{pmatrix}$$

Let $D = \mathrm{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{pp})$

$$R = D^{-1/2} \Sigma D^{-1/2}$$

Linear combination of elements of $\underset{\sim}{X}$

$$\underset{\sim}{X}_{p\times 1} \longrightarrow Y = \underline{\alpha}'\underset{\sim}{X} \qquad \underset{\sim}{\alpha} \in \mathbb{R}^p$$

i.e. $Y = \sum_{i=1}^{p} \alpha_i X_i$; Y is a random variable

$$E(Y) = E\left(\sum_1^p \alpha_i X_i\right)$$
$$= \sum_1^p \alpha_i \, E X_i$$
$$= \sum_1^p \alpha_i \mu_i = \underline{\alpha}'\underline{\mu}$$

i.e. $E(\underline{\alpha}'\underline{X}) = \underline{\alpha}' E(\underline{X}) = \underline{\alpha}'\underline{\mu}$

$$V(\underline{\alpha}'\underline{X}) = E(\underline{\alpha}'\underline{X} - \underline{\alpha}'\underline{\mu})^2$$
$$= E(\underline{\alpha}'\underline{X} - \underline{\alpha}'\underline{\mu})(\underline{\alpha}'\underline{X} - \underline{\alpha}'\underline{\mu})'$$
$$= E\,\underline{\alpha}'(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'\underline{\alpha} = \underline{\alpha}' E(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'\underline{\alpha}$$
$$= \underline{\alpha}'\Sigma\underline{\alpha}$$

Multivariate normal

Def^n: A $p\times 1$ random vector $\underset{\sim}{X} = (X_1, \ldots, X_p)'$ with $E(\underline{X}) = \underset{\sim}{\mu}$ and $Cov(\underline{X}) = \Sigma$ is said to follow a multivariate normal, $N_p(\underline{\mu}, \Sigma)$ iff $\forall\ \underline{\alpha} \in \mathbb{R}^p\ (\underline{\alpha} \neq \underline{0})$, $\underline{\alpha}'\underset{\sim}{X}$ follows univariate normal (i.e. $\underset{\sim}{X} \sim N_p(\underline{\mu}, \Sigma)$ iff $\underline{\alpha}'\underset{\sim}{X} \sim N_1\ \forall\ \underline{\alpha} \in \mathbb{R}^p (\underline{\alpha} \neq \underline{0})$).

Note: Marginal dist^{n}s.

Marginal of X_i: $X_i \sim N(\mu_i, \sigma_{ii})$ where, $\Sigma = ((\sigma_{ij}))$

follows from the def^n of $N_p(\underline{\mu}, \Sigma)$; take

$$\underline{\alpha} = (0, \ldots 0, \underset{\underset{i\text{th}}{\uparrow}}{1}, 0 \ldots 0)'.$$

jt marginal of any subset of $\underset{\sim}{X}$, say $X_1, \ldots X_q$ $(q < p)$

$$\underset{\sim}{Y}_{q\times 1} = \begin{pmatrix} X_1 \\ \vdots \\ X_q \end{pmatrix} \sim N_q(\underset{\sim}{\mu}_q, \Sigma_{11})$$

$\underset{\sim}{\mu}_q = E(\underline{Y})$ & $\Sigma_{11} = Cov(\underset{\sim}{Y})$.

follows from def^n of N_p as

$$\forall\ \underset{\sim}{\beta} \in \mathbb{R}^q;\quad \underline{\beta}'\underset{\sim}{Y} = (\underline{\beta}', \underset{\sim}{0}'_{p-q})\underset{\sim}{X} \sim N_1 \text{ as } \underset{\sim}{X} \sim N_p.$$

Note: If $\underset{\sim}{X}_{p\times 1} \sim N_p(\underline{\mu}, \Sigma)$

$$\underset{\sim}{X} \to \underset{\sim}{Y}_{q\times 1} = A_{q\times p}\underset{\sim}{X} \sim N_q(A\underline{\mu}, A\Sigma A')$$

$\Big(\forall\ \underline{\beta} \in \mathbb{R}^q;\ \underline{\beta}'\underset{\sim}{Y} = \underline{\beta}'A\underline{X} = \underline{\alpha}^{*\prime}\underset{\sim}{X} \sim N_1$ (as $\underline{\alpha}'\underline{X} \sim N_1\ \forall\ \underline{\alpha} \in \mathbb{R}^p$$\Big)$

$\underline{\alpha}^* \in \mathbb{R}^p \quad \Rightarrow \underline{Y} \sim N_q;\ E(\underline{Y}) = A\underline{\mu}$, $Cov(\underline{Y}) = A\Sigma A'$.

Note: If $\Sigma > 0$, then p.d.f. of $\underset{\sim}{X} \sim N_p(\underline{\mu}, \Sigma)$ is

$$f_{\underset{\sim}{X}}(\underline{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\underset{\sim}{x}-\underline{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underline{\mu})} \quad \underset{\sim}{x} \in \mathbb{R}^p.$$

$p = 2$; $\underline{N_2(\underline{\mu}, \Sigma)}$ $\quad \underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$; $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$.

or $\underline{BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)}$

$\sigma_{ii} = V(X_i) = \sigma_i^2$ (say)

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad \sigma_{12} = \sigma_{21} = \mathrm{Cov}(X_1, X_2)$$

$$\rho = \frac{\sigma_{12}}{[\sigma_{11} \cdot \sigma_{22}]^{1/2}}.$$

$$f_X(\underline{x}) = \frac{1}{2\pi\ \sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right\}\right\}$$

$$\underset{\sim}{x} \in \mathbb{R}^2$$

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

Remark:

Note that if X_1 & X_2 are uncorrelated, i.e. $\rho = 0$, then

$$f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1)\ f_{X_2}(x_2)$$

$\Rightarrow X_1$ & X_2 are indep.

(such a conclusion is not true in general; i.e. uncorrelated $\not\Rightarrow$ independence in general).

For $\underset{\sim}{X} \sim N_p(\underline{\mu}, \Sigma)$

$$\underset{\sim}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_q \\ \hline X_{q+1} \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} \underset{\sim}{X}^{(1)} \\ \hline \underset{\sim}{X}^{(2)} \end{pmatrix} \quad \underset{\sim}{\mu} = \begin{pmatrix} \underset{\sim}{\mu}^{(1)} \\ \hline \underset{\sim}{\mu}^{(2)} \end{pmatrix}.$$

$$\Sigma = \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right). \qquad \Sigma_{ii} = \text{Cov}(\underset{\sim}{X}^{(i)}).$$

$\underset{\sim}{X}^{(1)}$ & $\underset{\sim}{X}^{(2)}$ are indep iff $\Sigma_{12} = 0$.

<u>Conditional distⁿ</u>

$$\underset{\sim}{X}^{(1)} | \underset{\sim}{X}^{(2)} \sim N_q\left(\underset{\sim}{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}\left(\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)}\right), \ \Sigma_{11.2} \right)$$

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Sly

$$\underset{\sim}{X}^{(2)} | \underset{\sim}{X}^{(1)} \sim N_{p-q}\left(\underset{\sim}{\mu}^{(2)} + \Sigma_{21}\Sigma_{11}^{-1}\left(\underset{\sim}{x}^{(1)} - \underset{\sim}{\mu}^{(1)}\right), \ \Sigma_{22.1} \right)$$

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

For $p = 2$, i.e. $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$$X_1 | X_2 \sim N_1\left(\mu_1 + (\rho\sigma_1\sigma_2)(\sigma_2^2)^{-1}(x_2 - \mu_2), \ \sigma_1^2 - \frac{(\rho\sigma_1\sigma_2)^2}{\sigma_2^2} \right)$$

$$\Sigma = \left(\begin{array}{c|c} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \hline \rho\sigma_1\sigma_2 & \sigma_2^2 \end{array}\right)$$

i.e. $X_1 | X_2 \sim N_1\left(\mu_1 + \rho\dfrac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \ \sigma_1^2(1-\rho^2) \right)$

Sly

$$X_2 | X_1 \sim N_1\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \ \sigma_2^2(1-\rho^2) \right).$$

<u>Note</u> : If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$; $\Sigma > 0$

then $\quad (\underline{X} - \underline{\mu})' \Sigma^{-1} (\underline{X} - \underline{\mu}) \sim \chi^2_p$

Proof of $\underset{\sim}{X}^{(1)}$ & $\underset{\sim}{X}^{(2)}$ indep iff $\Sigma_{12} = 0$

If $\underset{\sim}{X}^{(1)}$ & $\underset{\sim}{X}^{(2)}$ are indep then for any X_i in $\underset{\sim}{X}^{(1)}$ and X_j in $\underset{\sim}{X}^{(2)}$,

X_i & X_j are indep $\Rightarrow \text{Cov}(X_i, X_j) = 0 \Rightarrow \Sigma_{12} = 0$

Alternately suppose $\Sigma_{12} = 0$, then

$$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} ; \quad \Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}$$

& $|\Sigma| = |\Sigma_{11}|\,|\Sigma_{22}|$

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}^{(1)}, \underset{\sim}{x}^{(2)}) = \frac{1}{(2\pi)^{p/2}|\Sigma_{11}|^{1/2}|\Sigma_{22}|^{1/2}} \exp\left(-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu})'\begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}(\underset{\sim}{x} - \underset{\sim}{\mu})\right)$$

$$= \left(\frac{1}{(2\pi)^{q/2}|\Sigma_{11}|^{1/2}} \exp\left(-\frac{1}{2}(\underset{\sim}{x}^{(1)} - \underset{\sim}{\mu}^{(1)})'\Sigma_{11}^{-1}(\underset{\sim}{x}^{(1)} - \underset{\sim}{\mu}^{(1)})\right)\right)$$

$$\times \left(\frac{1}{(2\pi)^{\frac{p-q}{2}}|\Sigma_{22}|^{1/2}} \exp\left(-\frac{1}{2}(\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)})'\Sigma_{22}^{-1}(\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)})\right)\right)$$

$$= f_{\underset{\sim}{X}^{(1)}}(\underset{\sim}{x}^{(1)})\, f_{\underset{\sim}{X}^{(2)}}(\underset{\sim}{x}^{(2)})$$

$\underset{\sim}{X}^{(1)} \sim N_q(\underset{\sim}{\mu}^{(1)}, \Sigma_{11})$; $\underset{\sim}{X}^{(2)} \sim N_{p-q}(\underset{\sim}{\mu}^{(2)}, \Sigma_{22})$

$\Rightarrow \underset{\sim}{X}^{(1)}$ & $\underset{\sim}{X}^{(2)}$ are indep.

$\underset{\sim}{X} \sim N_p(\underline{\mu}, \Sigma)$; $\Sigma > 0$

Derivation of conditional distⁿ

Let

$$\underset{\sim}{Z} = \underbrace{\begin{bmatrix} I_q & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-q} \end{bmatrix}}_{A} \begin{pmatrix} \underline{X}^{(1)} - \underline{\mu}^{(1)} \\ \underline{X}^{(2)} - \underline{\mu}^{(2)} \end{pmatrix}$$

$$= \begin{pmatrix} \underline{X}^{(1)} - \underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) \\ \underset{\sim}{X}^{(2)} - \underline{\mu}^{(2)} \end{pmatrix} \sim N_p(\underline{0}, A\Sigma A')$$

$$A\Sigma A' = \begin{bmatrix} I_q & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-q} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_q & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I_{p-q} \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_q & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I_{p-q} \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}$$

$\Sigma_{11.2}$

$\Rightarrow \left(\underset{\sim}{X}^{(1)} - \underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) \right)$ & $(\underline{X}^{(2)} - \underset{\sim}{\mu}^{(2)})$ are indep

$\Rightarrow \left(\underset{\sim}{X}^{(1)} - \underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) \right) \mid \underline{X}^{(2)}$ and unconditional distⁿ are identical

$\Rightarrow \underset{\sim}{X}^{(1)} - \underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) \mid \underline{X}^{(2)} \sim N_q(\underline{0}, \Sigma_{11.2}$

$\Rightarrow \underset{\sim}{X}^{(1)} \mid \underset{\sim}{X}^{(2)} \sim N_q\left(\underline{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}), \Sigma_{11.2}\right)$

joint moment generating f^n

$\underset{\sim}{X} = (X_1, \ldots, X_p)'$

joint m.g.f.

$$M_{\underset{\sim}{X}}(\underset{\sim}{t}) = E\left(e^{\underset{\sim}{t}'\underset{\sim}{X}}\right) = E\left(e^{t_1X_1 + \cdots + t_pX_p}\right).$$

$\underset{\sim}{t} \downarrow (t_1, \ldots, t_p)$

provided the expectation exists in some nbd of $\underset{\sim}{0}_{p\times 1}$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{\sum_1^p t_i x_i} f_X(\underline{x})\, dx_1 \ldots dx_p.$$

for continuous case

$$= \sum_{x_1} \cdots \sum_{x_p} e^{\sum t_i x_i} P(\underline{X} = \underset{\sim}{x})$$

for discrete case

Joint m.g.f.

$\underset{\sim}{X} = (X_1, \ldots, X_p)'$

joint m.g.f.

$$M_{\underset{\sim}{X}}(\underset{\sim}{t}) = E\left(e^{\underset{\sim}{t}'\underset{\sim}{X}}\right) = E\left(e^{t_1X_1 + \cdots + t_pX_p}\right)$$

$(t_1, \ldots t_p)$ provided the expectation exists in some nbd of $\underset{\sim}{0}_{p\times 1}$

$$M_X(\underset{\sim}{t}) = \int \cdots \int e^{\sum_1^p t_i x_i} f_{\underset{\sim}{X}}(\underset{\sim}{x})\, dx_1 \cdots dx_p \quad \text{for continuous case}$$

$$= \sum_{x_1} \cdots \sum_{x_p} e^{\sum_1^p t_i x_i} P(\underset{\sim}{X} = \underset{\sim}{x}) \quad \text{for discrete case}$$

$$\mu'_{k_1, \ldots, k_p} = E\left(X_1^{k_1} \cdots X_p^{k_p}\right) = \left.\frac{\partial^{k_1 + \cdots + k_p} M_{\underset{\sim}{X}}(\underset{\sim}{t})}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}}\right|_{\underset{\sim}{t} = \underset{\sim}{0}}$$

joint moment of order $k_1 + \cdots + k_p$

k_is are non-negative integers

Note that we can get marginal distn m.g.f s from the joint m.g.f.

$$M_{\underset{\sim}{X}}(0, \ldots 0, t_i, 0 \ldots 0) = E\left(e^{t_iX_i}\right) = M_{X_i}(t_i)$$

i^{th} position

Sly marginal joint m.g.f. for any subset can be obtained from $M_X(\underset{\sim}{t})$.

<u>Note</u>: m.g.f. of $X_1 + \cdots + X_k$ for $k \le p$ can be obtained from $M_{\underset{\sim}{X}}(\underset{\sim}{t})$

<u>Remark</u>: $X_1, \ldots, X_p$ are independent iff

$$M_{\underset{\sim}{X}}(\underset{\sim}{t}) = \prod_{i=1}^{p} M_{X_i}(t_i)$$

M.g.f. of $N_p(\underset{\sim}{\mu}, \Sigma)$

$$M_{\underset{\sim}{X}}(\underset{\sim}{t}) = E(e^{\underset{\sim}{t}'\underset{\sim}{X}})$$

$$= M_{\underset{\sim}{t}'\underset{\sim}{X}}(1)$$

$$= e^{\underset{\sim}{t}'\underset{\sim}{\mu} + \frac{1}{2}\underset{\sim}{t}'\Sigma\underset{\sim}{t}}$$

$\underset{\sim}{t}'\underset{\sim}{X} \sim N_1(\underset{\sim}{t}'\underset{\sim}{\mu}, \underset{\sim}{t}'\Sigma\underset{\sim}{t})$

m.g.f.

$$M_{\underset{\sim}{t}'\underset{\sim}{X}}(z) = E(e^{z(\underset{\sim}{t}'\underset{\sim}{X})})$$

$$= e^{z(\underset{\sim}{t}'\underset{\sim}{\mu}) + \frac{z^2}{2}(\underset{\sim}{t}'\Sigma\underset{\sim}{t})}$$

- Marginal m.g.f s can be derived using the above.

M.g.f. of trinomial (n, θ_1, θ_2)

$\underset{\sim}{X} = (X_1, X_2) \sim \text{trinomial}(n, \theta_1, \theta_2)$

$$M_{\underset{\sim}{X}}(t_1, t_2) = E(e^{t_1X_1 + t_2X_2})$$

$$= \sum_{x_1=0}^{n}\sum_{x_2=0}^{n-x_1} e^{t_1x_1+t_2x_2} \frac{n!}{x_1!\,x_2!\,(n-x_1-x_2)!}\theta_1^{x_1}\theta_2^{x_2}(1-\theta_1-\theta_2)^{n-x_1-x_2}$$

$$= \sum_{x_1=0}^{n} \frac{n!}{x_1!(n-x_1)!}(\theta_1e^{t_1})^{x_1}\sum_{x_2=0}^{n-x_1}\frac{(n-x_1)!}{x_2!(n-x_1-x_2)!}(\theta_2e^{t_2})^{x_2}(1-\theta_1-\theta_2)^{n-x_1-x_2}$$

$$= \sum_{x_1=0}^{n}\binom{n}{x_1}(\theta_1e^{t_1})^{x_1}(1-\theta_1-\theta_2+\theta_2e^{t_2})^{n-x_1}$$

$$= (1-\theta_1-\theta_2+\theta_1e^{t_1}+\theta_2e^{t_2})^n$$

Sly If $(X_1, \ldots, X_{p-1}) \sim \text{mult}(n, \theta_1, \theta_2, \ldots, \theta_{p-1})$

$$M_{\underset{\sim}{X}}(t_1, \ldots, t_{p-1}) = (1-\theta_1-\theta_2\cdots-\theta_{p-1}+\theta_1e^{t_1}+\cdots+\theta_{p-1}e^{t_{p-1}})^n$$

Marginal m.g.f.s can be derived.

Distⁿ of $X_i + X_j$ can be derived thro joint m.g.f.

Conditional expectation

Consider a bivariate setup $(X, Y) \to$ jt p.d.f $f_{X,Y}(x,y)$

marginal p.d.f.s : $f_X(x)$, $f_Y(y)$

Conditional p.d.f.s : $f_{X|Y}(x)$, $f_{Y|X}(y)$

Note that $f_{X,Y}(x,y) = f_X(x)\, f_{Y|X}(y)$

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

Let $g(X,Y)$ be some f^n of X & Y

$E(g(X,Y)|X)$: Conditional expectation of $g(X,Y)$ given X

$E(g(X,Y)|Y)$: Conditional expectation of $g(X,Y)$ given Y

$$E(g(X,Y)|X) = E_{Y|X}(g(X,Y)|X)$$

$$= \int_{-\infty}^{\infty} g(x,y)\, f_{Y|X}(y)\, dy \to f^n \text{ of } X \text{ only}$$

$$= \phi(X), \text{ say}$$

$$E_X(\phi(X)) = E_X\, E_{Y|X}(g(X,Y)|X)$$

$$= \int_{-\infty}^{\infty} \phi(x)\, f_X(x)\, dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{Y|X}(y)\, f_X(x)\, dy\, dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)\, dy\, dx$$

$$= E_{X,Y}(g(X,Y))$$

i.e $\quad E_X E_{Y|X}(g(X,Y)|X) = E_{X,Y}(g(X,Y))$

<u>sp. case</u> : $g(X,Y) = Y$ say

$$E\,E(Y|X) = E(Y)$$

sly $E\,E(X|Y) = E(X)$

<u>sp. case</u> : $g(X,Y) = (Y-E(Y))^2$

$$E(Y-E(Y))^2 = E_X E_{Y|X}((Y-E(Y))^2|X)$$

Note that

$E(Y-E(Y))^2|X$

$= E(Y-E(Y|X)+E(Y|X)-E(Y))^2|X$

$= E(Y-E(Y|X))^2|X + (E(Y|X)-E(Y))^2$

$\quad + 2\,\cancel{E((Y-E(Y|X))|X)}(E(Y|X)-E(Y))$

$\quad 0$ (as $E(Y-E(Y|X))|X$
$\quad = E(Y|X) - E(Y|X) = 0$)

$= V(Y|X) + (E(Y|X)-E(Y))^2$

$\Rightarrow E\,E[(Y-E(Y))^2|X]\ (= V(Y)$

$= E(V(Y|X)) + V(E(Y|X)))$

i.e. $\quad V(Y) = E(V(Y|X)) + V(E(Y|X))$

<u>Note</u>: Joint m.g.f. can be derived through m.g.f. of conditional distⁿ using conditional expectation

e.g. $(X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$$M_{X_1,X_2}(t_1,t_2) = E\left(e^{t_1X_1+t_2X_2}\right) \quad ; \quad g(x_1,x_2) = e^{t_1x_1+t_2x_2}$$

$$= E\, E\left(e^{t_1X_1+t_2X_2} \,\middle|\, X_1\right)$$

$$= E_{X_1}\left(e^{t_1X_1}\left(E_{X_2|X_1} e^{t_2X_2} \,|\, X_1\right)\right)$$

Now, use the fact that

$$X_2|X_1 \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1),\ \sigma_2^2(1-\rho^2)\right).$$

Random Sampling and functions of random variables

Suppose, we have

$X_1, \ldots, X_n$ a random sample from a dist^n with p.d.f f_θ
or p.m.f. f_θ
(here $\theta \in \Theta$ is the characterising parameter)

"random sample" $\Rightarrow X_1, \ldots, X_n$ are indep.

from the same dist^n $\Rightarrow X_1, \ldots, X_n$ have identical dist^n

Thus "$X_1, \ldots, X_n$ is a random sample from f_θ"

$\Leftrightarrow$ "$X_1, \ldots, X_n$ are independently and identically distributed with f_θ"

Let Y = function of the random sample (not involving the parameter)

We call such functions "statistic"

e.g. $Y_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \to$ sample mean r.v.

$Y_2 = \frac{1}{n} \sum (X_i - \bar{X})^2 = S_n^2 \to$ sample variance r.v.

or $(n-1)$

$Y_3 = \max(X_1, \ldots, X_n) = X_{(n)} \to$ maximum order statistic

$Y_4 = \min(X_1, \ldots, X_n) = X_{(1)} \to$ minimum order statistic

Point of interest: To know the prob law of such f^{n}s of the random sample

i.e. $(X_1, \ldots, X_n) \to Y = f(X_1, \ldots, X_n)$
what is the p.d.f. / p.m.f. of Y

(I) <u>M.g.f. based approach</u> (provided m.g.f. exists)

Applicable for standard distributions with readily identifiable m.g.f.

<u>EX1</u> : Additive property of standard distⁿs

(a) $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$

$$Y = \sum_{i=1}^{n} X_i$$

$$M_Y(t) = M_{\sum_1^n X_i}(t) = E\left(e^{t\sum X_i}\right)$$

$$= E\left(e^{tX_1} e^{tX_2} \cdots e^{tX_n}\right)$$

$$= \prod_{i=1}^{n} E\left(e^{tX_i}\right) \quad (\because X_1, \ldots X_n \text{ are indep}).$$

$$= \prod_{i=1}^{n} e^{t\mu + \frac{t^2}{2}\sigma^2} = e^{tn\mu + \frac{t^2}{2}n\sigma^2}$$

$\Rightarrow Y \sim N(n\mu, n\sigma^2)$ by uniqueness of m.g.f.

If $X_1, \ldots X_n$ are indep $N(\mu_i, \sigma_i^2)$, then $\sum_1^n X_i \sim N(\sum \mu_i, \sum \sigma_i^2)$

(b) $X_1, \ldots, X_n$ i.i.d. $P(\lambda)$

$$Y = \sum_{i=1}^{n} X_i \sim P(n\lambda) \; ; \quad M_Y(t) = \prod_{i=1}^{n} e^{\lambda(e^t - 1)} = e^{n\lambda(e^t-1)}$$

If $X_1, \ldots, X_n$ are indep $P(\lambda_i)$

then $Y = \sum_1^n X_i \sim P\left(\sum_{i=1}^{n} \lambda_i\right)$.

(c) $X_1, \ldots, X_n$ ind $X_i \sim B(n_i, p)$

same

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \left(q + pe^t\right)^{n_i} = \left(q + pe^t\right)^{\sum n_i}$$

$$\Rightarrow Y \sim B\left(\sum_{i=1}^{n} n_i, p\right).$$

(II) Dist^n f^n based approach

Let $X_1, \dots, X_n$ be indep continuous r.v.s with p.d.f. $f_X(x)$ and d.f. $F_X(x)$.

Let $Y = X_{(1)} = \min\{X_1, \dots, X_n\}$ — smallest order statistic

$Z = X_{(n)} = \max\{X_1, \dots, X_n\}$ — Largest order statistic

$$F_Y(y) = P(Y \le y) = P\left(\min\{X_1, \dots, X_n\} \le y\right)$$

d.f. of Y

$$= 1 - P\left(\min\{X_1, \dots, X_n\} > y\right)$$

$$= 1 - \prod_{i=1}^{n} P(X_i > y) \quad [\because \text{ of independence}]$$

$$= 1 - \left(1 - F_X(y)\right)^n \quad [\because X_1, \dots X_n \text{ have identical dist}^n]$$

(This step will hold for discrete setup also)

p.d.f. of Y:

$$f_Y(y) = n\left(1 - F_X(y)\right)^{n-1} f_X(y)$$

Example: $X_1, \dots, X_n$ i.i.d. (independent and identically distributed) $\exp(1)$

i.e. $f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & o/w \end{cases}$; $F_X(x) = 1 - e^{-x} \quad x > 0$

$$F_Y(y) = 1 - e^{-ny} \quad y > 0$$

$$f_Y(y) = \begin{cases} n e^{-ny} & y > 0 \\ 0, & o/w \end{cases}$$

$$Z = \max\{X_1, \ldots, X_n\}$$

$$F_Z(z) = P(Z \le z)$$

$$= P(\max\{X_1, \ldots, X_n\} \le z)$$

$$= P(X_1 \le z, \ldots, X_n \le z)$$

$$= \prod_{i=1}^{n} P(X_i \le z) \qquad [\because \text{ of independence}]$$

$$= (F_X(z))^n \qquad [\because \text{ of identical dist}^n]$$

(upto this ~~...~~ step same for discrete & cont dist^n)

p.d.f. of Z:

$$f_Z(z) = n\,(F_X(z))^{n-1} f_X(z).$$

Example: $f_X(x) = e^{-x} \quad x > 0 \quad ; \quad F_X(x) = 1 - e^{-x} \quad x > 0$

$$F_Z(z) = (1 - e^{-z})^n$$

$$f_Z(z) = \begin{cases} n(1 - e^{-z})^{n-1} e^{-z}, & z > 0 \\ 0, & o/w. \end{cases}$$

Remark: In general, suppose jt p.d.f. of $(X_1, \ldots X_n)$ is

$$f_{X_1, \ldots X_n}(x_1, \ldots, x_n)$$

$$(X_1, \ldots, X_n) \rightarrow u(X_1, \ldots, X_n) = Y.$$

$$F_Y(y) = P(U(X_1, \ldots, X_n) \le y)$$

$$= \int_{u(x_1, \ldots x_n) \le y} \cdots \int f_{X_1 \ldots X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

Remark: Distⁿ of r^{th} order statistic : $X_{(r)}$

$X_{(r)} = r^{th}$ smallest of $\{X_1, \ldots, X_n\}$ $r = 1, \ldots, n$

p.d.f. of $X_{(r)}$:

$$\frac{n!}{(r-1)!\,(n-r)!} (F(x))^{r-1} (1-F(x))^{n-r} f(x) \qquad x \in \mathbb{R}$$

jt p.d.f. of $X_{(r)}$ & $X_{(s)}$ $\quad 1 \le r < s \le n$

$$f_{X_{(r)}, X_{(s)}}(x, y) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} (F(x))^{r-1} (F(y) - F(x))^{s-r-1} \times$$

$$(1 - F(y))^{n-s} f(x) f(y)$$

$$-\infty < y < x < \infty$$

e.g. jt distⁿ of $(X_{(1)}, X_{(n)})$

$$f_{X_{(1)}, X_{(n)}}(x, y) = \frac{n!}{(n-2)!} (F(y) - F(x))^{n-2} f(x) f(y)$$

can be used to obtain p.d.f of range statistic $R = X_{(n)} - X_{(1)}$

(iii) p.m.f. based approach for discrete setup

Suppose $(X_1, \dots, X_n)$ have joint p.m.f.

$$P(\underset{\sim}{X} = \underset{\sim}{x}) = P(X_1 = x_1, \dots, X_n = x_n) \qquad \underset{\sim}{x} \in \mathcal{X}$$

$(X_1, \dots, X_n) \to Y = u(X_1, \dots, X_n)$ with $\mathcal{Y}$ as possible values of Y

$$P(Y = y) = \sum_{\substack{\underset{\sim}{x} \in \mathcal{X} \\ \ni u(\underset{\sim}{x}) = y}} P(\underset{\sim}{X} = \underset{\sim}{x})$$

In case we have $X_1, \dots, X_n$ a random sample (implying independence) with a common p.m.f (identical distn) then $P(\underset{\sim}{X} = \underset{\sim}{x})$ factors into n components with identical marginal p.m.f.s.

Example: Let X_1 and X_2 are indep with

$$X_i \sim P(\lambda_i) \qquad i = 1, 2$$

$$Y = X_1 + X_2 \qquad \mathcal{Y} = \{0, 1, 2, \dots\}$$

$$P(Y = y) = \sum_{x=0}^{y} P(X_1 = x, X_2 = y - x)$$

$$= \sum_{x=0}^{y} P(X_1 = x) \; P(X_2 = y - x)$$

$$= \sum_{x=0}^{y} \frac{e^{-\lambda_1} \lambda_1^x}{x!} \; \frac{e^{-\lambda_2} \lambda_2^{y-x}}{(y-x)!}$$

$$= e^{-(\lambda_1 + \lambda_2)} \frac{1}{y!} \sum_{x=0}^{y} \binom{y}{x} \lambda_1^x \lambda_2^{y-x}$$

$$= e^{-(\lambda_1 + \lambda_2)} \frac{1}{y!} (\lambda_1 + \lambda_2)^y$$

i.e. $Y = X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$

Remark : If $X_1, \dots, X_n$ are indep with $X_i \sim P(\lambda_i)$; $i=1,\dots,n$

then $Y = \sum_{i=1}^{n} X_i \sim P\left(\sum_{i=1}^{n} \lambda_i\right)$

$\overline{V}$ <u>Jacobian based approach</u> (cont case only)

$\underset{\sim}{X} = (X_1, \dots, X_n)'$

jt p.d.f. of $X_1, \dots, X_n$: $f_{X_1,\dots X_n}(x_1, \dots, x_n)$

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) > 0 \text{ for } \underset{\sim}{x} \in \mathcal{X} \subset R^n$$

suppose
$$y_1 = h_1(x_1, \dots, x_n)$$
$$\vdots$$
$$y_n = h_n(x_1, \dots, x_n)$$

be 1-1 transformation $\mathcal{X} \to \mathcal{Y}$ with inverse as

$$x_i^{-1} = h_i^{-1}(y_1, \dots, y_n) \;;\; i=1,\dots,n$$

$$\mathcal{Y} = \underset{\sim}{h}(\mathcal{X}) = \{\underset{\sim}{h}(\underset{\sim}{x}) \in R^n : \underset{\sim}{x} \in \mathcal{X}\}$$

Suppose further that

(i) $\dfrac{\partial h_i^{-1}(\underset{\sim}{y})}{\partial y_j}$ exists $\forall i, j$ and are continuous

& (ii) the Jacobian determinant

$$J = \begin{vmatrix} \dfrac{\partial h_1^{-1}(\underset{\sim}{y})}{\partial y_1} & \cdots & \dfrac{\partial h_1^{-1}(\underset{\sim}{y})}{\partial y_n} \\ - & - & - \\ \dfrac{\partial h_n^{-1}(\underset{\sim}{y})}{\partial y_1} & \cdots & \dfrac{\partial h_n^{-1}(\underset{\sim}{y})}{\partial y_n} \end{vmatrix} \neq 0.$$

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \to \begin{pmatrix} Y_1 = h_1(\underset{\sim}{X}) \\ \vdots \\ Y_n = h_n(\underline{X}) \end{pmatrix}$$

The joint p.d.f. of new r.v.s $(Y_1, \dots, Y_n)$ is

$$f_{Y_1 \dots Y_n}(y_1, \dots, y_n) = f_{\underset{\sim}{X}}\left(h_1^{-1}(\underset{\sim}{y}), \dots, h_n^{-1}(\underline{y})\right) |J| \underbrace{I_{(\underset{\sim}{y} \in \underset{\sim}{h}(\underset{\sim}{x}))}}_{\text{inf on the support of } \underset{\sim}{Y}}$$

Remark: From $f_{\underset{\sim}{Y}}(\underline{y})$ we can obtain marginal of Y_i; $i=1,\dots,n$

Remark: Based on $(X_1, \dots, X_n)$, suppose we are interested in jt dist^n of $(Y_1, \dots, Y_K)$ $K < n$, we define $n-K$ dummy transformations (key is to keep the dummies simple!) and integrate out the dummies to get the jt dist^n of $(Y_1, \dots, Y_K)$

i.e. $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \to \begin{pmatrix} Y_1 \\ \vdots \\ Y_K \\ \vdots \\ Y_n \end{pmatrix} \to \begin{pmatrix} Y_1 \\ \vdots \\ Y_K \end{pmatrix}$. $\} \to$ dummies

Remark: The above approach can be extended to the case where we have 1-1 mapping in mutually disjoint regions of $\mathcal{X}$

$$\mathcal{X} = \bigcup_{i=1}^{K} \mathcal{X}_i \quad \& \quad \mathcal{X}_i \cap \mathcal{X}_j = \phi$$

$\underset{\sim}{h} = (h_1, \ldots, h_n)$ is 1-1 with the inverse

$$h_i^{-1}(\underset{\sim}{y}) = \left(h_{1,i}^{-1}(\underset{\sim}{y}), \ldots, h_{n,i}^{-1}(\underset{\sim}{y}) \right) \text{ on } \mathcal{X}_i$$

$$\frac{\partial h_{k,i}^{-1}(\underset{\sim}{y})}{\partial y_j}$$ exists and are continuous with Jacobian

determinants

$$J_i = \begin{vmatrix} \dfrac{\partial h_{1,i}^{-1}(\underset{\sim}{y})}{\partial y_1} & \cdots & \dfrac{\partial h_{1,i}^{-1}(\underset{\sim}{y})}{\partial y_n} \\ \vdots & & \vdots \\ \dfrac{\partial h_{n,i}^{-1}(\underset{\sim}{y})}{\partial y_1} & \cdots & \dfrac{\partial h_{n,i}^{-1}(\underset{\sim}{y})}{\partial y_n} \end{vmatrix} \neq 0$$

$i = 1, \ldots, K$

then

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \sum_{j=1}^{K} f_{\underset{\sim}{X}}\left(h_{1,j}^{-1}(\underset{\sim}{y}), \ldots, h_{n,j}^{-1}(\underset{\sim}{y}) \right) |J_j|$$

Example:

X_1, X_2 i.i.d. $\exp(1)$

i.e. $f_X(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & o/w \end{cases}$

$Y = X_1 - X_2$ — interested to know p.d.f. of Y

Define a 1-1 transformation (with a dummy)

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \to \begin{pmatrix} Y = X_1 - X_2 \\ Z = X_1 + X_2 \end{pmatrix} \leftarrow \text{can be different}$$

Inverse transformation

$$X_1 = \frac{Y+Z}{2} = h_1^{-1}(y, z)$$

$$X_2 = \frac{Z-Y}{2} = h_2^{-1}(y, z)$$

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y} & \frac{\partial x_1}{\partial z} \\ \frac{\partial x_2}{\partial y} & \frac{\partial x_2}{\partial z} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{vmatrix} = \frac{1}{2}$$

Support calculation : y

Note that unconditionally $0 < z < \infty$ & $-\infty < y < \infty$

Note further that from inverse transformation we have:

$$\left.\begin{aligned} 0 < x_1 < \infty; \quad \text{i.e. } & 0 < \frac{y+z}{2} < \infty \\ \text{i.e. } & -y < z < \infty \;/\; -z < y < \infty \end{aligned}\right\} (*^1)$$

$$\left.\begin{aligned} \& \quad 0 < x_2 < \infty; \quad \text{i.e. } & 0 < \frac{z-y}{2} < \infty \\ \text{i.e. } & y < z < \infty \;/\; -\infty < y < z \end{aligned}\right\} (*^2)$$

Combining $(*^1)$ & $(*^2)$, we get

$$\left.\begin{aligned} &\max(y,-y) < z < \infty \\ \& \quad &-z < y < z \end{aligned}\right\} \text{ with } \begin{aligned} &0 < z < \infty \\ -&\infty < y < \infty \end{aligned}$$

Thus, if $-\infty < y < 0$ then $-y < z < \infty$

& if $0 < y < \infty$ then $y < z < \infty$

$$\Rightarrow f_{Y,Z}(y,z) = \begin{cases} \frac{1}{2} e^{-z}; & (-\infty < y < 0 \text{ and } -y < z < \infty) \\ & \text{or } (0 < y < \infty \text{ and } y < z < \infty) \\ 0, & \text{o/w.} \end{cases}$$

$\Rightarrow$ Marginal p.d.f. of Y (variable of interest)

$$f_Y(y) = \frac{1}{2} \int_{-y}^{\infty} e^{-z}\, dz = \frac{1}{2} e^{y}; \qquad \text{if } -\infty < y < 0$$

$$= \frac{1}{2} \int_{y}^{\infty} e^{-z}\, dz = \frac{1}{2} e^{-y}; \qquad \text{if } 0 < y < \infty$$

i.e. $f_Y(y) = \frac{1}{2} e^{-|y|}; \quad y \in (-\infty, \infty)$

Some important dist^n obtained through transformations

(I) $X_1, \ldots, X_n$ i.i.d $N(0,1)$

$X_i^2 \sim \chi_1^2$ and are i.i.d.

$\sum_{i=1}^{n} X_i^2 \sim \chi_n^2 \leftarrow$ chi-square on n degrees of freedom

(II)

$U \sim N(0,1)$

$V \sim \chi_r^2$

U & V are indep

$$T = \frac{U}{\sqrt{V/r}} \sim \text{Student's t-dist}^n \text{ on } r \text{ degrees of freedom}$$

$T \sim t_r$

(III)

$U \sim \chi_r^2$

$V \sim \chi_s^2$

U & V are indep

$$F = \frac{U/r}{V/s} \sim F \text{ dist}^n \text{ with } (r,s) \text{ degrees of freedom}$$

$F \sim F_{r,s}$

Remark : $F \sim F_{r,s} \Leftrightarrow \frac{1}{F} \sim F_{s,r}$

$F_{1,s} \overset{d}{=} t_s^2$

An important result: Sampling from $N(\mu, \sigma^2)$

Suppose $X_1, \ldots, X_n$ be a random sample (i.e $X_1, \ldots X_n$ are independent and identically distributed) from $N(\mu, \sigma^2)$ distribution; $\mu \in \mathbb{R}$, $\sigma > 0$

Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$: sample mean random variable

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$: sample variance random variable.

i.e. $\bar{X} = f_1(X_1, \ldots, X_n)$

$S^2 = f_2(X_1, \ldots, X_n)$

Then

(i) $\bar{X} \sim N(\mu, \sigma^2/n)$

(ii) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

& (iii) $\bar{X}$ & S^2 are independent

Remark:

$$\left.\begin{array}{l} \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \\[2ex] \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \end{array}\right\rangle \text{independent}$$

$$\Rightarrow \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

Proof of the results (i)-(iii)

jt p.d.f. of $(X_1, \ldots, X_n)$

$$f_{X_1,\ldots X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

i.e. $f_{\underset{\sim}{X}}(\underset{\sim}{x}) = K \exp\left(-\frac{1}{2\sigma^2} \sum_1^n (x_i - \mu)^2\right)$

K does not depend on $(x_1, \ldots, x_n)$

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = K \exp\left(-\frac{1}{2\sigma^2} \sum_1^n \left(x_i - \bar{x} + \bar{x} - \mu\right)^2\right)$$

$$= K \exp\left(-\frac{1}{2\sigma^2} \Sigma (x_i - \bar{x})^2 - \frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right)$$

Make the following transformation

$$\begin{pmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{pmatrix} \longrightarrow \begin{pmatrix} Y_1 = \frac{1}{n}(X_1 + \cdots + X_n) \\ Y_2 = X_2 - \bar{X} \\ \vdots \\ Y_n = X_n - \bar{X} \end{pmatrix} \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Inverse transformation:

$$X_1 = Y_1 - Y_2 - \cdots - Y_n$$

$$X_2 = Y_1 + Y_2$$

$$X_3 = Y_1 + Y_3$$

$$\vdots$$

$$X_n = Y_1 + Y_n$$

Jacobian determinant

$$J = \begin{vmatrix} 1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 \cdots & 0 \\ \vdots & & & & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

$$= \begin{vmatrix} n & 0 & 0 & \cdots & 0 \\ 1 & & & & \\ \vdots & & \left(\quad I_n \quad \right) & & \\ 1 & & & & \end{vmatrix} = n$$

From the inverse transformation note that

$$(x_i - \bar{x}) = y_i \quad \text{for } i = 2, 3, \ldots, n$$

and $x_1 - \bar{x} = (y_1 - y_2 - \cdots - y_n) - y_1 = \left(-\sum_{i=2}^{n} y_i\right)$

Thus the jt p.d.f. of the random variables $Y_1, \ldots, Y_n$ is

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = K' \exp\left(-\frac{1}{2\sigma^2}(-y_2 - y_3 - \cdots - y_n)^2 - \frac{1}{2\sigma^2}\sum_{i=2}^{n} y_i^2 - \frac{n}{2\sigma^2}(y_1 - \mu)^2\right)$$

$$= f_{Y_2,\ldots,Y_n}(y_2, \ldots, y_n)\, f_{Y_1}(y_1) \qquad \underset{\sim}{y} \in \mathbb{R}^n$$

$\Rightarrow (Y_2, \ldots, Y_n)$ and Y_1 are independent

$\Rightarrow Y_1$ & (any function of $Y_2, \ldots, Y_n$) are independent

Note further that $Y_1 = \bar{X}$

& $s^2 = \frac{1}{n-1}\left[(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2\right]$

i.e. $s^2 = \frac{1}{n-1}\left[(-Y_2 \cdots - Y_n) + Y_2^2 + \cdots + Y_n^2\right]$

f^n of $(Y_2, \ldots, Y_n)$

$\Rightarrow \bar{X}$ & s^2 are independent

Further $f_{Y_1}(y_1) = K_1 \exp\left(-\frac{n}{2\sigma^2}(y_1-\mu)^2\right) \quad -\infty < y_1 < \infty$

$\Rightarrow Y_1 = \bar{X} \sim N(\mu, \sigma^2/n)$

Note that

$X_i \sim N(\mu, \sigma^2)$ i.i.d.

$\frac{X_i-\mu}{\sigma} \sim N(0,1)$ i.i.d.

$\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_1^2$ i.i.d.

$\Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi_n^2$

and $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 + \frac{n(\bar{X}-\mu)^2}{\sigma^2}$

$= \frac{(n-1)s^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2} \quad — (*)$

Note that m.g.f. of l.h.s. of (*) is $(1-2t)^{-n/2}$

(m.g.f. of χ^2_n)

i.e. $(1-2t)^{-n/2} = E\left(e^{t\left(\frac{1}{\sigma^2}\sum_1^n (x_i-\mu)^2\right)}\right)$

$$= E\left(e^{t\left\{\frac{(n-1)s^2}{\sigma^2} + \frac{n(\bar{x}-\mu)^2}{\sigma^2}\right\}}\right)$$

$$= E\left(e^{t\frac{(n-1)s^2}{\sigma^2}}\right) E\left(e^{t\frac{n(\bar{x}-\mu)^2}{\sigma^2}}\right)$$

(using independence)

$$= E\left(e^{t\frac{(n-1)s^2}{\sigma^2}}\right) (1-2t)^{-1/2}$$ m.g.f. of χ^2_1

$$\Rightarrow E\left(e^{t\frac{(n-1)s^2}{\sigma^2}}\right) = (1-2t)^{-\frac{(n-1)}{2}} = M_{\frac{(n-1)s^2}{\sigma^2}}(t)$$

By uniqueness of m.g.f. it follows that

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

—x—

Convergence of sequence of random variables :

Modes of convergence :

Convergence in probability - to be covered in this course
Convergence in distribution - to be covered in this course
Convergence almost surely
Convergence in r^{th} mean

Convergence in probability

Let $\{X_n\}$ be a sequence of random variables on $(\Omega, \mathcal{F}, P)$
$\{X_n\}$ is said to converge in probability to a random variable X (we write $X_n \xrightarrow{p} X$ as $n \to \infty$) if

$$P(|X_n - X| > \epsilon) \to 0 \quad \text{as } n \to \infty \quad \forall \epsilon > 0$$

Some important results

(i) If $X_n \xrightarrow{p} X$ and 'a' is a constant, then

$$a X_n \xrightarrow{p} a X$$

(ii) If $X_n \xrightarrow{p} X$ and $g(.)$ is any continuous function, then

$$g(X_n) \xrightarrow{p} g(X)$$

(iii) If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then

$$X_n \pm Y_n \xrightarrow{p} X + Y$$

$$X_n Y_n \xrightarrow{p} X Y$$

$$\frac{X_n}{Y_n} \xrightarrow{p} \frac{X}{Y} \quad (\text{provided } P(Y=0)=0)$$

<u>Remark</u>: Approaches to verify convergence in prob

(i) Direct approach (by calculating limiting prob)

(ii) Using Chebyshev's inequality (provided 2^{nd} order moment exists)

<u>Examples</u>

(1) $X_1, \ldots, X_n$ are i.i.d. Bernoulli $(1, \theta)$; $0 < \theta < 1$

Let $Z_n = \sum_{i=1}^{n} X_i \sim B(n, \theta)$

Consider the r.v. $Y_n = \frac{Z_n}{n}$

$$P(|Y_n - \theta| > \epsilon) \leq \frac{E(Y_n - \theta)^2}{\epsilon^2}$$

$$= \frac{E\left(\frac{Z_n}{n} - \theta\right)^2}{\epsilon^2} = \frac{E(Z_n - n\theta)^2}{n^2 \cdot \epsilon^2} = \frac{V(Z_n)}{n^2 \epsilon^2}$$

$$= \frac{n\theta(1-\theta)}{n^2 \epsilon^2} \to 0 \text{ as } n \to \infty$$
$$\forall \epsilon > 0$$

$$\Rightarrow Y_n \xrightarrow{p} \theta$$

i.e. $\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{p} \theta \; (= E(X_i))$

Convergence of sequence of random variables

Modes of convergence:

Convergence in probability - to be covered in this course
Convergence in distribution - to be covered in this course
Convergence almost surely
Convergence in r^{th} mean

Convergence in probability

Let $\{X_n\}$ be a sequence of random variables on $(\Omega, \mathcal{F}, P)$.
$\{X_n\}$ is said to converge in probability to a random variable X (we write $X_n \xrightarrow{p} X$ as $n \to \infty$) if

$$P(|X_n - X| > \epsilon) \to 0 \text{ as } n \to \infty \quad \forall \epsilon > 0$$

Some important results

(i) If $X_n \xrightarrow{p} X$ and 'a' is a constant, then

$$aX_n \xrightarrow{p} aX$$

(ii) If $X_n \xrightarrow{p} X$ and $g(.)$ is any continuous function, then

$$g(X_n) \xrightarrow{p} g(X)$$

(iii) If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then

$$X_n \pm Y_n \xrightarrow{p} X \pm Y$$

$$X_n Y_n \xrightarrow{p} XY$$

$$\frac{X_n}{Y_n} \xrightarrow{p} \frac{X}{Y} \quad (\text{provided } P(Y=0)=0)$$

Remark: Approaches to verify convergence in prob

(i) Direct approach (by calculating limiting prob)

(ii) Using Chebyshev's inequality (provided 2^{nd} order moment exists)

Examples

(1) $X_1, \ldots, X_n$ are i.i.d. Bernoulli $(1, \theta)$; $0 < \theta < 1$

Let $Z_n = \sum_{i=1}^{n} X_i \sim B(n, \theta)$

Consider the r.v. $Y_n = \frac{Z_n}{n}$

$$P(|Y_n - \theta| > \epsilon) \leq \frac{E(Y_n - \theta)^2}{\epsilon^2}$$

$$= \frac{E\left(\frac{Z_n}{n} - \theta\right)^2}{\epsilon^2} = \frac{E(Z_n - n\theta)^2}{n^2 \epsilon^2} = \frac{V(Z_n)}{n^2 \epsilon^2}$$

$$= \frac{n\theta(1-\theta)}{n^2 \epsilon^2} \to 0 \text{ as } n \to \infty$$

$\forall \epsilon > 0$

$\Rightarrow Y_n \xrightarrow{p} \theta$

i.e. $\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{p} \theta \; (= E(X_i))$

(2) $X_1, \ldots, X_n$ i.i.d $U(0, \theta)$ $\theta > 0$

$Y_n = \max\{X_1, \ldots, X_n\} = X_{(n)}$

$$F_{X_{(n)}}(x) = \begin{cases} 0, & x < 0 \\ \left(\frac{x}{\theta}\right)^n, & 0 \leq x \leq \theta \\ 1, & x > \theta \end{cases}$$

$$P(|X_{(n)} - \theta| > \epsilon) = 1 - P(|X_{(n)} - \theta| \leq \epsilon)$$

$$= 1 - P(\theta - \epsilon \leq X_{(n)} \leq \theta + \epsilon)$$

$$= 1 - \left(F_{X_{(n)}}(\theta + \epsilon) - F_{X_{(n)}}(\theta - \epsilon)\right)$$

$$= 1 - \left(1 - \left(\frac{\theta-\epsilon}{\theta}\right)^n\right) \quad \forall \epsilon > 0$$

$$\to 0 \quad \text{as } n \to \infty$$

$$\Rightarrow \quad Y_n = X_{(n)} \xrightarrow{p} \theta$$ (can be proved through Chebyshev ineq also)

$$e^{X_{(n)}} \xrightarrow{p} e^{\theta}$$

$$\sqrt{X_{(n)}} \xrightarrow{p} \sqrt{\theta}$$

(3) Let $\{X_n\}$ be sequence of r.v.s with p.m.f.

$$P(X_n = 1) = \frac{1}{n} \text{ and } P(X_n = 0) = 1 - \frac{1}{n}$$

For $\epsilon > 0$,

$$P(|X_n| > \epsilon) = \begin{cases} P(X_n = 1) = \frac{1}{n}, & 0 < \epsilon < 1 \\ 0, & \epsilon \geq 1 \end{cases}$$

$$\to 0 \quad \text{as } n \to \infty \quad \forall \epsilon > 0$$

$$\Rightarrow X_n \xrightarrow{p} 0$$

Changing support example

$$P(X_n = 0) = 1 - \frac{1}{n^r} \quad \& \quad P(X_n = n) = \frac{1}{n^r} \quad r > 0$$

$$P(|X_n| > \epsilon) = \begin{cases} P(X_n = n), & 0 < \epsilon < n \\ 0, & \epsilon \geq n \end{cases}$$

$$\to 0 \quad \text{as } n \to \infty$$

$$X_n \xrightarrow{p} 0$$

(4) $X_1, \ldots, X_n$ be i.i.d. with p.d.f. (or p.m.f) f_X

X_i with mean μ and variance $\sigma^2 (< \infty)$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad : \text{ Sample mean r.v.}$$

$$E\bar{X}_n = \frac{1}{n}\sum_i EX_i = \mu$$

$$V(\bar{X}_n) = E(\bar{X}_n - \mu)^2$$
$$= E\left(\frac{1}{n}\sum_{1}^{n} X_i - \mu\right)^2$$
$$= E\left(\frac{1}{n}\sum (X_i - \mu)\right)^2$$
$$= \frac{1}{n^2} E\left(\sum (X_i - \mu)\right)^2$$
$$= \frac{1}{n^2}\sum_{i=1}^{n} E(X_i - \mu)^2 \quad (\because X_1, \ldots X_n \text{ are indep})$$
$$= \frac{\sigma^2}{n}$$

$$P(|\bar{X}_n - \mu| > \epsilon) \le \frac{E(\bar{X}_n - \mu)^2}{\epsilon^2} = \frac{V(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty$$

$$\Rightarrow \bar{X}_n \xrightarrow{p} \mu$$

Remark: This convergence is irrespective of the underlying distⁿ.

Weak Law of Large Numbers (WLLN)

Defⁿ: Let $\{X_n\}$ be a seq of r.v.s. We say that $\{X_n\}$ satisfies WLLN if $\exists$ constants $\{a_n\}$ and $\{b_n\}$ where $b_n > 0$ and $b_n \uparrow \infty \ni$

$$\frac{S_n - a_n}{b_n} \xrightarrow{p} 0 \quad \text{as } n \to \infty$$

where, $S_n = \sum_{i=1}^{n} X_i$

Khintchine's WLLN

If $\{X_n\}$ is $\ni X_1, X_2, \ldots$ is i.i.d. seq of r.v.s with $E|X_i| < \infty$, then WLLN holds and

$$\frac{1}{n}\sum X_i \xrightarrow{p} \mu = E(X_1) \quad \left(\begin{array}{c} S_n = \sum_1^n X_i \\ a_n = n\mu \\ b_n = n \end{array}\right)$$

Remark: Khintchine's WLLN does not require existence of 2nd moment

Applications

(i) $X_1, \ldots, X_n$ i.i.d. $B(1,\theta)$; θ unknown $0<\theta<1$

$E X_i$ exists ; $E X_i = \theta$

Khintchine's WLLN $\Rightarrow \frac{1}{n}\sum X_i \xrightarrow{p} \theta$ as $n \to \infty$

(ii) $X_1, \ldots, X_n$ i.i.d. with p.d.f./p.m.f. f_X having mean θ and variance σ^2

by WLLN sample mean $\frac{1}{n}\sum_1^n X_i \xrightarrow{p} \theta$ (popn mean)

Further, let

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

$$= \frac{1}{n-1}\left(\sum X_i^2 - n\bar{X}_n^2\right)$$

$$= \frac{1}{n-1}\sum X_i^2 - \frac{n}{n-1}\bar{X}_n^2$$

Now $\bar{X}_n \xrightarrow{p} \theta \Rightarrow \bar{X}_n^2 \xrightarrow{p} \theta^2$

Note that $X_1, \ldots, X_n$ i.i.d. with mean θ & var σ^2

$\Rightarrow X_1^2, \ldots, X_n^2$ i.i.d. with mean $(\sigma^2+\theta^2) < \infty$

by WLLN

$$\frac{1}{n}\sum_1^n X_i^2 \xrightarrow{p} E X_1^2 = \sigma^2 + \theta^2$$

$$\Rightarrow \frac{1}{n-1}\sum_1^n X_i^2 = \frac{n}{n-1}\,\frac{1}{n}\sum X_i^2 \xrightarrow{p} \theta^2 + \sigma^2$$

$$\Rightarrow S^2 = \frac{1}{n-1}\sum X_i^2 - \frac{n}{n-1}\bar{X}_n^2 \xrightarrow{p} (\theta^2+\sigma^2) - \theta^2 = \sigma^2$$

$\Rightarrow \quad S^2 \xrightarrow{p} \sigma^2$; i.e. sample variance random variable converges in prob to corresponding population variance.

<u>Note</u>: $$s_n^2 = \frac{1}{n}\sum(x_i-\bar{x})^2 = \frac{(n-1)s^2}{n} \xrightarrow{p} \sigma^2$$

<u>Remark</u>: WLLN for non i.i.d. setup

Suppose $X_1, X_2, \ldots$ be a seq of <u>uncorrelated</u> r.v.s with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$; $i = 1, 2, \ldots$

If $\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 \to 0$ as $n \to \infty$, then WLLN holds for $\{X_n\}$.

Take $a_n = \sum_{i=1}^{n}\mu_i$ & $b_n = n$,

then

$$P\left(\left|\frac{S_n - a_n}{b_n}\right| > \epsilon\right) = P\left(\left|\frac{1}{n}\sum_1^n (X_i - \mu_i)\right| > \epsilon\right)$$

$$\leq \frac{E\left(\frac{1}{n}\sum(X_i-\mu_i)\right)^2}{\epsilon^2} = \frac{E\left(\sum(X_i-\mu_i)\right)^2}{n^2\epsilon^2}$$

$$= \frac{\sum_1^n \sigma_i^2}{n^2\epsilon^2} \quad (\because \text{ of uncorrelatedness})$$

$$\to 0 \text{ as } n \to \infty$$

$$\Rightarrow \frac{S_n - a_n}{b_n} \xrightarrow{p} 0$$

i.e. WLLN holds for $\{X_n\}$.

Remark: under the same setup

i.e. $X_1, X_2, \ldots$ uncorrelated with $E X_i = \mu_i$

and $V(X_i) = \sigma_i^2$

suppose $\sum_1^n \sigma_i^2 \to \infty$, then we can take

$$a_n = \sum_{i=1}^n \mu_i \quad \& \quad b_n = \sum_1^n \sigma_i^2$$

with $\frac{S_n - a_n}{b_n} \xrightarrow{p} 0$

Remark: If $\sigma_i^2 = \sigma^2 \ \forall i$, then the condition

$\frac{1}{n^2} \sum \sigma_i^2 \to 0$ as $n \to \infty$ is automatically

satisfies and WLLN holds

Remark: If $X_1, \ldots$ is a seq of i.i.d. r.v.s with

mean μ and variance $\sigma^2 < \infty$, then

WLLN holds for $\{X_n\}$ &

$$\bar{X}_n \xrightarrow{p} \mu = E X_1$$

Note that finiteness of variance is not reqd by

'Khintchine's WLLN.

Convergence in distribution (or law)

Defⁿ: We say that a seq of r.v.s $\{X_n\}$ converges in distribution to X ($X_n \xrightarrow{L} X$ as $n \to \infty$)

if $F_{X_n}(x) \longrightarrow F_X(x)$ $\forall x$ at which the limiting distⁿ fⁿ is continuous.

$F_{X_n}(\cdot)$: d.f. of X_n

$F_X(\cdot)$: d.f. of X

Examples

(1) $X_1, \ldots, X_n$ i.i.d. $N(0,1)$

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

$$Y_n \sim N\left(0, \frac{1}{n}\right) \; ; \; Y_n \xrightarrow{P} 0.$$

$$F_{Y_n}(y) = P(Y_n \le y) = P(\sqrt{n}\, Y_n \le \sqrt{n}\, y)$$

$$= \Phi(\sqrt{n}\, y) \longrightarrow \begin{cases} 0, & y < 0 \\ \frac{1}{2}, & y = 0 \\ 1, & y > 0 \end{cases}$$

Consider a d.f. $F_Y(y) = \begin{cases} 0, & y < 0 \\ 1, & y \ge 0 \end{cases}$

i.e. $P(Y = 0) = 1$

then

$$F_{Y_n}(y) \to F_Y(y). \quad \forall\, y \text{ at which } F_Y(\cdot) \text{ is continuous}$$

$$\Rightarrow \quad Y_n \xrightarrow{L} Y \ (\leftarrow \text{ a degenerate r.v.})$$

i.e. $Y_n \xrightarrow{L} 0$

(2) $X_1, X_2 \cdots$ i.i.d. $U(0, \theta)$; $\theta > 0$

$$Y_n = X_{(n)} = \max\{X_1, \ldots, X_n\} \quad ; \quad Y_n \xrightarrow{P} \theta$$

$$F_{Y_n}(y) = \begin{cases} 0, & y < 0 \\ \left(\frac{y}{\theta}\right)^n, & 0 \le y < \theta \\ 1, & y \ge \theta \end{cases} \longrightarrow \begin{cases} 0, & y < \theta \\ 1, & y \ge \theta \end{cases}$$

$\Rightarrow \quad Y_n \xrightarrow{L} Y$, a degenerate r.v.; degenerate at θ

i.e. $Y_n \xrightarrow{L} \theta$

Consider $\quad Z_n = n(\theta - X_{(n)}) = n(\theta - Y_n)$

$$\begin{aligned} F_{Z_n}(x) &= P(Z_n \le x) \\ &= P(n(\theta - X_{(n)}) \le x) \\ &= P\left(X_{(n)} \ge \theta - \frac{x}{n}\right) = 1 - F_{X_{(n)}}\left(\theta - \frac{x}{n}\right) \\ &= \begin{cases} 0, & x < 0 \\ 1 - \left(\frac{\theta - x/n}{\theta}\right)^n & 0 \le x < n\theta \\ 1, & x \ge n\theta \end{cases} \end{aligned}$$

$$F_{Z_n}(x) \to \begin{cases} 0, & x < 0 \\ 1 - e^{-x/\theta}, & x \geq 0 \end{cases}$$

i.e. $Z_n \xrightarrow{L} Z$; where $f_Z(z) = \begin{cases} 0, & z < 0 \\ \frac{1}{\theta} e^{-z/\theta}, & z \geq 0 \end{cases}$

i.e. $Z_n \xrightarrow{L} Z \sim \exp(0, \theta)$

scale θ.

(3) $\{X_n\}$ - seq of discrete r.v.s.

$$P(X_n = x) = \begin{cases} 1, & \text{if } x = 2 + \frac{1}{n} \\ 0, & \text{o/w} \end{cases}$$

$$F_{X_n}(x) = \begin{cases} 0, & x < 2 + \frac{1}{n} \\ 1, & x \geq 2 + \frac{1}{n} \end{cases}$$

$$\to F_X(x) = \begin{cases} 0, & x < 2 \\ 1, & x \geq 2 \end{cases}$$

$$X_n \xrightarrow{L} 2$$

<u>Remark</u>: Ex 1, 2, 3 are direct approaches to prove convergence in law/distribution.

Convergence in law can also be proved using m.g.f. convergence.

(4) $X_n \sim Bin(n, \theta)$

Suppose $n \to \infty$ $\ni$ $np = \lambda$ is fixed i.e. $\theta = \frac{\lambda}{n}$

$$M_{X_n}(t) = \left((1-\theta) + \theta e^t\right)^n$$

$$= \left(1 + \frac{\lambda}{n}(e^t - 1)\right)^n$$

$$\longrightarrow e^{\lambda \cancel{\theta}(e^t-1)} \quad \text{as } n \to \infty$$

$\Rightarrow X_n \xrightarrow{L} X$; where $X \sim P(\lambda)$

(5) $X_1, \dots$ i.i.d. $N(0,1)$

$\bar{X}_n \sim N(0, \frac{1}{n})$

$M_{\bar{X}_n}(t) = e^{\frac{t^2}{2n}} \to 1$ as $n \to \infty$

$\Rightarrow \bar{X}_n \xrightarrow{L} X$; where X is degenerate at 0

<u>Remark</u>: Ex 1, 2, 3 are direct approaches to prove convergence in law/distribution.

Convergence in law can also be proved using m.g.f. convergence.

(4) $X_n \sim \text{Bin}(n, \theta)$

Suppose $n \to \infty \ni n\theta = \lambda$ is fixed i.e. $\theta = \frac{\lambda}{n}$

$$M_{X_n}(t) = \left((1-\theta) + \theta e^t\right)^n$$

$$= \left(1 + \frac{\lambda}{n}(e^t - 1)\right)^n$$

$$\longrightarrow e^{\lambda\,\cancel{\theta}(e^t-1)} \quad \text{as } n \to \infty$$

$\Rightarrow X_n \xrightarrow{L} X$; where $X \sim P(\lambda)$

(5) $X_1, \ldots$ i.i.d. $N(0,1)$

$\bar{X}_n \sim N(0, \frac{1}{n})$

$M_{\bar{X}_n}(t) = e^{\frac{t^2}{2n}} \rightarrow 1$ as $n \to \infty$

$\Rightarrow \bar{X}_n \xrightarrow{L} X$; where X is degenerate at 0

(6) $X_n \sim \chi^2_n$; $E X_n = n$; $V X_n = 2n$

$M_{X_n}(t) = (1-2t)^{-n/2}$ $\quad t < \frac{1}{2}$

$$Y_n = \frac{X_n - n}{\sqrt{2n}}$$

$$M_{Y_n}(t) = E\left(e^{t\left(\frac{X_n - n}{\sqrt{2n}}\right)}\right)$$

$$= e^{-\frac{tn}{\sqrt{2n}}} E\left(e^{\frac{tX_n}{\sqrt{2n}}}\right)$$

$$= e^{-\frac{tn}{\sqrt{2n}}} \left(1 - \frac{2t}{\sqrt{2n}}\right)^{-n/2} \qquad t < \sqrt{\frac{n}{2}}$$

$$= \left(e^{t\sqrt{\frac{2}{n}}}\right)^{-n/2} \left(1 - \sqrt{\frac{2}{n}}\, t\right)^{-n/2}$$

$$= \left[\left(1 + t\sqrt{\frac{2}{n}} + \frac{t^2\, 2/n}{2!} + \frac{t^3\, 2^{3/2}/n^{3/2}}{3!} \ldots\right) - t\sqrt{\frac{2}{n}}\left(1 + t\sqrt{\frac{2}{n}} + \frac{t^2\, 2/n}{2!} + \ldots\right)\right]^{-n/2}$$

$$= \left(1 - \frac{t^2}{n} + \frac{k_1}{n^{3/2}} - \ldots\right)^{-n/2}$$

$$= \left(1 - \frac{t^2}{n} + o\left(\frac{1}{n}\right)\right)^{-n/2}$$

$$\rightarrow e^{t^2/2} \leftarrow \text{m.g.f. of } N(0,1)$$

$$\Rightarrow \frac{X_n - n}{\sqrt{2n}} \xrightarrow{L} N(0,1) \text{ r.v.}$$

<u>Some important results</u>

(i) If $X_n \xrightarrow{p} X$ then $X_n \xrightarrow{L} X$

Converse is not true in general.

However, if $X_n \xrightarrow{L} c$ (a const), then

$$X_n \xrightarrow{p} c$$

(ii) <u>Slutsky's Lemma</u>

If $X_n \xrightarrow{L} X$ and $Y_n \xrightarrow{p} c$, then

(i) $X_n \pm Y_n \xrightarrow{L} X \pm c$

(ii) $X_n Y_n \xrightarrow{L} cX$

(iii) $\frac{X_n}{Y_n} \xrightarrow{L} X/c$ (for $c \neq 0$)

(iii) <u>Δ-method or Δ-rule</u>

Let $\{X_n\}$ be a seq of random variables $\ni$

$$\sqrt{n}(X_n - \theta) \xrightarrow{L} N(0, \sigma^2)$$

Suppose g be real valued f^n differentiable at $\theta \ni$ $g'(\theta) \neq 0$, then

$$\sqrt{n}\left(g(X_n) - g(\theta)\right) \xrightarrow{L} N(0, (g'(\theta))^2 \sigma^2)$$

Central Limit Theorem (CLT)

In WLLN, we investigated

$$\frac{S_n - a_n}{b_n} \xrightarrow{p} 0$$

i.e convergence of $\frac{S_n - a_n}{b_n}$ to a degenerate distⁿ (degenerate at 0)

In CLT, we investigate convergence of $\frac{S_n - a_n}{b_n}$ to a non-degenerate distⁿ.

Defⁿ: Let $X_1, X_2, \ldots$ be a seq of i.i.d. r.v.s with common d.f F. We say that F belongs to the domain of attraction of a distⁿ V if there exists centering constants $\{A_n\}$ and norming constants $\{B_n\}$ $(B_n > 0)$ $\ni$

as $n \to \infty$

$$P\left(\frac{S_n - A_n}{B_n} \le x\right) \to V(x) \quad \text{at all continuity points of } V$$

$$S_n = \sum_{i=1}^{n} X_i$$

i.e. $\frac{S_n - A_n}{B_n} \xrightarrow{L} X$; where d.f. of X is $V(\cdot)$

Lindeberg-Levy CLT

Let $\{X_n\}$ be a seq of i.i.d. r.v.s with $E(X_i) = \mu$ and $V(X_i) = \sigma^2 < \infty$, then for $S_n = \sum_{i=1}^{n} X_i$

$$\frac{S_n - ES_n}{\sqrt{V(S_n)}} \xrightarrow{L} N(0,1) \text{ r.v.}$$

i.e. $$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \xrightarrow{L} X \quad ; \quad X \sim N(0,1)$$

i.e. $$\frac{n\bar{X}_n - n\mu}{\sqrt{n}\,\sigma} \xrightarrow{L} X \quad ; \quad X \sim N(0,1)$$

i.e. $$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{L} X ; \quad X \sim N(0,1)$$

i.e. $$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} Y ; \quad Y \sim N(0,\sigma^2)$$

Applications of CLT

(1) $X_1, \dots X_n$ are i.i.d. $Exp(\theta)$ (exponential with mean θ)

i.e. $f_X(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x > 0 \\ 0, & o/w \end{cases}$

$E(X_i) = \theta \quad ; \quad V(X_i) = \theta^2$

By CLT

$$\frac{\sqrt{n}(\bar{X}_n - \theta)}{\theta} \xrightarrow{L} N(0,1)$$

i.e. $$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{L} N(0,\theta^2)$$

Further, suppose we are interested in asymptotic dist^n $(n\to\infty)$ of $\frac{1}{\bar{X}_n}$; we can apply Δ-rule on the CLT result with

$$g(x) = \frac{1}{x} \qquad g'(x) = -\frac{1}{x^2} \qquad g'(\theta) \neq 0$$

Δ-rule $\Rightarrow$ $$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\theta}\right) \xrightarrow{L} N\Big(0, \underbrace{\left(\frac{1}{\theta^4}\right)}_{(g'(\theta))^2} \cdot \theta^2\Big)$$

i.e. $$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\theta}\right) \xrightarrow{L} N\left(0, \frac{1}{\theta^2}\right)$$

(2) Let $X_1, \ldots X_n$ be i.i.d. $U(0,1)$

suppose, we are interested in asymptotic distⁿ $(n \to \infty)$ of

$$G_n = \left(\prod_{i=1}^{n} X_i \right)^{1/n}$$

Note that $H_n = -\log G_n = \frac{1}{n} \sum_{i=1}^{n} (-\log X_i)$

$$G_n = e^{-H_n}$$

Realize that $X_i \sim U(0,1)$

$$Y_i = -\log X_i \sim exp(1) \; ; \; E Y_i = 1, \; V Y_i = 1$$

<u>By CLT</u> $\sqrt{n}\,(\bar{Y}_n - 1) \xrightarrow{L} N(0,1)$

i.e. $\sqrt{n}\,(H_n - 1) \xrightarrow{L} N(0,1)$

Take $g(x) = e^{-x}$ $\quad g'(x) = -e^{-x}$ $\quad g'(1) = -e^{-1} \neq 0$

<u>Δ-rule</u> $\Rightarrow \sqrt{n}\,(e^{-H_n} - e^{-1}) \xrightarrow{L} N(0, e^{-2})$

i.e. $\sqrt{n}\,(G_n - e^{-1}) \xrightarrow{L} N(0, e^{-2})$

i.e. $\sqrt{n}\left((\prod X_i)^{1/n} - e^{-1} \right) \xrightarrow{L} N(0, e^{-2})$

(3) $X_1, \ldots X_n$ i.i.d χ^2_m

$$Y_n = \sum_{i=1}^{n} X_i \sim \chi^2_{nm}$$

suppose we want to find approximate value of

$P(a < Y_n < b)$ for large n $(a < b)$

By CLT $\quad \dfrac{Y_n - nm}{\sqrt{2nm}} \xrightarrow{L} N(0,1)$

$$P(a < Y_n < b) = P\left(\frac{a-nm}{\sqrt{2nm}} < \frac{Y_n - nm}{\sqrt{2nm}} < \frac{b-nm}{\sqrt{2nm}}\right)$$

$$\underset{\text{for large } n}{\simeq} \Phi\left(\frac{b-nm}{\sqrt{2nm}}\right) - \Phi\left(\frac{a-nm}{\sqrt{2nm}}\right)$$

(4) $X_1, \ldots X_n$ i.i.d $P(\lambda)$

$$S_n = \sum_1^n X_i \sim P(n\lambda) \qquad E S_n = n\lambda \quad V S_n = n\lambda$$

By CLT $\quad \dfrac{S_n - n\lambda}{\sqrt{n\lambda}} \xrightarrow{\mathcal{L}} N(0,1)$

Take $n = 64$, $\lambda = 0.125$; $n\lambda = 8$

$P(S_n = 10) = 0.099 \rightarrow$ from Poisson p.m.f.

Approximate value of the above can be obtained thro CLT

$P(S_n = 10) = P(9.5 < S_n < 10.5) \leftarrow$ Continuity correction

$$\underset{\text{for large } n}{\simeq} \Phi\left(\frac{10.5 - \overset{n\lambda}{8}}{\sqrt{8}}\right) - \Phi\left(\frac{9.5-8}{\sqrt{8}}\right) = .108$$

for $n = 96$; $\lambda = 0.125$, $n\lambda = 12$

$P(S_n = 10) = .105$ (exact)

$P(S_n = 10) \simeq .101$ (thro normal approximation)

Higher the value of n, closer will be the approximation.

Statistical Inference

Let X be a random variable describing a characteristics

We assume an "appropriate" prob model for X, in the sense that we say X has a prob distⁿ with p.d.f. or p.m.f $f_\theta(x)$; where θ is an unknown parameter (or a parameter vector) characterizing the distⁿ of the r.v. X

e.g. $X \sim N(\mu, \sigma^2)$ $\quad \underset{\sim}{\theta} = (\mu, \sigma)'; \ \mu \in \mathbb{R}, \sigma > 0$

$X \sim B(n, \theta)$ $\quad 0 < \theta < 1$

$X \sim exp(\theta)$, $\quad \theta > 0$

$X \sim P(\theta)$, $\quad \theta > 0$

In each of these cases θ is unknown and is assumed to vary in a space called parameter space, Θ, say

e.g. $X \sim N(\mu, \sigma^2)$; $\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$

$X \sim exp(\theta)$; $\Theta = \{\theta : \theta > 0\}$

Family of distⁿs:

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}.$$

P_θ is a distⁿ characterised by θ.
p.m.f. or p.d.f. of P_θ is $f_\theta(x)$ say

Inference problem: to make inference about the unknown parameter(s).

Random Sample from P_θ :

Let $X_1, \ldots, X_n$ be a random sample from P_θ

i.e. $X_1, \ldots, X_n$ are i.i.d. with p.d.f. (or p.m.f) $f_\theta(x)$.

We use the random sample $(X_1, \ldots, X_n)$ for inference about θ.

3 approaches of statistical inference

(i) **Point estimation**

$\delta(X_1, \ldots, X_n)$ - estimator (f^n of r.v.s); a statistic

$\delta(\underset{\sim}{X}) = \hat{\theta}(\underset{\sim}{X})$; e.g. Maximum likelihood estimator (MLE)
Bayes estimator, unbiased estimator
Uniformly minimum variance unbiased estimator (UMVUE),
Least Squares estimator (LSE)
- - -

~~For~~

For an observed sample
$(x_1, \ldots, x_n) \longrightarrow \delta(\underset{\sim}{x})$; an estimate

(ii) **Interval estimation** : Confidence interval

Construct a random interval

$[\hat{\theta}_L(\underset{\sim}{X}), \hat{\theta}_U(\underset{\sim}{X})] \ni$

$P(\theta \in [\hat{\theta}_L(\underset{\sim}{X}), \hat{\theta}_U(\underset{\sim}{X})]) \geq 1 - \alpha$; say $\alpha = 0.05$ or 0.01

Estimated interval from observed sample $(x_1, \ldots, x_n)$

$[\hat{\theta}_L(\underset{\sim}{x}), \hat{\theta}_U(\underset{\sim}{x})]$

(iii) Hypothesis testing

Validate / test some prior belief about parameter

$H_0 : \theta = \theta_0$ against $H_A : \theta = \theta_1$

null hypothesis

θ_0 & θ_1 known consts

alternate hypothesis

or $H_0 : \theta = \theta_0$ against $H_A : \theta > \theta_0$ or $H_A : \theta < \theta_0$

or $H_0 : \theta = \theta_0$ against $H_A : \theta \neq \theta_0$.

Point Estimation

$X_1, \dots X_n$ random sample from P_θ with p.d.f. (or p.m.f.) f_θ

$\theta \in \Theta$

$g(\theta)$: parametric f^n of ~~the~~ interest, called estimand

$\delta(X_1, \dots, X_n)$: Estimator (a f^n of r.v.s $X_1, \dots, X_n$)

Def^n: Unbiased estimator

$\delta(\underline{X})$ is an unbiased estimator for $g(\theta)$ if

$$E\,\delta(\underline{X}) = g(\theta) \quad \forall\ \theta \in \Theta$$

Examples

(i) $N(\theta, 1)$ pop^n $\theta \in \mathbb{R}$; $g(\theta) = \theta$

$X_1, \dots, X_n$ random sample

$$\left.\begin{aligned} \delta_1(\underline{X}) &= \bar{X} \\ \delta_2(\underline{X}) &= X_1 \\ \delta_3(\underline{X}) &= \frac{X_1 + X_2}{2} \\ \delta_4(\underline{X}) &= \sum_{i=1}^{n} a_i X_i \ni \sum a_i = 1 \end{aligned}\right\}$$

$E\,\delta_i(\underline{X}) = \theta$

$\delta_1, \delta_2, \delta_3, \delta_4$ are all unbiased estimator of θ

(ii) $U(0, \theta)$ pop^n $\theta > 0$ $g(\theta) = \theta$

$X_1, \dots X_n$ random sample

$$\left.\begin{aligned} \delta_1(\underline{X}) &= 2X_1 \\ \delta_2(\underline{X}) &= X_1 + X_2 \\ \delta_3(\underline{X}) &= 2\bar{X} \\ \delta_4(\underline{X}) &= \frac{n+1}{n} X_{(n)} \end{aligned}\right\}$$

all are unbiased estimator for θ

(iii) $B(1, \theta)$ pop^n $0 < \theta < 1$.

$g(\theta) = \theta$ $\delta_i(\underline{X}) = X_i$ $i = 1(1)n$

$\delta_2(\underline{X}) = \frac{X_1 + X_2}{2}$; $\delta_3(\underline{X}) = \sum_1^n X_i / n$

all are u.e. for θ

$g(\theta) = \theta(1-\theta)$

$$\delta(\underline{X}) = \begin{cases} 1, & X_1 = 1,\ X_2 = 0 \\ 0 & \text{o/w} \end{cases}$$

$\delta(\underline{X})$ is u.e. of $\theta(1-\theta)$

Sufficient statistic

Data dimension reduction without loss of information

$X_1, \ldots, X_n$ random sample from a $dist^n$ with p.d.f. or p.m.f. $f_\theta(x)$

$T(\underset{\sim}{X})$: statistic

$T(\underset{\sim}{X})$ is sufficient for θ if $T(\underline{X})$ contains all information about θ, that is contained in the entire sample $(X_1, \ldots, X_n)$; i.e. given $T(\underline{X})$, $(X_1, \ldots X_n)$ does not contain any information about θ.

<u>Def^n</u>: A statistic $T(\underline{X})$ is said to be sufficient for θ if the conditional $dist^n$ of $(X_1, \ldots, X_n)$ given $T=t$ is independent of θ.

<u>Example</u>:

(i) $X_1, \ldots, X_n$ i.i.d. random sample from $B(1, \theta)$

Claim: $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is sufficient for θ

$$\sum_{i=1}^{n} X_i \sim B(n, \theta)$$

$$P(X_1 = x_1, \ldots, X_n = x_n \mid T = t)$$

$$= \frac{P(X_1 = x_1, \ldots, X_n = x_n \,;\, T = t)}{P(T = t)} = 0 \text{ if } \sum_{i=1}^{n} x_i \neq t$$

indep of θ

otherwise if $\sum_{i=1}^{n} x_i = t$

$$= \frac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(T = t)}$$

$$= \frac{P(X_1 = x_1) \cdots P(X_n = x_n)}{P(T = t)} \quad \text{(independence)}$$

$$= \frac{\left(\theta^{x_1}(1-\theta)^{1-x_1}\right) \cdots \left(\theta^{x_n}(1-\theta)^{1-x_n}\right)}{\binom{n}{t} \theta^{t} (1-\theta)^{n-t}}$$

$$= \frac{1}{\binom{n}{t}} \leftarrow \text{indep of } \theta$$

$\Rightarrow T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is sufficient for θ

Example

(ii) $X_1, \ldots, X_n$ i.i.d. random sample from $P(\theta)$; $\theta > 0$

Claim: $T(\underline{X}) = \sum_{i=1}^{n} X_i$ is sufficient for θ

$$P(X_1 = x_1, \ldots, X_n = x_n \mid T = t) \qquad T \sim P(n\theta)$$

$$= \frac{P(X_1 = x_1, \ldots, X_n = x_n \,;\, T = t)}{P(T = t)} = 0 \text{ if } \sum_{i=1}^{n} x_i \neq t$$

indep of θ

otherwise if $\sum_{i=1}^{n} x_i = t$

$$= \frac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(T = t)}$$

$$= \frac{P(X_1 = x_1) \cdots P(X_n = x_n)}{P(T = t)} \quad \text{(independence)}$$

$$= \frac{\cancel{P(X_1 = x_1)} - \ldots}{}$$

$$= \frac{\left(e^{-\theta}\theta^{x_1}/x_1!\right) \cdots \left(e^{-\theta}\theta^{x_n}/x_n!\right)}{e^{-n\theta}(n\theta)^t/t!}$$

$$= \frac{t!}{\prod_{i=1}^{n} x_i!} \cdot \frac{1}{n^t} \rightarrow \text{indep of } \theta$$

$$\Rightarrow T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i \quad \text{is sufficient for } \theta$$

<u>Remark</u>: The above defn of sufficient statistic is not a constructive definition.

<u>Neyman-Fisher Factorization Theorem</u>

$X_1, \ldots, X_n$ be a random sample with p.d.f. or p.m.f. $f_\theta(x)$ $\theta \in \Theta$. A statistic $T(\underset{\sim}{X})$ is sufficient for θ iff $f_\theta(\underset{\sim}{x})$ can be factored as

$$f_\theta(\underset{\sim}{x}) = h(\underset{\sim}{x})\, g_\theta(t(\underset{\sim}{x}))$$

where, $h(\underset{\sim}{x}) > 0$ is a fn of $(x_1, \ldots, x_n)$ only and indep of θ

and $g_\theta(t(\underset{\sim}{x}))$: fn of θ and depends on $(x_1, \ldots, x_n)$ only through $t(x_1, \ldots, x_n)$.

<u>Remark</u>: Every 1-1 fn of a sufficient statistic $T(\underset{\sim}{X})$ is also ~~of~~ a sufficient statistic.

<u>Remark</u>: T & T^* be 2 statistic $\ni T = \psi(T^*)$

T is suff $\Rightarrow T^*$ is also sufficient statistic

Examples

(1) $X_1, \dots X_n$ r.s. from $N(\theta, 1)$ $\quad \theta \in \mathbb{R}$

$$f_\theta(\underset{\sim}{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_1^n (x_i-\theta)^2} \qquad -\infty < x_1, \dots, x_n < \infty$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\left(\Sigma x_i^2 + n\theta^2 - 2\theta\Sigma x_i\right)\right)$$

$$= \underbrace{\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\Sigma x_i^2\right)\right)}_{h(\underset{\sim}{x})} \underbrace{\left(\exp\left(-\frac{n}{2}\theta^2 + \theta\sum_1^n x_i\right)\right)}_{g_\theta(\Sigma x_i)}$$

By NFFT, $T(\underset{\sim}{X}) = \sum_{i=1}^n X_i$ is sufficient for θ

$\bar{X}$ is also sufficient for θ

$(X_1, \dots, X_n)$ is suff for θ (it's the trivial suff stat)

$(X_1, \sum_{i=2}^n X_i)$ is also suff for θ

(2) $X_1, \dots, X_n$ r.s. from $P(\theta)$

$$f_\theta(\underset{\sim}{x}) = \prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \frac{e^{-n\theta}\,\theta^{\sum_1^n x_i}}{\prod_{i=1}^n x_i!}$$

i.e. $f_\theta(\underset{\sim}{x}) = \left(\frac{1}{\prod_1^n x_i!}\right)\left(e^{-n\theta}\,\theta^{\sum_1^n x_i}\right)$

$$= h(\underset{\sim}{x}) \; g_\theta\left(\sum_1^n x_i\right)$$

By NFFT, $T(\underset{\sim}{X}) = \sum_1^n X_i$ is sufficient for θ

(3) $X_1, \ldots, X_n$ random sample from $U(0,\theta)$, $\theta > 0$

jt p.d.f.

$$f_\theta(\underline{x}) = \begin{cases} \frac{1}{\theta^n}, & 0 < x_1, \ldots, x_n < \theta \\ 0, & o/w \end{cases}$$

i.e. $$f_\theta(\underline{x}) = \begin{cases} \frac{1}{\theta^n}, & 0 < x_{(1)} < \cdots < x_{(n)} < \theta \\ 0, & o/w \end{cases}$$

i.e. $$f_\theta(\underline{x}) = \frac{1}{\theta^n} I(0, x_{(1)}) \, I(x_{(n)}, \theta)$$

where $I(a,b) = \begin{cases} 1, & a < b \\ 0, & o/w \end{cases}$

$$f_\theta(\underline{x}) = \underbrace{\left(I(0, x_{(1)}) \right)}_{h(\underline{x})} \left(g_\theta(x_{(n)}) \right)$$

$$g_\theta(x_{(n)}) = \frac{1}{\theta^n} I(x_{(n)}, \theta)$$

By NFFT, $T(\underline{X}) = X_{(n)}$ is suff for θ

(4) $X_1, \ldots X_n$ random sample from $N(\mu, \sigma^2)$

$\mu \in \mathbb{R}, \ \sigma > 0$

$$\underset{\sim}{\theta} = (\mu, \sigma) \in \Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$$

$$f_{\underset{\sim}{\theta}}(\underline{x}) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2} \Sigma (x_i - \mu)^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\left(\Sigma x_i^2 + n\mu^2 - 2\mu \Sigma x_i\right)\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^n} \exp\left(-\frac{\Sigma x_i^2}{2\sigma^2} - \frac{n}{2}\frac{\mu^2}{\sigma^2} + \frac{\mu}{\sigma^2}\Sigma x_i\right)\right)$$

$$= h(\underset{\sim}{x}) \quad g_{\underset{\sim}{\theta}}(\Sigma x_i, \Sigma x_i^2)$$

$T(\underset{\sim}{X}) = \left(\sum_1^n X_i, \sum_1^n X_i^2\right)$ is jointly sufficients for $\underset{\sim}{\theta}$

(5) $X_1, \ldots X_n$ is a random sample from $N(\theta, \theta^2)$ $\theta > 0$

jt p.d.f.

$$f_\theta(\underline{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\theta} \exp\left(-\frac{1}{2\theta^2}[x_i - \theta)^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\theta^n} \exp\left(-\frac{1}{2\theta^2}\left(\Sigma x_i^2 + n\theta^2 - 2\theta \Sigma x_i\right)\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\theta^n} e^{-\frac{\Sigma x_i^2}{2\theta^2}} e^{-\frac{n}{2}} e^{\frac{\Sigma x_i}{\theta}}$$

$$= \underbrace{\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-n/2}\right)}_{h(\underline{x})} \underbrace{\left(\frac{1}{\theta^n} e^{-\frac{\Sigma x_i^2}{2\theta^2}} e^{\frac{\Sigma x_i}{\theta}}\right)}_{g_\theta(\Sigma x_i,\, \Sigma x_i^2)}$$

By NFFT, $T(\underline{X}) = \left(\sum_{i=1}^{n} X_i, \Sigma X_i^2\right)$ is jointly sufficient for θ

(6) $X_1, \ldots, X_n$ is a random sample from $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ $\theta \in \mathbb{R}$

jt p.d.f.

$$f_\theta(\underline{x}) = \begin{cases} 1, & \theta - \frac{1}{2} < x_1, \ldots, x_n < \theta + \frac{1}{2} \\ 0, & o/w \end{cases}$$

$$= \begin{cases} 1, & \theta - \frac{1}{2} < x_{(1)} - \cdots < x_{(n)} < \theta + \frac{1}{2} \\ 0, & o/w \end{cases}$$

i.e. $f_\theta(\underset{\sim}{x}) = I(\theta - \frac{1}{2}, x_{(1)})\, I(x_{(n)}, \theta + \frac{1}{2})$

By NFFT, $T(\underline{X}) = (X_{(1)}, X_{(n)})$ is ~~..~~ jointly sufficient for θ

$h(\underline{x}) = 1$

$g_\theta(x_{(1)}, x_{(n)}) = I(\theta - \frac{1}{2}, x_{(1)})\, I(x_{(n)}, \theta + \frac{1}{2})$

<u>Remark</u>: Any 1-1 f^n of sufficient statistic is sufficient

<u>Remark</u>: Any f^n of sufficient statistic is NOT necessarily a sufficient statistic

<u>Remark</u>: The factorization (as in NFFT) is not unique.

<u>Remark</u>: <u>sufficient statistic is not unique</u>. We look for the sufficient statistic that provides the maximum possible reduction of the data, this leads us to the concept of minimal sufficient statistic

<u>Ex</u>: $X_1, \ldots X_n$ r.s. from $N(\theta, 1)$ $\theta \in \mathbb{R}$

By NFFT, all the following statistics are sufficient for θ.

$T_1(\underline{X}) = (X_1, \ldots, X_n)$

$T_2(\underline{X}) = (X_1 + X_2, X_3, X_4, \ldots, X_n)$

$T_3(\underline{X}) = (X_1 + X_2, X_3 + X_4, X_5, \ldots, X_n)$

$T_4(\underline{X}) = (X_1 + X_2 + X_3 + X_4, X_5, \ldots, X_n)$

$$T_5(\underline{X}) = \left(X_1 + X_2, \sum_{i=3}^{n} X_i\right)$$

$$\vdots$$

$$T(\underline{X}) = \sum_{i=1}^{n} X_i$$

Intuitively, among all the above sufficient statistics, $T(\underline{X}) = \sum_{i=1}^{n} X_i$ appears to be "best", giving maximum possible reduction.

<u>Def</u>^{n}:

<u>Minimal sufficient statistic</u>

Let $X_1, \ldots, X_n$ be a random sample from a dist^n P_θ, $\theta \in \Theta$, having p.d.f. or p.m.f. $f_\theta(x)$. A statistic $T(\underline{X})$ is said to be minimal sufficient for θ if

(i) $T(\underline{X})$ is sufficient.

& (ii) $T(\underline{X})$ is a function of every other sufficient statistics.

<u>Note</u>: Note that in the above example, $T(\underline{X}) = \sum_{1}^{n} X_i$ is a f^n of all other listed sufficient statistics.

An <u>important result to find</u> minimal sufficient statistic

Let $f_\theta(\underset{\sim}{x})$ be the joint p.d.f (or p.m.f) of $X_1, \ldots, X_n$ from P_θ $(\theta \in \Theta)$, $\underset{\sim}{x} \in \mathcal{X}$. Suppose $\exists$ a function $T(.)$ such that for every two sample points $\underset{\sim}{x}$ and $\underset{\sim}{y}$ $(\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X})$, the ratio $f_\theta(\underset{\sim}{x}) / f_\theta(\underset{\sim}{y})$ is independent of θ iff $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$. Then $T(\underset{\sim}{X})$ is a minimal sufficient statistic for θ.

<u>Remark</u>: The above result can be used to find minimal sufficient statistic.

<u>Example</u>: $X_1, \ldots, X_n$ random sample from $N(\theta, 1)$; $\theta \in \mathbb{R}$

$$f_\theta(\underset{\sim}{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_1^n (x_i - \theta)^2} \qquad \underset{\sim}{x} \in \mathbb{R}^n$$

$$f_\theta(\underset{\sim}{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum x_i^2 - \frac{n}{2}\theta^2 + \theta \sum x_i}$$

$\forall\, \underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$, we have

$$\frac{f_\theta(\underset{\sim}{x})}{f_\theta(\underset{\sim}{y})} = \frac{\cancel{\left(\frac{1}{\sqrt{2\pi}}\right)^n} e^{-\sum x_i^2/2} \cancel{e^{-n/2\,\theta^2}} e^{\theta \sum x_i}}{\cancel{\left(\frac{1}{\sqrt{2\pi}}\right)^n} e^{-\sum y_i^2/2} \cancel{e^{-n/2\,\theta^2}} e^{\theta \sum y_i}}$$

i.e. $$\frac{f_\theta(\underset{\sim}{x})}{f_\theta(\underset{\sim}{y})} = \underbrace{e^{-\frac{1}{2}\left(\Sigma x_i^2 - \Sigma y_i^2\right)}}_{\text{indep of } \theta} \; e^{\theta\left(\sum_1^n x_i - \Sigma y_i\right)}$$

$\Rightarrow$ $\dfrac{f_\theta(\underset{\sim}{x})}{f_\theta(\underset{\sim}{y})}$ is indep of θ <u>iff</u> $\sum_1^n x_i = \sum_1^n y_i$

$\Rightarrow$ (by the previous result) $T(\underset{\sim}{X}) = \sum_{i=1}^n X_i$ is minimal sufficient statistic for θ.

Example

$X_1, \ldots, X_n$ random sample from $N(\mu, \sigma^2)$

$\mu \in \mathbb{R}, \sigma > 0 \quad \underset{\sim}{\theta} = (\mu, \sigma)'$

$\forall \underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$

$$\frac{f_{\underset{\sim}{\theta}}(\underset{\sim}{x})}{f_{\underset{\sim}{\theta}}(\underset{\sim}{y})} = \frac{\exp\left(-\frac{1}{2}\left(\frac{\Sigma x_i^2}{\sigma^2} + \frac{n\mu^2}{\sigma^2} - \frac{2\mu}{\sigma^2}\Sigma x_i\right)\right)}{\exp\left(-\frac{1}{2}\left(\frac{\Sigma y_i^2}{\sigma^2} + \frac{n\mu^2}{\sigma^2} - \frac{2\mu}{\sigma^2}\Sigma y_i\right)\right)}$$

$$= \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma^2}\left(\Sigma x_i^2 - \Sigma y_i^2\right) - \frac{2\mu}{\sigma^2}\left(\Sigma x_i - \Sigma y_i\right)\right)\right)$$

indep of θ iff $\Sigma x_i = \Sigma y_i$ & $\Sigma x_i^2 = \Sigma y_i^2$

$\Rightarrow T(\underset{\sim}{X}) = \left(\sum_1^n X_i, \sum_1^n X_i^2\right)$ or $\left(\bar{X}, \Sigma(X_i - \bar{X})^2\right)$ is m.s.s.

(i.e. jointly minimal suff for θ)

Example: $X_1, \ldots, X_n$ random sample from $N(\theta, \theta^2)$

$\theta > 0$

$\forall \underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$

$$\frac{f_\theta(\underset{\sim}{x})}{f_\theta(\underset{\sim}{y})} = \exp\left(-\frac{1}{2}\left(\frac{1}{\theta^2}\left(\Sigma x_i^2 - \Sigma y_i^2\right) - \frac{2}{\theta}\left(\Sigma x_i - \Sigma y_i\right)\right)\right)$$

indep of θ iff $\Sigma x_i = \Sigma y_i$ & $\Sigma x_i^2 = \Sigma y_i^2$

$\Rightarrow T(\underset{\sim}{X}) = \left(\Sigma X_i, \Sigma X_i^2\right)$ is jointly minimal suff statistic for θ.

Example: $X_1, \ldots, X_n$ random sample from $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$

$\theta \in \mathbb{R}$

$\forall \underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$

$$\frac{f_\theta(\underset{\sim}{x})}{f_\theta(\underset{\sim}{y})} = \frac{I_{(\theta - \frac{1}{2}, x_{(1)})} I_{(x_{(n)}, \theta + \frac{1}{2})}}{I_{(\theta - \frac{1}{2}, y_{(1)})} I_{(y_{(n)}, \theta + \frac{1}{2})}}$$

indep of θ iff $x_{(1)} = y_{(1)}$ & $x_{(n)} = y_{(n)}$

$\Rightarrow T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ is jointly minimal sufficient statistic for θ.

Improvement of an unbiased estimator using sufficient statistic

Consider the problem of estimation of θ in $N(\theta,1)$; $\theta\in\mathbb{R}$

$(X_1, \ldots, X_n)$ random sample

X_i $(i=1,\ldots n)$, $\frac{X_1+X_2}{2}$, $\frac{X_1+X_n}{2}$, ..., $\bar{X}$ are all unbiased estimator for θ; $\exists$ infinite number of unbiased estimators in this case.

A natural criterion to pickup the "best" unbiased estimator would be to look for unbiased estimator having least variance $\forall\theta$, i.e. find $\delta^*(\underline{X})$ $\ni$

(i) $E_\theta\, \delta^*(\underline{X}) = g(\theta)$ ← the estimand (θ in $N(\theta,1)$ example)

& (ii) $V_\theta\, \delta^*(\underline{X}) \le V_\theta(\delta(\underline{X}))$ $\forall\, \theta\in\Theta$ and $\forall\,\delta$ satisfying (i)

Such a $\delta^*(\underline{X})$ is called Uniformly minimum variance unbiased estimator for $g(\theta)$ (or UMVUE for $g(\theta)$)

Note: The following result provides a way to improve upon an unbiased estimator, in terms of lower variance, using information of sufficient statistic.

Rao - Blackwell Theorem

Let $\delta(\underline{X})$ be any unbiased estimator of $g(\theta)$ and $T(\underline{X})$ be a sufficient statistic for θ.

Define $\eta(T) = E(\delta(\underline{X})\,|\,T)$

then

(i) $\eta(T)$ is a statistic as T is sufficient

(ii) $E(\eta(T)) = E\, E(\delta(\underline{X}) \mid T)$

$$= E(\delta(\underline{X})) = g(\theta)$$

i.e. $\eta(T)$ is an unbiased estimator of $g(\theta)$

and (iii) $V(\eta(T)) \le V(\delta(\underline{X}))$ (equality iff $\eta(T) = \delta(\underline{X})$ w.p. 1)

(with probability)

<u>Remark</u>: The above th^m. leads to a new estimator $\eta(T) = E(\delta(X) \mid T)$; which is called Rao-Blackwellized version of $\delta(\underline{X})$.

Example:

$X_1, \ldots, X_n$ random sample from $B(1, \theta)$
$0 < \theta < 1$

$g(\theta) = \theta$

$T(\underline{X}) = \sum_1^n X_i$ is suff (also minimal suff)

$\delta(\underline{X}) = X_1$ an unbiased estimator for θ

$$\eta(T) = E(X_1 \mid T)$$

$$= 0 \; P(X_1 = 0 \mid T) + 1 \cdot P(X_1 = 1 \mid T)$$

$$= P(X_1 = 1 \mid T)$$

$$= \frac{P(X_1 = 1, \sum_1^n X_i = t)}{P(\sum_1^n X_i = t)}$$

$$= \frac{P(X_1 = 1, \sum_2^n X_i = t-1)}{P(\sum_1^n X_i = t)}$$

$$= \frac{P(X_1=1)\, P\left(\sum_2^n X_i = t-1\right)}{P\left(\sum_1^n X_i = t\right)} \quad \left(X_1, \ldots X_n \text{ are indep}\right)$$

$$= \frac{\left(\theta^1 (1-\theta)^{1-1}\right)\left(\binom{n-1}{t-1} \theta^{t-1} (1-\theta)^{n-1-(t-1)}\right)}{\binom{n}{t} \theta^t (1-\theta)^{n-t}}$$

$$= \frac{\binom{n-1}{t-1}}{\binom{n}{t}} = \frac{t}{n}$$

$$\Rightarrow \eta(T) = E(\delta(\underline{X}) \mid T) = \frac{T}{n}$$

By Rao-Blackwell th^{m} $V(\eta(T)) \le V(\delta)$

Remark: Start with an other unbiased estimator $\tilde{\delta}(\underline{X}) \to \eta(T)$ would be the same !!

This however may not happen for other suff stat, say, $(X_1, \ldots, X_n)$ as suff statistic.

Example: $X_1, \ldots, X_n$ i.i.d. random sample from $P(\theta)$ $\theta > 0$

$g(\theta) = e^{-\theta} \left(= P(X_1 = 0)\right)$

$T(\underline{X}) = \sum_1^n X_i$ is sufficient for θ (also minimal suff)

Let $\delta(\underline{X}) = \begin{cases} 1, & X_1 = 0 \\ 0, & o/w \end{cases}$ $\quad E\,\delta(\underline{X}) = P(X_1 = 0) = e^{-\theta}$

$\Rightarrow \delta(\underline{X})$ is unbiased estimator for $e^{-\theta}$

$$\eta(T) = E(\delta(\underline{X}) \mid T)$$

$$= P(X_1 = 0 \mid T)$$

$$= \frac{P(X_1 = 0, T = t)}{P(T = t)}$$

$$= \frac{P\left(X_1 = 0, \sum_2^n X_i = t\right)}{P\left(\sum_1^n X_i = t\right)} \qquad \left(T \sim P(n\theta)\right)$$

$$\sum_2^n X_i \sim P((n-1)\theta)$$

$$= \frac{P(X_1 = 0)\, P\left(\sum_2^n X_i = t\right)}{P\left(\sum_1^n X_i = t\right)}$$

$$= \frac{\left(e^{-\theta}\right)\left(e^{-(n-1)\theta}((n-1)\theta)^t / t!\right)}{\dfrac{e^{-n\theta}(n\theta)^t}{t!}}$$

$$= \left(\frac{n-1}{n}\right)^t$$

$$\Rightarrow \eta(T) = \left(\frac{n-1}{n}\right)^T.$$

By Rao-Blackwell th^{m} $V(\eta(T)) < V(\delta)$

<u>Remark</u>: Once again if we start with any other starting unbiased estimator δ and use $T = \Sigma X_i$, we would get same $\eta(T) = \left(\frac{n-1}{n}\right)^T$

<u>Remark</u>: An additional property of sufficient statistic that ensures existence of unique unbiased estimator based on sufficient statistic, which has minimum variance among all other unbiased estimators, is "Completeness".

Complete Statistic

Def^n: A statistic T is said to be complete if for any real valued f^n g

$$E\,g(T) = 0 \quad \forall\, \theta \in \Theta$$

$$\Rightarrow \quad g(t) = 0 \quad \text{with probability 1}$$

(i.e. almost everywhere)

Remark: Completeness of sufficient statistic is important if we are trying to find the "best" unbiased estimator, i.e. the UMVUE.

Let $g(\theta)$ be an estimand and T be complete sufficient statistic.

Suppose $\exists$ an unbiased estimator of $g(\theta)$. Then $g(\theta)$ has one and only one unbiased estimator that is a f^n of T.

If $\delta(\underset{\sim}{X})$ is u.e. of $g(\theta)$, then

$\eta(T) = E(\delta(\underset{\sim}{X}) \mid T)$ is also u.e. (Rao-Blackwell)

$\eta(T)$ is an u.e. based on T

Let $\eta^*(T)$ be another u.e. of $g(\theta)$ based on T

$$E(\eta(T) - \eta^*(T)) = 0 \quad \forall\, \theta \in \Theta$$

As T is complete, the above implies that

$$\eta(T) = \eta^*(T) \quad \text{with prob 1}$$

i.e. essentially $\exists$ one u.e. of $g(\theta)$ based on T which has the lowest variance among all u.e.s

Remark: If T is thus complete sufficient, unbiased estimator of $g(\theta)$ based on T would be the unique UMVUE – the "best" unbiased estimator.

Remark: If T is complete sufficient

$g(\theta)$: estimand

$\exists$ an u.e. of $g(\theta)$

Unique UMVUE for $g(\theta)$ is

$$\eta(T) = E(\delta(X) \mid T)$$

Approaches to prove completeness of suff statistic

(I) s-parameter exponential family argument

Def^n: X has a dist^n of s-parameter exponential family if it's p.d.f. or p.m.f. is of the form

$$f(x) = h(x) \exp\left(\sum_{i=1}^{s} \eta_i(\theta)\, T_i(x) - \beta(\theta)\right)$$

or

$$f(x) = h(x) \exp\left(\sum_{i=1}^{s} \eta_i\, T_i(x) - A(\eta)\right)$$

in the reparametrized form (in terms of η parametrization)

$\{\underset{\sim}{\eta} : \underset{\sim}{\theta} \in \Theta\}$ – is called the natural parameter space

Approaches to prove completeness of suff statistic

(I) s-parameter exponential family argument

Def^n

X has a dist^n of s-parameter exponential family if it's p.d.f. or p.m.f. is of the form

$$f(x) = h(x) \exp\left(\sum_{i=1}^{s} \eta_i(\theta)\, T_i(x) - \beta(\theta) \right)$$

or

$$f(x) = h(x) \exp\left(\sum_{i=1}^{s} \eta_i\, T_i(x) - A(\eta) \right)$$

in the reparametrized form (in terms of η parametrization)

$\{ \underset{\sim}{\eta} : \underset{\sim}{\theta} \in \Theta \}$ – is called the natural parameter space

Remark: Many of the common dist^{n}s follow exponential family dist^n setup for some 's'.

e.g. Normal, exponential (scale), gamma, chi-square, log-normal, beta, Bernoulli, Binomial (n known), Poisson, geometric, negative binomial, etc.

Remark: Distributions for which range is dependent on parameter, do not ~~be~~ belong to exponential family dist^n setup.

e.g. $U(0, \theta)$, $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $U(-\theta, \theta)$ exponential (location), exponential (location-scale).

Remark: If the natural parameter space associated with an s-parameter exponential family dist^n contains an s-dimensional open rectangle (open interval for $s=1$), then the s-parameter exponential family dist^n is said to be of "full rank".

An important result:

If an s-parameter exponential family dist^n is of full rank, then the associated minimal sufficient statistic is complete.

i.e. $T(\underline{X}) = \left(\sum_{j=1}^{n} T_1(X_j), \sum_{j=1}^{n} T_2(X_j), \ldots, \sum_{j=1}^{n} T_s(X_j) \right)$

is complete sufficient

<u>Remark</u>: The above result can be used to prove completeness of minimal sufficient for all such distributions.

<u>Examples</u>

(i) $X \sim P(\theta)$ $\theta > 0$ $\Theta = \{\theta : \theta > 0\}$

p.m.f. $f_\theta(x) = \dfrac{e^{-\theta}\theta^x}{x!}$ $x = 0, 1, \ldots$

$= \dfrac{1}{x!} e^{x \log \theta - \theta}$

$h(x) = \dfrac{1}{x!}$; $T_1(x) = x$; $\eta_1(\theta) = \log \theta$; $\beta(\theta) = \theta$

i.e. $f_\eta(x) = \dfrac{1}{x!} \exp(\eta_1 T_1(x) - A(\eta))$

This is 1-parameter exponential family form with natural parameter space as

$$\{\eta : \eta \in \mathbb{R}\}$$

The above natural parameter space contains open rectangles and hence the 1-parameter exponential family is of full rank

$\Rightarrow T(\underline{X}) = \sum_{i=1}^{n} X_i$ is complete suff stat

(2) $X \sim B(1,\theta)$ $1 > \theta > 0$ $\quad \Theta = \{\theta : 0 < \theta < 1\}$

p.m.f $\quad f_\theta(x) = \theta^x (1-\theta)^{1-x} \qquad x = 0, 1$

$$= \exp\left(x \log\left(\frac{\theta}{1-\theta}\right) + \log(1-\theta) \right)$$

with $h(x) = 1$, $T_1(x) = x$, $\eta(\theta) = \log \frac{\theta}{1-\theta}$; $\beta(\theta) = -\log(1-\theta)$

The above is 1-parameter exponential family

$$f_\eta(x) = \exp(x\eta - A(\eta))$$

Natural parameter space : $\{\eta : \eta \in \mathbb{R}\}$ contains open intervals

$\Rightarrow$ The above 1-param expo family distⁿ is of full rank

$\Rightarrow$ $T(\underline{X}) = \sum_{i=1}^{n} X_i$ is complete suff statistic

(3) $X \sim N(\mu, \sigma^2)$ $\quad \underset{\sim}{\theta} = (\mu, \sigma)'$

$\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$

p.d.f. $\quad f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2\sigma^2}(x^2 + \mu^2 - 2\mu x)\right)$

$$= \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left[-\frac{1}{2\sigma^2} x^2 + \frac{\mu}{\sigma^2} x - \frac{\mu^2}{2\sigma^2} - \log\sigma\right)$$

$$= h(x) \exp\left(\sum_{i=1}^{2} T_i(x) \eta_i(\underset{\sim}{\theta}) - \beta(\underset{\sim}{\theta}) \right)$$

$h(x) = \frac{1}{\sqrt{2\pi}}$; $T_1(x) = x^2$; $\eta_1(\underline{\theta}) = -\frac{1}{2\sigma^2} \; (= \eta_1)$

$T_2(x) = x$; $\eta_2(\underline{\theta}) = \frac{\mu}{\sigma^2} \; (= \eta_2)$

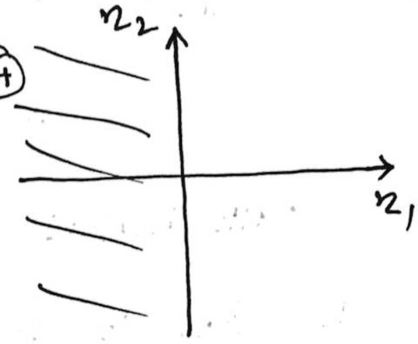
$\beta(\underline{\theta}) = \frac{\mu^2}{2\sigma^2} + \log\sigma$

The above is 2-parameter exponential family representation

$$f_{\underset{\sim}{\eta}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\eta_1 T_1(x) + \eta_2 T_2(x) - A(\underline{\eta}) \right)$$

with natural parameter space as Θ

$$\{ (\eta_1, \eta_2) : \eta_1 < 0 ,\ \eta_2 \in \mathbb{R} \}$$



which contains 2-dim open rectangles

$$\Rightarrow T(\underline{X}) = \left(\sum_{i=1}^{n} X_i ,\ \sum X_i^2 \right) \text{ is complete sufficient stat}$$

$$\Leftrightarrow \left(\bar{X} ,\ \frac{1}{n-1} \sum (X_i - \bar{X})^2 = S^2 \right) \text{ is complete suff stat}$$

(4) example of non-full rank s-parameter expo family distⁿ

$X \sim N(\theta, \theta) \qquad \theta > 0 \ ; \ \Theta = \{ \theta : \theta > 0 \}$

p.d.f. $\quad f_\theta(x) = \frac{1}{\sqrt{2\pi}\,\theta} \exp\left(-\frac{1}{2\theta^2} (x^2 + \theta^2 - 2\theta x) \right)$

$$= \underbrace{\left(\frac{e^{-1/2}}{\sqrt{2\pi}} \right)}_{h(x)} \exp\left(-\frac{1}{2\theta^2} x^2 + \frac{1}{\theta} x - \log \theta \right)$$

$T_1(x) = x^2 \quad \eta_1(\theta) = -\frac{1}{2\theta^2} \ (= \eta_1)$

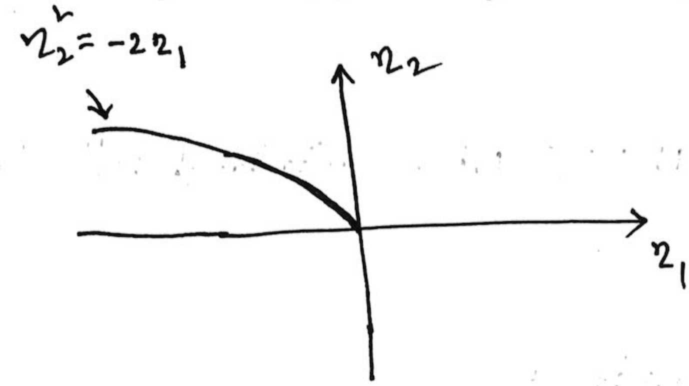
$T_2(x) = x \ ; \ \eta_2(\theta) = \frac{1}{\theta} \ (= \eta_2)$

The above is 2-parameter exponential family

$$f_{\underset{\sim}{\eta}}(x) = \left(\frac{e^{-1/2}}{\sqrt{2\pi}} \right) \exp\left(T_1(x)\eta_1 + T_2(x)\eta_2 - A(\underline{\eta}) \right)$$

The natural parameter space is

$$\{(\eta_1, \eta_2) : \eta_2^2 = -2\eta_1, \eta_1 < 0, \eta_2 > 0\}$$



The natural parameter space is a curve and does not contain an open 2-dim rectangle

$\Rightarrow$ The 2-parameter expo family is not of full rank

Remark: We can show that here $\underset{\sim}{T} = (T_1, T_2)$; $T_1 = \Sigma X_i$ & $T_2 = \Sigma X_i^2$

is not complete

$$E\,T_2 = n\,E X_1^2 = n\,2\theta$$

$$E\,T_1^2 = V(T_1) + (E\,T_1)^2 = n\theta^2 + (n\theta)^2$$

$$= \theta^2 n(n+1)$$

$$\left.\begin{array}{l} \Rightarrow E\left(\dfrac{T_1^2}{n(n+1)} - \dfrac{T_2}{2n}\right) = \theta^2 - \theta^2 = 0 \quad \forall \theta \in \Theta \\[2ex] \not\Rightarrow \dfrac{T_1^2}{n(n+1)} = \dfrac{T_2}{2n} \quad \text{w.p. 1 (a.e.)} \end{array}\right\} - (*)$$

In fact $P\left(\dfrac{T_1^2}{n(n+1)} = \dfrac{T_2}{2n}\right) = 0$!!

(note that $\dfrac{T_1^2}{n(n+1)} - \dfrac{T_2}{2n}$ is a cont r.v.)

Remark: We can show that here $\underset{\sim}{T} = (T_1, T_2)$; $T_1 = \sum X_i$
& $T_2 = \sum X_i^2$

is not complete

$$E\, T_2 = n\, E X_1^2 = n\, 2\theta$$

$$E\, T_1^2 = V(T_1) + (E\, T_1)^2 = n\theta^2 + (n\theta)^2$$

$$= \theta^2 n(n+1)$$

$$\left.\begin{aligned} &\Rightarrow E\left(\frac{T_1^2}{n(n+1)} - \frac{T_2}{2n}\right) = \theta^2 - \theta^2 = 0 \quad \forall \theta \in \Theta \\ &\not\Rightarrow \frac{T_1^2}{n(n+1)} = \frac{T_2}{2n} \quad \text{w.p. } 1 \text{ (a.e.)} \end{aligned}\right\} - (*)$$

In fact $P\left(\frac{T_1^2}{n(n+1)} = \frac{T_2}{2n}\right) = 0$!!

(note that $\frac{T_1^2}{n(n+1)} - \frac{T_2}{2n}$ is a cont r.v.)

$(*) \Rightarrow T(\underline{X}) = \left(\sum_{i=1}^{n} X_i , \sum_{i=1}^{n} X_i^2 \right)$ is NOT complete although

It is minimal suff.

Remark: Approaches to find UMVUE when minimal suff stat is complete

Step I : Find complete suff statistic

Step II : Find a function of complete suff stat which is unbiased for the estimand - this will be the UMVUE

Step II Calculations

- For simple estimands it is easy to find u.e. based on c.s.s.
- For complicated estimands use Rao-Blackwellization or solve for the eqn

$$E(\delta(T)) = g(\theta) \quad \forall \theta \in \Theta$$

- Find UMVUE thro Cramér-Rao Lower Bound (If the bound is attainable)

Cramer-Rao Lower Bound (CRLB)

CRLB provides lower bound for the variance of any unbiased estimator of $g(\theta)$.

$X_1, \ldots, X_n$ be i.i.d. random sample from $f_\theta(x)$

$$\theta \in \Theta$$

$g(\theta)$: estimand

$g(\theta)$ is $\ni$ $\exists$ unbiased estimator of $g(\theta)$

Suppose the following regularity conditions hold

(i) support of the r.v.s does not depend on θ

(ii) $g(\theta)$ is differentiable

(iii) derivate of $\frac{\partial}{\partial\theta} f_\theta(x)$ exists and is finite

(iv) derivate of $\int f_\theta(x)\,dx$, w.r.t. θ, can be obtained by differentiating under the integral.

Let $\delta(\underline{X})$ be any unbiased estimator of $g(\theta)$

then

$$V(\delta(\underline{X})) \geq \left(\frac{(g'(\theta))^2}{n\, E\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2}\right) \quad \text{CRLB}$$

~~Alternate~~

Alternate form of CRLB

$$V(\delta(\underline{X})) \geq \frac{(g'(\theta))^2}{-n\, E\left(\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right)}$$

provided that 2nd derivate conditions (existence and interchange of differentiation and integration) holds

Remark: $E\left(\frac{\partial}{\partial\theta}\log f_\theta(x)\right)^2 = I(\theta)$ is called the Fisher information

$\nwarrow V\left(\frac{\partial}{\partial\theta}\log f_\theta(x)\right)$ as $E\left[\frac{\partial}{\partial\theta}\log f_\theta(x)\right] = 0$

Alt form: $I(\theta) = -E\left(\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2}\right)$

Remark: If $\exists$ an unbiased estimator whose variance equals CRLB, then it is UMVUE.

Remark: There can be situations wherein UMVUE has variance higher than CRLB. In such cases, CRLB is not achievable.

Examples:

(i) $X_1, \ldots, X_n$ i.i.d. $B(1,\theta)$

$$f_\theta(x) = \theta^x (1-\theta)^{1-x}$$

$$\log f_\theta(x) = x\log\theta + (1-x)\log(1-\theta)$$

$$\frac{\partial \log f_\theta(x)}{\partial\theta} = \frac{x}{\theta} + (1-x)\frac{1}{1-\theta}(-1)$$

$$= \frac{x}{\theta} - (1-x)\frac{1}{1-\theta}$$

$$\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2} = -\frac{x}{\theta^2} - (1-x)\frac{1}{(1-\theta)^2}$$

$$E\left(\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2}\right) = -\frac{\theta}{\theta^2} - \frac{1-\theta}{(1-\theta)^2} = -\frac{1}{\theta(1-\theta)}$$

$$I(\theta) = -E\left(\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2}\right) = \frac{1}{\theta(1-\theta)}$$

Estimand: $g(\theta) = \theta$

$$\text{CRLB} = \frac{(g'(\theta))^2}{n\,I(\theta)} = \frac{\theta(1-\theta)}{n}$$

$\delta(\underline{x}) = \frac{\sum X_i}{n}$ u.e. for $g(\theta)$

$$V\left(\frac{\Sigma X_i}{n}\right) = \frac{1}{n^2}\Sigma V(X_i)$$

$$= \frac{1}{n^2} n\,\theta(1-\theta) = \frac{\theta(1-\theta)}{n} = CRLB$$

$\Rightarrow \bar{X}$ is UMVUE for θ.

Example (ii)

$X_1, \dots X_n$ r.s $N(\theta, 1)$

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$$

$$\log f_\theta(x) = c - \frac{1}{2}(x-\theta)^2$$

$$\frac{\partial \log f_\theta(x)}{\partial\theta} = -\frac{1}{\cancel{2}}\cancel{2}(x-\theta)(-1) = x-\theta$$

$$E\left(\frac{\partial \log f_\theta(X)}{\partial\theta}\right)^2 = E(X-\theta)^2 = 1 = I(\theta)$$

or $$\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2} = -1 \Rightarrow -E\left(\frac{\partial^2 \log f_\theta(X)}{\partial\theta^2}\right) = I(\theta) = 1$$

$g(\theta) = \theta$ say

$$CRLB = \frac{(g'(\theta))^2}{n\,I(\theta)} = \frac{1}{n}$$

$$V(\bar{X}) = \frac{1}{n} = CRLB$$

$\Rightarrow \bar{X}$ is UMVUE for θ

if $g(\theta) = \theta^2$

$$CRLB = \frac{(g'(\theta))^2}{n\,I(\theta)} = \frac{4\theta^2}{n}$$

Consistent Estimator

A large sample optimal property

Def^n: An estimator $\delta(\underline{X})$ is said to be consistent for $g(\theta)$

If $\delta(\underline{X}) \xrightarrow{p} g(\theta)$ as $n \to \infty$

Remark: Use WLLN to prove consistency or use definition of $\xrightarrow{p}$

EX1: $X_1, \ldots X_n$ r.s. $N(\mu, \sigma^2)$

By WLLN (i) $\frac{1}{n}\sum X_i \xrightarrow{p} \mu$

$\Rightarrow \bar{X}$ is consistent est of μ

S_n^2 (say)

(ii) $\frac{1}{n}\sum_1^n (X_i - \bar{X})^2 \xrightarrow{p} \sigma^2$

$\Rightarrow \frac{1}{n}\sum (X_i - \bar{X})^2$ is consistent for σ^2

$$\frac{\bar{X}}{S_n^2} \xrightarrow{p} \frac{\mu}{\sigma^2}$$

$\Rightarrow \frac{\bar{X}}{S_n^2}$ is a consistent estimator for $\frac{\mu}{\sigma^2}$

EX: $X_1, \ldots X_n$ r.s. from $U(0, \theta)$ $\theta > 0$

$X_{(n)} \xrightarrow{p} \theta$ (proved earlier)

$\Rightarrow X_{(n)}$ is consistent for θ

Remark: $E\left(\frac{\partial}{\partial\theta}\log f_\theta(x)\right)^2 = I(\theta)$ is called the Fisher information

$V\left(\frac{\partial}{\partial\theta}\log f_\theta(x)\right)$ as $E\left(\frac{\partial}{\partial\theta}\log f_\theta(x)\right) = 0$

Alt form: $I(\theta) = -E\left(\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2}\right)$

Remark: If $\exists$ an unbiased estimator whose variance equals CRLB, then it is UMVUE.

Remark: There can be situations wherein UMVUE has variance higher than CRLB. In such cases, CRLB is not achievable.

Examples:

(i) $X_1, \ldots, X_n$ i.i.d. $B(1,\theta)$

$$f_\theta(x) = \theta^x (1-\theta)^{1-x}$$

$$\log f_\theta(x) = x\log\theta + (1-x)\log(1-\theta)$$

$$\frac{\partial \log f_\theta(x)}{\partial\theta} = \frac{x}{\theta} + (1-x)\frac{1}{1-\theta}(-1)$$

$$= \frac{x}{\theta} - (1-x)\frac{1}{1-\theta}$$

$$\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2} = -\frac{x}{\theta^2} - (1-x)\frac{1}{(1-\theta)^2}$$

$$E\left(\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2}\right) = -\frac{\theta}{\theta^2} - \frac{1-\theta}{(1-\theta)^2} = -\frac{1}{\theta(1-\theta)}$$

$$I(\theta) = -E\left(\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2}\right) = \frac{1}{\theta(1-\theta)}$$

Estimand: $g(\theta) = \theta$

$$\text{CRLB} = \frac{(g'(\theta))^2}{n\,I(\theta)} = \frac{\theta(1-\theta)}{n}$$

$\delta(\underline{x}) = \frac{\sum X_i}{n}$ u.e. for $g(\theta)$

$$V\left(\frac{\sum X_i}{n}\right) = \frac{1}{n^2}\sum V(X_i)$$

$$= \frac{1}{n^2}\, n\, \theta(1-\theta) = \frac{\theta(1-\theta)}{n} = CRLB$$

$\Rightarrow \bar{X}$ is UMVUE for θ.

Example (ii)

$X_1, \ldots X_n$ r.s $N(\theta, 1)$

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$$

$$\log f_\theta(x) = c - \frac{1}{2}(x-\theta)^2$$

$$\frac{\partial \log f_\theta(x)}{\partial \theta} = -\frac{1}{\cancel{2}}\cancel{2}(x-\theta)(-1) = x-\theta$$

$$E\left(\frac{\partial \log f_\theta(X)}{\partial \theta}\right)^2 = E(X-\theta)^2 = 1 = I(\theta)$$

or $$\frac{\partial^2 \log f_\theta(x)}{\partial \theta^2} = -1 \Rightarrow -E\left(\frac{\partial^2 \log f_\theta(X)}{\partial \theta^2}\right) = I(\theta) = 1$$

$g(\theta) = \theta$ say

$$CRLB = \frac{(g'(\theta))^2}{n \cdot I(\theta)} = \frac{1}{n}$$

$$V(\bar{X}) = \frac{1}{n} = CRLB$$

$\Rightarrow \bar{X}$ is UMVUE for θ

If $g(\theta) = \theta^2$

$$CRLB = \frac{(g'(\theta))^2}{n\, I(\theta)} = \frac{4\theta^2}{n}$$

Maximum Likelihood Estimator (MLE)

Let $X_1, \ldots, X_n$ be an i.i.d. random sample from $f_\theta(x)$ (p.d.f. or p.m.f), $\theta \in \Theta$

jt p.d.f. (or p.m.f)

$$f_{X_1,\ldots,X_n} = \prod_{i=1}^{n} f_\theta(x_i)$$

Likelihood f^n:

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) \quad \text{viewed as a function of } \theta$$

given the observations $x_1, \ldots, x_n$

Note that for a discrete dist^n setup $L(\theta)$ is probability of observing $(x_1, \ldots, x_n)$ and for a continuous dist^n setup $L(\theta)$ is proportional to a probability statement.

MLE approach: Find θ which maximises the likelihood (linked with the above prob statement interpretation)

Def^n: $\hat{\theta}$ is an MLE of θ if

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max}\ L(\theta)$$

Remark: MLE is a function of sufficient statistic

T is suff $\Rightarrow$ by NFFT

$$f_\theta(\underline{x}) = h(\underline{x})\, g_\theta(t(\underline{x})) = L(\theta)$$

Thus, maximization of L w.r.t θ

$\Longleftrightarrow$ maximization of $g_\theta(t(\underline{x}))$ w.r.t θ

Hence, MLE is a f^n of suff stat $T(\underline{X})$

Remark: Note that (as log is a monotone f^n),

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max} \log L(\theta)$$

It is often convenient to work with $\log L(\theta)$ to find MLE

$l(\theta) = \log L(\theta)$ – log likelihood f^n.

Remark: Invariance property of MLE

$\theta \in \Theta \subseteq \mathbb{R}^k$, say

Let $\hat{\theta}$ be MLE of θ and $g(.)$ be a f^n from Θ to a subset of $\mathbb{R}^m$ (say). Then $g(\hat{\theta})$ is MLE of $g(\theta)$.

Remark: Suppose $L(\theta)$ (or $l(\theta)$) is differentiable w.r.t. θ and the maximum of $L(\theta)$ ($l(\theta)$) is an interior point of Θ and not a point on the boundary then $\hat{\theta}$ satisfies

$$\left.\frac{\partial\, l(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0$$

and

$$\left.\frac{\partial^2 l(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} < 0$$

Similar conditions for multiparameter setup.

In such a situation, MLE can be obtained by solving

$$\frac{\partial \ell(\theta)}{\partial \theta} = 0$$

and verifying that $\left.\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right|_{\theta = \hat{\theta}} < 0$.

Computation of MLE

Examples

(i) $X_1, \dots, X_n$ r.s. from $B(1,\theta)$; $0<\theta<1$

Likelihood f^n : $L(\theta) = \theta^{\sum_1^n x_i} (1-\theta)^{n-\sum_1^n x_i}$

log likelihood f^n : $l(\theta) = L(\theta) = \sum x_i \log\theta + (n - \sum_1^n x_i) \log(1-\theta)$

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{\sum x_i}{\theta} - \frac{(n-\sum x_i)}{1-\theta}$$

likelihood eq^n : $\frac{\partial l(\theta)}{\partial\theta} = 0$

$$\Rightarrow \hat\theta = \sum_1^n x_i / n$$

Further $$\frac{\partial^2 l(\theta)}{\partial\theta^2} = -\frac{\sum x_i}{\theta^2} - \frac{(n-\sum x_i)}{(1-\theta)^2}$$

$$\Rightarrow \left.\frac{\partial^2 l(\theta)}{\partial\theta^2}\right|_{\hat\theta} < 0$$

$$\Rightarrow \hat\theta_{MLE} = \frac{1}{n}\sum_1^n X_i$$

(ii) $X_1, \dots, X_n$ r.s. from $P(\theta)$; $\theta>0$

Likelihood f^n : $L(\theta) = \frac{e^{-n\theta}\,\theta^{\sum_1^n x_i}}{\pi x_i!}$

Log likelihood f^n : $l(\theta) = -n\theta + \sum x_i \log\theta - \log(\pi x_i!)$

$$\frac{\partial l(\theta)}{\partial\theta} = -n + \frac{\sum x_i}{\theta}$$

$$\frac{\partial l(\theta)}{\partial\theta} = 0 \Rightarrow \hat\theta = \frac{\sum x_i}{n}$$

Further $\left.\frac{\partial^2 l(\theta)}{\partial \theta^2}\right|_{\hat\theta} = -\left.\frac{\Sigma x_i}{\theta^2}\right|_{\hat\theta} < 0$

$\Rightarrow \hat\theta_{MLE} = \frac{1}{n}\sum_1^n X_i$

(3) $X_1, \ldots, X_n$ r.s. from a distⁿ (exp) with

p.d.f. $f_\theta(x) = \begin{cases} \theta e^{-\theta x}, & x > 0 \\ 0, & o/w \end{cases}$

$L(\theta) = \theta^n e^{-\theta \Sigma x_i}$

$l(\theta) = n \log \theta - \theta \sum_1^n x_i$

$\frac{\partial l(\theta)}{\partial \theta} = \frac{n}{\theta} - \Sigma x_i$

$\frac{\partial l(\theta)}{\partial \theta} = 0 \Rightarrow \hat\theta = \frac{n}{\sum_1^n x_i}$

$\left.\frac{\partial^2 l(\theta)}{\partial \theta^2}\right|_{\hat\theta} = -\left.\frac{n}{\theta^2}\right|_{\hat\theta} < 0$

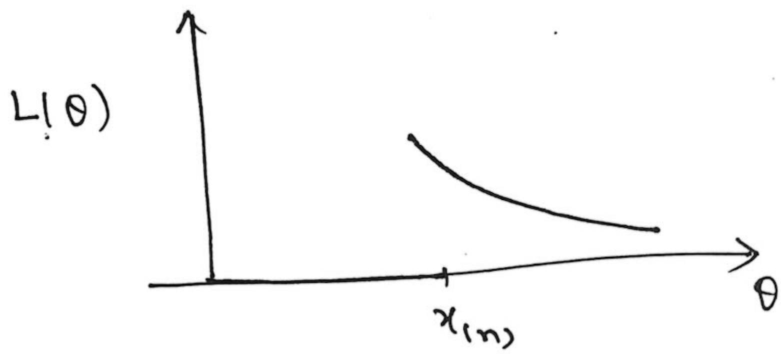
$\Rightarrow \hat\theta_{MLE} = \frac{n}{\sum_1^n X_i}$

(4) $X_1, \ldots, X_n$ r.s. from $U[0,\theta]$; $\theta > 0$

$$f_\theta(x) = \begin{cases} 1/\theta, & 0 \le x \le \theta \\ 0, & o/w \end{cases}$$

$$L(\theta) = \frac{1}{\theta^n} I(0, x_{(1)})\, I(x_{(n)}, \theta) \; ; \; I(a,b) = \begin{cases} 1, & a \le b \\ 0, & o/w \end{cases}$$

Note that $L(\theta)$ is not differentiable at $x_{(n)}$



i.e $L(\theta) = 0 \quad \forall\, \theta < x_{(n)}$

> 0 & $\downarrow \theta \quad \forall\, \theta \geq x_{(n)}$

$\Rightarrow L(\theta)$ is maximized at $x_{(n)}$

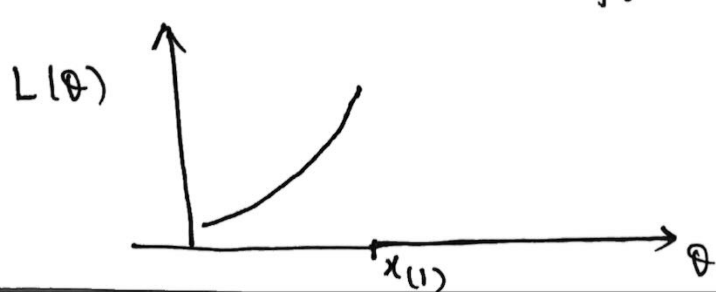
$$\Rightarrow \quad \hat{\theta}_{MLE} = X_{(n)}$$

(5) $X_1, \ldots, X_n$ r.s. from expo with location parameter θ

$$f_\theta(x) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta \\ 0, & o/w \end{cases}$$

$$L(\theta) = e^{-\sum x_i} e^{n\theta} I(\theta, x_{(1)})$$

Note that $L(\theta)$ is not differentiable at $x_{(1)}$



$$L(\theta) = 0 \quad \forall \ \theta > x_{(1)}$$
$$> 0 \ \& \ \uparrow \quad \forall \ \theta \leq x_{(1)}$$

$\Rightarrow L(\theta)$ is maximised at $x_{(1)}$

$$\Rightarrow \hat{\theta}_{MLE} = X_{(1)}$$

<u>Remark</u>: As the examples show, MLE need not be unbiased