1.1

<u>Sample space</u>: S = the set of all possible outcomes of an experiment

<u>Countable</u> if $\exists$ a 1-1 cooresponodence w/ a subset of integers

<u>Uncountable</u> if $!\exists$ a 1-1 correspondence (ie reaction time)

<u>Event</u>: any collection of possible outcomes of an experiment, any subset of S which includes S itself

<u>Containment</u> $A \subset B \leftrightarrow x \in A \rightarrow x \in B$

<u>Equality</u> $A = B \leftrightarrow A \subset B$ & $A \supset B$

<u>Union</u>: $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$

<u>Intersection</u> $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$

<u>Complement</u> $A^c = \{x \mid x \notin A\}$

<u>Null set</u> $\emptyset$ = the set consisting of no elements

Theorem 1.1.4

a. $A \cup B = B \cup A$

• Let $x \in A \cup B$ then, by definition, $x \in A$ or $x \in B$. then $x \in B \cup A$ because $x \in A$ or $x \in B$ and def. of $\cup$ //

$A \cap B = B \cap A$

• Let $x \in A \cap B$ then, by definition, $x \in A$ and $x \in B$. then $x \in B \cap A$ because $x \in A$ and $x \in B$ and def. of $\cap$ /

b. $A \cup (B \cup C) = (A \cup B) \cup C$

• Let $x_1 \in A \cup (B \cup C)$ then $x_1 \in A$ or $x_1 \in B \cup C$ which is the same as $x_1 \in A$ or $x \in B$ or $x \in C$ by def $\cup$.
Let $x_2 \in (A \cup B) \cup C$ then $x_2 \in A \cup B$ or $x_2 \in C$ which is the same as $x_2 \in A$ or $x_2 \in B$ or $x_2 \in C$ by def $\cup$.
$\therefore A \cup (B \cup C) = (A \cup B) \cup C$ b/c they are defined the same way //

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

• Let $x_1 \in A \cap (B \cap C)$ then $x_1 \in A$ and $x_1 \in B \cap C$ which is the same as $x_1 \in A$ and $x_1 \in B$ and $x_1 \in C$ by def $\cap$.
Let $x_2 \in (A \cap B) \cap C$ then $x_2 \in (A \cap B) \cap C$ which is the same as $x_2 \in A$ and $x_2 \in B$ and $x_2 \in C$ by def $\cap$
$\therefore A \cap (B \cap C) = (A \cap B) \cap C$ b/c the are defined the same way, thus containment is establised

c. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

• Let $x_1 \in A \cap (B \cup C)$ then $x_1 \in A$ and $x \in B$ or $x \in C$
Let $x_2 \in (A \cap B) \cup (A \cap C)$ then $x_2 \in A$ and $x_2 \in B$ or $x_2 \in A$ or $x_2 \in C$
which is the same as $x_2 \in A$ and $x_2 \in B$ or $x_2 \in C$
$\therefore A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ b/c they are defined the same way, thus containment is established

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

- Let $x_1 \in A \cup (B \cap C)$ then $x \in A$ or $x \in B$ and $x \in C$
- Let $x_2 \in (A \cup B) \cap (A \cup C)$ then $x_2 \in A$ or $x_2 \in B$ and $x_2 \in A$ or $x_2 \in C$ which is the same as $x_2 \in A$ or $x_2 \in B$ and $x_2 \in C$

$\therefore A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ because they are defined the same way thus containment is established

d. $(A \cup B)^c = A^c \cap B^c$

- Let $x_1 \in (A \cup B)^c$ then $x_1 \notin A \cup B$ which is the same as $x_1 \notin A$ and $x_1 \notin B$ which is equivalent to $x_1 \in A^c$ and $x_1 \in B^c$
- Let $x_2 \in A^c \cap B^c$ then $x \in A^c$ and $x \in B^c$

$\therefore (A \cup B)^c = A^c \cap B^c$ because they are defined the same way thus containment is established

$(A \cap B)^c = A^c \cup B^c$

- Let $x_1 \in (A \cap B)^c$ then $x_1 \notin A \cap B$ which is the same as $x_1 \notin A$ or $x_1 \notin B$ which implies $x \in A^c$ or $x_1 \in B^c$
- Let $x_2 \in A^c \cup B^c$ then $x_2 \in A^c$ or $x_2 \in B^c$

$\therefore (A \cap B)^c = A^c \cup B^c$ because they are defined the same way thus containment is established

$\bigcup_{i}^{\infty} A_i = \{x \mid x \in A_i \text{ for some } i\}$

$\bigcap_{i=1}^{\infty} A_i = \{x \mid x \in A_i \text{ for all } i\}$

<u>Disjoint</u> or <u>Mutually Exclusive</u>: if $A \cap B = \phi$

<u>Pairwise Disjoint or Mutually Exclusive</u>: if $A_i \cap A_j = \phi \;\; \forall\, i \neq j$

<u>Partition</u> if $A_i \cap A_j = \phi \;\forall\, i \neq j$ and $\bigcup_{i}^{\infty} A_i = S$

- $B = \{B \cap A\} \cup \{B \cap A^c\}$ $\boxed{\text{Proof}}$ $x \in B$ and $x \in A$ or $x \in B$ and A^c $\equiv x \in B$ and $x \in S$ $\therefore x \in B$ //
- $A \cup B = A \cup \{B \cap A^c\}$ $\boxed{\text{Proof}}$ $A \cup B = A \cup [(B \cap A) \cup (B \cap A^c)]$ (by the proof above) $= A \cup B \cap A^c$ $(A \cup (\text{a subset of } A)) = A$

1.2

Def 1.2.1 σ algebra ($\mathcal{B}$) or Borel field follows these axioms

1) $\emptyset \in \mathcal{B}$
2) If $A \in \mathcal{B}$ then $A^c \in \mathcal{B}$ (closed under complementation)
3) If $A_1, A_2 \ldots \in \mathcal{B}$ then $\bigcup_i^\infty A_i \in \mathcal{B}$ (closed under countable unions)

Def 1.2.4 A probability function w/ S = samplespace $\mathcal{B}$ = σ algebra domain = $\mathcal{B}$ $\Rightarrow$

Kolmogorov axioms
(1) $P(A) \geq 0 \ \forall A \in \mathcal{B}$
(2) $P(S) = 1$
(3) If $A_1, A_2 \ldots \in \mathcal{B}$ are pairwise disjoint (if $A_i \cap A_j = \emptyset \ \forall i \neq j$), then $P(\bigcup_i^\infty A_i) = \sum_i^\infty P(A_i)$

Axiom of Finite additivity $\rightarrow$ (3R) If $A \in \mathcal{B}$ & $B \in \mathcal{B}$ are disjoint then $P(A \cup B) = P(A) + P(B)$

Theorem 1.2.6 Let $S = \{s_1, s_2, \ldots s_n\}$ and let $\mathcal{B}$ be any sigma algebra of subsets of S. Let $p_1, p_2, \ldots p_n \geq 0 \ni \sum p_i = 1$. Then $\forall A \in \mathcal{B}$ $P(A) = \sum_{i|s_i \in A} p_i$

$\boxed{\text{Proof}}$
- Axiom 1: $\forall A \in \mathcal{B}$, $P(A) = \sum_{i|s_i \in A} p_i \geq 0$ (by $p_1, \ldots \geq 0$)
- Axiom 2: $P(S) = \sum_{i \in S} p_i = \sum_i^n p_i = 1$ //
- Axiom 3: Let $A_1, \ldots, A_k$ denote pairwise disjoint events then $P(\bigcup_i^k A_i) = \sum_{j=1}^k \sum_{i|s_i \in A_j} p_i = \sum_i^k P(A_i)$ (Def pairwise disjoint) //

Theorem 1.28 If P is a probability function and A is any set in $\mathcal{B}$ then

a) $P(\emptyset) = 0$
b) $P(A) \leq 1$
c) $P(A^c) = 1 - P(A)$

$\boxed{\text{Proof}}$
- A and A^c form a partition so $S = A \cup A^c$ so $P(A \cup A^c) = P(S) = 1$ (by Kolmogorov 2)
- Since A and A^c are disjoint $1 = P(A \cup A^c) = P(A) + P(A^c)$
  $\therefore P(A^c) = 1 - P(A)$ //c
- $S \cup \emptyset = S$ so $P(S \cup \emptyset) = P(S) = 1$ (Kolmogorov 2)
- Since S and $\emptyset$ are disjoint $1 = P(S \cup \emptyset) = P(S) + P(\emptyset)$
  $\therefore P(\emptyset) = 0$ //a
- Since $P(A^c) = 1 - P(A)$ $P(A) = 1 - P(A^c)$
  $P(A^c) \geq 0$ (by 1.24 a)
  $\therefore P(A) \leq 1$ //b

Bonferroni's Inequality $P(A \cap B) \geq P(A) + P(B) - 1$

NOTE: USEFUL FOR LARGE P(A) and P(B)

$\boxed{\text{Proof}}$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(A \cap B) = P(A) + P(B) - P(A \cup B)$ (rearranging)
$P(A \cap B) \geq P(A) + P(B) - 1$ (b/c $P(A \cup B) \leq 1$) (1.24 a)

Theorem 1.2.9 If P is a probability function and A and B are sets in $\mathcal{B}$ then

a. $P(B\cap A^c) = P(B) - P(A\cap B)$

b. $P(A\cup B) = P(A) + P(B) - P(A\cap B)$

c. If $A\subseteq B$ then $P(A)\leq P(B)$

$\boxed{\text{Proof}}$

• $P(B) = P((B\cap A)\cup(B\cap A^c))$ [b/c pg 2 $B=(B\cap A)\cup(B\cap A^c)$]

$= P(B\cap A) + P(B\cap A^c)$ [b/c they are disjoint]

$\therefore P(B\cap A^c) = P(B) - P(B\cap A)$ //a

• $P(A\cup B) = P(A) + P(B\cap A^c)$ [b/c pg 2 and A, A^c are disjoint $A\cup B = A\cup(B\cap A^c)$]

$= P(A) + P(B) - P(A\cap B)$ by a //b

• If $A\subseteq B$ then $A\cap B = A$

$0\leq P(B\cap A^c) = P(B) - P(A)$ (by above and a + def 1.2.4)

$\therefore P(A)\leq P(B)$ //c

Theorem 1.2.11 If P is a probability function then

a) $P(A) = \sum_1^\infty P(A\cap C_i)$ for any partition $C_1, \dots,$ on S

b) $P(\bigcup_1^\infty A_i) \leq \sum_1^\infty P(A_i)$ for $A_1, A_2, \dots$

NOTE $A\setminus B = A\cap B^c$

$\boxed{\text{Proof}}$

• Since C_i forms a partition $C_i\cap C_j = \emptyset\ \forall i\neq j$ and $\bigcup_1^\infty C_i = S$ so $A = A\cap S \overset{\text{by distrib}}{=} A\cap(\bigcup_1^\infty C_i) = \bigcup_1^\infty(A\cap C_i)$

$\therefore P(A) = \sum_1^\infty P(A\cap C_i)$ //a

• Let $A_1^*, A_2^* \dots$ be a disjoint collection $\ni \bigcup_1^\infty A_i^* = \bigcup_1^\infty A_i$

• $A_1^* = A_1$, $A_i^* = A_i \setminus (\bigcup_1^{i-1} A_j)$ $i = 2, 3, \dots$

$\therefore P(\bigcup_1^\infty A_i) = P(\bigcup_1^\infty A_i^*) = \sum_1^\infty P(A_i^*)$ since (A_i^* are disjoint)

• $A_i^*\subseteq A_i$ so $P(A_i^*)\leq P(A_i)$

doesn't A_i have to be disjoint? → $\therefore \sum_1^\infty P(A_i^*)\leq \sum_1^\infty P(A_i)$

$\therefore P(\bigcup_1^\infty A_i)\leq \sum_1^\infty P(A_i)$ for $A_1, A_2, \dots$ //b

Theorem 1.2.14 If a job consists of K separate tasks the i^{th} of which can be done in n_i ways $\forall i = 1, 2, \dots, K$ then the entire job can be done in $n_1\times n_2\times\dots\times n_K$ ways

$\boxed{\text{Proof}}$ Base cases: $K=1$ we can do 1 job n_1 ways

$K=2$ n_1, n_2 ways → $n_1\times n_2$ ways by counting

Assume: $K=n$ $n_1, n_2, \dots n_n$ ways → $n_1\times n_2\times\dots\times n_n$ ways

Show: $K=n+1$ $n_1, n_2 \dots n_n, n_{n+1}$ ways → $n_1\times n_2\times\dots\times n_n\times n_{n+1}$ ways

K_{n+1} ways = (ways to do K_n ways) × (n_{n+1}) (by counting)

$= (n_1\times n_2\times\dots\times n_n)\times n_{n+1}$ // EMS

Counting w/ replacement - choice is the same each time you make it (i.e. chosen items are replaced restoring original set)

Counting w/o replacement - choice is different each time as the original set changes w/ every choice

Ordered - the order of the choices matter (Powerball)

Unordered - the order of the choices doesn't matter (3 play w box option)

Def 1.2.16 For $n>0$ $n! = n\cdot(n-1)\cdot(n-2)\cdots 1$. For $n=0$ $0!=1$

Counting Procedures w/ n items and r selections

- Ordered w/o replacement: $_nP_r = \frac{n!}{(n-r)!}$
- Ordered w/ replacement: n^r
- Unordered w/o replacement: $_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$
- Unordered w/ replacement: $_{n+r-1}C_r = \frac{(n+r-1)!}{r!(n-1)!}$

Def if all outcomes are equally likely ie $P(s_i) = \frac{1}{n}$ $\forall s_1, s_2, \ldots, s_n$ then

$$P(A) = \sum_{s_i \in A} P(s_i) = \sum \frac{1}{n} = \frac{\#\text{ of elements in } A}{\#\text{ of elements in } S}$$

$\boxed{1.3}$

Def 1.3.2 • $P(A|B)$ = Prob of A given that B has occured = $\frac{P(A\cap B)}{P(B)}$

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

Def 1.3.5 • $P(A|B) = P(B|A)\frac{P(A)}{P(B)}$ by rearranging the above

Theorem 1.3.5 Bayes' Rule: Let $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum P(B|A_j)P(A_j)}$ for A_i a partition of S

$\boxed{\text{Proof}}$ $P(A|B) = \frac{P(A\cap B)}{P(B)}$, $P(B|A) = \frac{P(A\cap B)}{P(A)}$ by def 1.3.5 and disjointness

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

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

$P(B) = \sum^{\infty} (B\cap A_i)$ (by 1.2.11 b/c A is a partition)

$(B\cap A_i) = P(B|A)P(A)$ (by 1.3.2)

$P(A\cap B) = P(B|A)P(A)$ (by 1.3.2)

$\therefore P(B|A)P(A) / \sum P(B|A)P(A)$ by substitution //ems

Def 1.3.7 A, B are statistically independent if $P(A\cap B) = P(A)P(B)$

$\boxed{\text{Proof}}$ $P(A\cap B) = P(A|B)P(B)$

$P(A|B) = P(A)$ if A, B are independent

$\therefore$ If independent $P(A\cap B) = P(A)P(B)$ //ems

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

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

Theorem 1.3.9: If A and B are independent events then the following are also independent
a) A and B^c
b) A^c and B
c) A^c and B^c

$\boxed{\text{Proof}}$ • $P(A \cap B^c) = P(A)P(B^c)$ (Want to show)

$P(A \cap B^c) = P(A) - P(A \cap B)$ (Theorem 1.2.9a)
$= P(A) - P(A)P(B)$ (b/c A & B are ind)
$= P(A)(1 - P(B))$ (dist)
$= P(A)P(B^c)$ (complement) //a

• $P(B \cap A^c) = P(B) - P(B \cap A)$
$= P(B) - P(B)P(A)$ (b/c A & B are ind)
$= P(B)(1 - P(A))$ (dist)
$= P(B)P(A^c)$ (complement) //b

• $P(A^c \cap B^c) = P(A^c)P(B^c)$ (WTS)
$= P(A^c) - P(A^c \cap B)$ (artful complication)
$= P(A^c) - P(A^c)P(B)$ (by b)
$= P(A^c)(1 - P(B))$ (distribution)
$= P(A^c)P(B^c)$ (complement) //

Definition 1.3.12: A collection of events $A_1, \ldots, A_n$, mutually independent for any sub collection $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ then we have: $P(\bigcap_1^k A_{i_j}) = \prod_1^k P(A_{i_j})$

1.3 w/ σ algebras

• from $(\underset{\text{sample}}{S}, \underset{\sigma\text{ algebra}}{\mathcal{B}}, P) \rightarrow P(A|B) = P(A \cap B)/P(B)$ Assuming $P(B) > 0$
= from $(\underset{\text{subset provided by given event}}{B}, \underset{\sigma\text{ algebra } \subseteq \mathcal{B}}{\mathcal{B}^*}, P_B) \rightarrow P(A|B) = P_B(A \cap B)$ $\forall A \cap B \in \mathcal{B}^*$

Add to 1.2.11 • $P(A) = \sum_1^\infty P(A \cap C_i)$ for partition C_i (by theorem 1.2.11)
$= \sum_1^\infty P(A|C_i)P(C_i)$ (Multiplication Rule)

• Pairwise independent if: 1) $P(A \cap B) = P(A)P(B)$
2) $P(A \cap C) = P(A)P(C)$
3) $P(B \cap C) = P(B)P(C)$

• Mutually independent if: 1–3
4) $P(A \cap B \cap C) = P(A)P(B)P(C)$

1.4

Random Variable:

- $X: S \Rightarrow \mathbb{R}$ & $X^{-1}(B) = \{w \in S \mid X(w) \in B\} \in \mathcal{B} \;\; \forall B \in \mathcal{B}(\mathbb{R})$
- $(S, \mathcal{B}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_x)$
- Domain $\to$ Range
- NOTE 1) $\sigma(x) = \{x^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$ is a σ algebra on S
  2) $x^{-1}(B) = \{w \in S \mid X(w) \in B\} \in \mathcal{B} \;\forall B \in \mathcal{B}(\mathbb{R})$ says all $\{x \in B\}$ (range) can be assigned a probability according to the measure P on $(S, \mathcal{B})$
  3) X is a $\mathcal{B}-\mathcal{B}(\mathbb{R})$ measurable mapping from S to $\mathbb{R}$
- Support of a random variable is the range ie $\mathcal{X} = \{x_1, \dots, x_m\}$
  - finite support if $m < \infty$
  - countably infinite support if $m = \infty$
- Induced Measure $P_x(X = x_i) = P(\omega \in S \mid X(w) = x_i)$ measured on $\mathbb{R}, \mathcal{B}(\mathbb{R})$ 'the range'

1.5

Cumulative Distribution Function: $F_X(x) = P(X \leq x)$ ← induced measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

NOTE: $F(x)$ is defined on all $x \in \mathbb{R}$ not just $X \subset \mathcal{X}$

Theorem 1.5.3 - $F_X : \mathbb{R} \to [0,1]$ is a cdf iff

a) $\lim_{x \to -\infty} F_X(x) = 0$ & $\lim_{x \to \infty} F_X(x) = 1$

b) $F_X(x)$ is non decreasing

c) $F_X(x) = P(X \leq x)$ is right continuous that is $\lim_{x \to x_0^+} F(x) = F(x_0) \;\forall x_0$

Proof a) $\lim_{x \to -\infty} F_X(x) = \lim_{x \to -\infty} P(X \leq x) = 0$

$\lim_{x \to \infty} F_X(x) = \lim_{x \to \infty} P(X \leq x) = 1$

b) Let $x_0 \leq x_1$ then $F(x_0) = P(X \leq x_0) \leq P(X \leq x_1) = F(x_1)$

$\therefore F(x)$ is non decreasing

c) $F(x) = P(X \leq x)$ is right continuous because we define $F(x)$ using '$\leq$' so even if $F(x)$ is discontinuous (a step function) $\lim_{x \to x_0^+} F(x) = F(x_0)$ //

Definition 1.5.8 The random variables X and Y are identically distributed if for every set $A \in \mathcal{B}^1$, $P(X \in A) = P(Y \in A)$ where $X = Y$ or $X \neq Y$. Denote $(X \stackrel{d}{=} Y)$

Definition 1.5.7 A random variable is continuous if $F(x)$ is continuous and discrete if $F(x)$ is a step function

Theorem 1.5.10 $F_X(x) = F_Y(Y) \equiv X \overset{c}{=} Y$ $\boxed{\text{Proof}}$ $\Rightarrow F_X(x) = P(X \le x) = P(Y \le y) = F_Y(x)\ \forall x$

$\Leftarrow$ Too hard to show

$\boxed{1.6}$

Def 1.6.1 • The <u>Probability Mass Function</u> of a discrete random variable X is given by

$$\underset{PMF}{f_X(x)} = P(X = x) = \underset{CDF}{F_X(x)} = \sum_{t \mid t \le x} \underset{PMF}{f_X(t)}$$

• If X is a continuous r.v. w/ CDF $F_X(x)$ then $P_X(X=x) = 0\ \forall x \in \mathbb{R}$

$\boxed{\text{Proof}}$ $\{X = x\} \subseteq \{x - \epsilon < X \le x\}$ $\ni$ We fix $\epsilon > 0$

$P(X=x) \le P(x-\epsilon < X \le x)$

$P(X=x) = F_X(x) - F_X(x-\epsilon)$

$= 0$ as $\epsilon \to 0$ //

Def 1.6.3 • The <u>Probability Density Function</u> of a (absolutely) continuous random variable X is given by

$$\underset{CDF}{F_X(x)} = \int_{-\infty}^{x} f(x)\,dx$$

Theorem 1.6.5 • <u>Properties of PDF/PMF</u> valid iff

a) $f_X(x) \ge 0\ \forall x \in \mathbb{R}$

b) discrete: $\sum_{x \in \mathbb{R}} f_X(x) = 1$

continuous: $\int_{\mathbb{R}} f_X(x)\,dx = 1$

$\boxed{\text{Proof}}$ $\Rightarrow$ discrete: $f_X(x) = P(X = x) \ge 0$

$\sum_{\mathbb{R}} f_X(x) = \sum_{\mathbb{R}} P_X(X=x) = 1$

continuous: $f_X(x) \ge 0$ (b/c = derivative of CDF is non decreasing

$\int_{\mathbb{R}} f_X(x)\,dx = \lim_{x \to \infty} F_X(x) = 1$ //

$\Leftarrow$

2.1

Binomial Transformation $f_X(x) = P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots n$

Uniform Transformation $f_X(x) = 1/2\pi \quad 0 < x < 2\pi$

Theorem 2.1.3 — nice cdf → cdf

Let X have cdf $F_X(x)$, let $Y = g(x)$ and let $\mathcal{X} = \{x \mid f_X(x) > 0\}$ & $\mathcal{Y} = \{y \mid y = g(x)\ x \in \mathcal{X}\}$

Then a) If g is an increasing function on $\mathcal{X}$ $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$

b) If g is a decreasing function on Continuous Variable $\mathcal{X}$ $F_Y(y) = 1 - F_X(g^{-1}(y))$ $y \in \mathcal{Y}$

Theorem 2.1.5 — nice pdf → pdf

Let X have pdf $f_X(x)$ and $Y = g(x) \ni g(x)$ is a monotone function. Let $\mathcal{X} = \{x \mid f_X(x) > 0\}$ & $\mathcal{Y} = \{y \mid y = g(x)\ x \in \mathcal{X}\}$. Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ and that $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$ then

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & ow \end{cases}$$

↑ This comes directly from theorem 2.1.3

Theorem 2.1.8 — bad pdf → pdf

Let X have pdf $f_X(x)$, let $Y = g(x)$ and $\mathcal{X} = \{x \mid f_X(x) > 0\}$. Suppose $\exists$ a partition $A_0, A_1, A_2, \dots A_n$ of $\mathcal{X} \ni P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Also suppose $\exists$ functions $g_1(x), \dots g_n(x)$ on $A_1(x), \dots A_n$ respectively $\ni$

i) $g(x) = g_i(x) \quad \forall x \in A_i$

ii) $g_i(x) =$ is monotone on A_i

iii) The set $\mathcal{Y} = \{y \mid y = g_i(x)$ for some $x \in A_i\}$ is the same for each $i = 1, \dots k$

iv) $g_i^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$ $\forall i = 1, \dots, k$

$$f_Y = \begin{cases} \sum_{1}^{k} f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & ow \end{cases}$$

Theorem 2.1.10 Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(x)$ then Y is uniformly distributed on $(0,1)$ that is $P(Y \le y) = y$ for $0 \le y \le 1$

- Proof pg 54

2.2

Def 2.2.1 Expected Value : $E(x)$ = The mean of a random variables

- Discrete : $E(g(x)) = \sum_{x \in \mathcal{X}} g(x) f_X(x)$
- Continuous: $E(g(x)) = \int_{-\infty}^{\infty} g(x) f_X(x)$

Theorem 2.2.5 Let X be a random variable and let a, b, c be constants then for any functions g_1 and g_2 whose expectations exist (finite)

a) $E(ag_1(x) + bg_2(x) + c) = aE(g_1(x)) + bE(g_2(x)) + c$

b) if $g_1(x) \ge 0 \ \forall x$ then $E(g_1(x)) \ge 0$

c) if $g_1(x) \ge g_2(x) \ \forall x$ then $E(g_1(x)) \ge E(g_2(x))$

d) If $a \le g_1(x) \le b \ \forall x$ then $a \le Eg_1(x) \le b$

- For $Y = g(x)$:

$E(Y) = \int_{-\infty}^{\infty} Y f_Y(y) \, dy$

$E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx$

2.3

Def 2.3.1

- K^{th} moment $= \mu'_k = E(X^k)$
- K^{th} central moment $= \mu_k = E((X-\mu)^k)$

Def 2.3.2

- The variance of a random variable X is its second central moment

$Var(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$

- The positive square root of the $var(x)$ is the std. deviation

Theorem 2.3.4 $Var(aX + b) = a^2 Var X$

$\boxed{\text{Proof}}$ $E((aX+b) - E(aX+b))^2$

$= E(aX - aEX)^2$

$= b^2 E(X - E(x))^2$

$= a^2 var(x)$ //

Def 2.3.6 Let X be a random variable with cdf F_X. The moment generating function, mgf, of X (or F_X) denoted by $M_X(t)$ is $M_X(t) = E(e^{tX})$ provided the expectation exists for t in some neighborhood of 0.
ie $\exists\, h > 0 \ni \forall t \in (-h, h)$ $E(e^{tX})$ exists, otherwise DNE

Continuous: $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$

Discrete: $M_X(t) = \sum_x e^{tx} f_X(x)\,dx$

Theorem 2.3.7 If X has mgf $M_X(t)$ then $E(X^n) = M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\big|_0$

Proof pg (62)

Theorem 2.3.11 Let $F_X(x)$ and $F_Y(y)$ be two cdfs whose moments exist

a) If X and Y have bounded support then $F_X(u) = F_Y(u)$ $\forall u$ iff $E(X^r) = E(Y^r)$ for $r = 0, 1, 2, \dots$

b) If the mgfs exist and $M_X(t) = M_Y(t)$ $\forall t$ in some neighborhood of 0 then $F_X(u) = F_Y(u)$ $\forall u$

Theorem 2.3.12 Suppose $\{X_i, i = 1, 2, \dots\}$ is a sequence of random variables each with mgf $M_{X_i}(t)$. Further suppose that $\lim_{i\to\infty} M_{X_i}(t) = M_X(t)$ $\forall t$ in a neighborhood of zero and $M_X(t)$ is an mgf. Then $\exists$ a unique cdf F_X whose moments are determined by $M_X(t)$ and, $\forall x \ni F(x)$ is continuous we have $\lim_{i\to\infty} F_{X_i}(x) = F_X(x)$

Lemma 2.3.14 Let $a_1, a_2, \dots$ be a sequence of numbers converging to a, that is, $\lim_{n\to\infty} a_n = a$ then $\lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$

Theorem 2.3.15 For any constants a and b the mgf of the random variable $aX+b$ is $M_{aX+b}(t) = e^{bt} M_X(at)$

Proof p68

2.4

Theorem 2.4.1 If $f(x,\theta)$, $a(\theta)$ and $b(\theta)$ are differentiable w/ respect to θ then
$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,dx = f(b(\theta),\theta)\frac{d}{d\theta}b(\theta) - f(a(\theta),\theta)\frac{d}{d\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)\,dx$$

Leibnitz's Rule ** when $a(\theta)$ and $b(\theta)$ are constants $\frac{d}{d\theta}\int_a^b f(x,\theta)\,dx = \int_a^b \frac{\partial}{\partial\theta} f(x,\theta)\,dx$ **

Theorem 2.4.2 Suppose $h(x,y)$ is continuous at y_0 $\forall$ x and $\exists$ a function $g(x)$ that satisfies

i) $|h(x,y)| \leq g(x)$ $\forall$ x and y

ii) $\int_{-\infty}^{\infty} g(x)\,dx < \infty$

then $\lim_{y \to y_0} \int_{-\infty}^{\infty} h(x,y)\,dx = \int_{-\infty}^{\infty} \lim_{y \to y_0} h(x,y)\,dx$

Theorem 2.4.3 Suppose $f(x,\theta)$ is differentiable at $\theta = \theta_0$, that is $\lim_{\delta \to 0} \frac{f(x,\theta+\delta) - f(x,\theta_0)}{\delta} = \frac{\partial}{\partial\theta} f(x,\theta)\big|_{\theta=\theta_0}$ exists $\forall x$ and $\exists$ a function $g(x,\theta_0)$ and a constant $\delta_0 > 0$ $\ni$

i) $\frac{f(x,\theta+\delta) - f(x,\theta_0)}{\delta} \leq g(x,\theta_0)$ $\forall$ x and $|\delta| \leq \delta_0$

ii) $\int_{-\infty}^{\infty} g(x,\theta_0)\,dx < \infty$

then $\frac{\partial}{\partial\theta} \int_{-\infty}^{\infty} f(x,\theta)\,dx \big|_{\theta=\theta_0} = \int \left[\frac{\partial}{\partial\theta} f(x,\theta)\big|_{\theta=\theta_0}\right] dx$

Corollary 2.4.4 Suppose $f(x,\theta)$ is differentiable in θ and $\exists$ $g(x,\theta) \geq \left|\frac{\partial}{\partial\theta} f(x,\theta)\right|_{\theta=\theta'}$ $\forall$ $\theta' \ni |\theta'-\theta| \leq \delta_0$ and $\int_{-\infty}^{\infty} g(x,\theta)\,dx < \infty$ 2.4.3 holds

Theorem 2.4.8 Suppose that the series $\sum_0^{\infty} h(\theta,x)$ converges for all θ in an interval (a,b) of real numbers and

i) $\frac{\partial}{\partial\theta} h(\theta,x)$ is continuous in θ for each x

ii) $\sum_0^{\infty} \frac{\partial}{\partial\theta} h(\theta,x)$ converges uniformly on every closed and bounded subinterval of (a,b)

Then $\frac{\partial}{\partial\theta} \sum_0^{\infty} h(\theta,x) = \sum_0^{\infty} \frac{\partial}{\partial\theta} h(\theta,x)$

Theorem 2.4.10 Suppose the series $\sum_{x=0}^{\infty} h(\theta,x)$ converges uniformly on $[a,b]$ and that $\forall$ x, $h(\theta,x)$ is a continuous function of θ then

$$\int_a^b \sum_0^{\infty} h(\theta,x)\,d\theta = \sum_0^{\infty} \int_a^b h(\theta,x)\,dx$$

2.6

- A distribution is not necessarily determined by its moments
- Cumulant generating function $= \log[M_X(t)]$
- Factorial mgf $= E(t^X)$

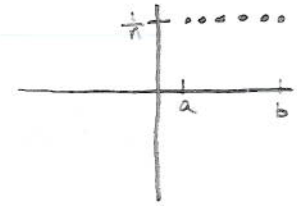
Chapter 3

Family - type of distribution whose members are decided by parameters

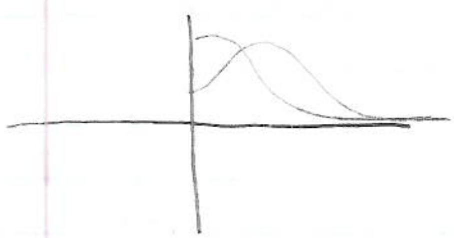
Discrete Distribution if it's support set is countable

Discrete Uniform $(1, N)$, $X = 1, 2, \dots, N$, $P(X=x|N) = \frac{1}{N}$

- This distribution puts equal weights on each outcome
- $E(x) = \sum_1^N x\, P(X=x|N) = \frac{N+1}{2}$
- $E(x^2) = \sum_1^N x^2 P(X=x|N) = \frac{(N+1)(2N+1)}{6}$
- $Var(x) = \frac{(N+1)(N-1)}{12}$

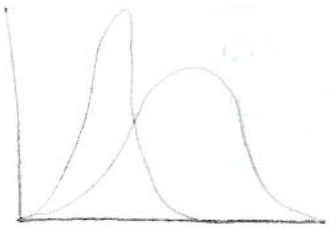


Hypergeometric Distribution



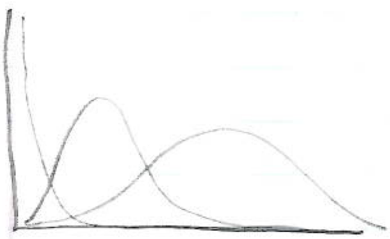
- This distribution is used in the 'urn model'
- We have an urn w/ N balls (M red, N-M green) we then select K balls at random - what is the probability that x of the K balls are red
  - # of samples of size K $\binom{N}{K}$
  - # of ways x of M balls are red $\binom{M}{x}$ and K-x green
- $P(X=x|N,M,K) = \left[\binom{M}{x}\binom{N-M}{K-x}\right] / \binom{N}{K}$ $\quad X = 0, 1, \dots, K$
- $E(x) = \sum_0^K x\, P(X=x|N,M,K)$
- $E(x^2) = \sum_0^K x^2 P(X=x|N,M,K)$
- $Var(x) = \frac{KM}{N}((N-M)(N-K))/(N(N-1))$

Binomial Distribution



- Based on Bernoulli trials (Success/fail)
- $y = \begin{cases} 1 & \text{w/ probability } p \\ 0 & \text{w/ probability } 1-p \end{cases}$ // success // failure
- $P(Y=y \mid n, p) = \binom{n}{y} p^y (1-p)^{n-y}$ $y = 0, 1, 2 \dots n$
- $E(x) = np$
- $var(x) = np(1-p)$
- All trials must be identical, independent and of fixed size
- $M_x(t) = [(1-p) + pe^t]^n$

Poisson Distribution



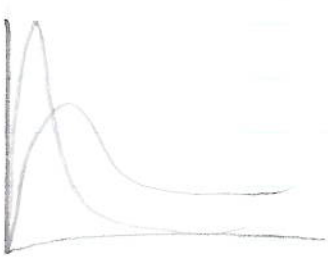
- used for wait time and spatial statistics
- $P(X=x \mid \lambda) = e^{-\lambda} \lambda^x / x!$ $x = 0, 1, \dots$ λ = intensity
- $E(x) = \lambda$
- $var(x) = \lambda$
- $M_x(t) = e^{\lambda(e^t - 1)}$

Negative Binomial Distribution

- Just about the opposite of a binomial
- Concerned with # of Bernoulli trials to see x successes

1. Let X = trial at which the r^{th} success occurs
   - $P(X=x \mid r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$ $x = r, r+1, \dots$
2. Let X = # of failures before r^{th} success occurs
   - $P(X = x \mid r, p) = \binom{r+y-1}{y} p^r (1-p)^y$ $y = 0, 1, \dots$
   - $E(X) = r\frac{(1-p)}{p}$
   - $var(x) = r\frac{(1-p)}{p^2}$ $M_x(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$

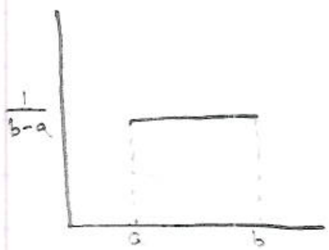
Geometric Distribution



- Special case of negative binomial $r=1 \rightarrow$ trials until 1^{st} success
- $P(X=x \mid p) = p(1-p)^{x-1}$ $x = 1, 2, 3, \dots$
- $E(x) = \frac{1}{p}$
- $var(x) = \frac{1-p}{p^2}$
- $M_x(t) = \frac{pe^t}{1-(1-p)e^t}$
- memoryless

Continuous Distributions

Uniform Distribution



- Spreads mass uniformly over the interval $[a,b]$
- $f_X(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{o.w.} \end{cases}$
- $E(X) = \frac{b+a}{2}$
- $var(X) = \frac{(b-a)^2}{12}$
- $M_X(t) = (e^{bt} - e^{\alpha t}) / (b-\alpha)t$

Gamma Distribution

- gamma function = $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt \qquad \alpha > 0 \in \mathbb{Z}^+$
  - $\Gamma(\alpha+1) = \alpha(\Gamma(\alpha))$
  - $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

- $f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \qquad 0 < x < \infty \qquad \alpha, \beta > 0$
  - α shape
  - β scale
- $E(X) = \alpha\beta$
- $var(X) = \alpha\beta^2$
- $M_X(t) = \left(\frac{1}{1-\beta t}\right)^\alpha \qquad t < \frac{1}{\beta}$

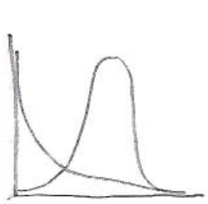
Chi-Squared

- Special case of gamma $\alpha = p/2 \quad \beta = 2$
- $f(x|p) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-x/2} \qquad 0 < x < \infty$
- $E(X) = p$
- $Var(X) = 2p$
- $M_X(t) = \left(\frac{1}{1-2t}\right)^{p/2}$

Exponential

- Special case of gamma $\alpha = 1 \quad \beta = \beta$
- $f_X(x|\beta) = \frac{1}{\beta} e^{-x/\beta} \qquad 0 < x < \infty$
- $E(X) = \beta$
- $var(X) = \beta^2$
- memoryless

Weibull Used for modeling hazard functions



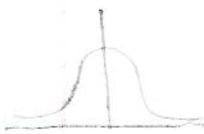
$f(x|\beta) = \frac{\gamma}{\beta} x^{\gamma-1} e^{-x^{\gamma}/\beta} \quad x>0 \quad \gamma, \beta > 0$

$E(x) = \beta^{1/\gamma} \Gamma\left(1+\frac{1}{\gamma}\right)$

$var(x) = \beta^{2/\gamma}\left[\Gamma\left(1+\frac{2}{\gamma}\right) - \Gamma^2\left(1+\frac{1}{\gamma}\right)\right]$

$M_x(t) = \beta^{n/\gamma}\Gamma\left(1+\frac{n}{\gamma}\right) \quad \gamma \geq 1$

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



$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

$Z = \frac{x-\mu}{\sigma} \sim N(0,1)$ Standard Normal

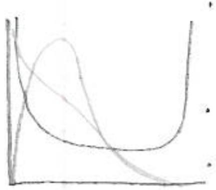
$E(x) = \mu$

$var(x) = \sigma^2$

We can use the normal distribution to approximate the binomial when heuristic $\min(np, np(1-p)) \geq 5$ ie large n and small p

Binomial $(n,p) \sim N(np, np(1-p))$

Beta ·for interval data



$\cdot f_x(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$; $0<x<1 \quad \alpha>0 \quad \beta>0$

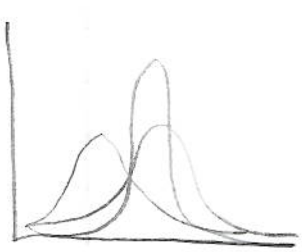
$\ni B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$\cdot E(x) = \frac{\alpha}{\alpha+\beta}$

$\cdot var(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$\cdot M_x(t) = 1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!}$

Cauchy Dist ·symmetric bellshaped on $(-\infty, \infty)$

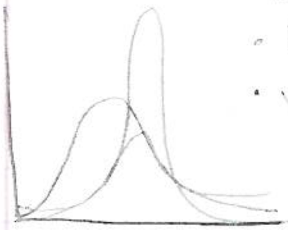


$\cdot f_x(x|\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2} \quad -\infty < x < \infty \quad -\infty < \theta < \infty$

$\cdot E(x)$ DNE – We use θ as the center of the distribution

$P(X \leq \theta) = .5$; $P(X \geq \theta) = .5$

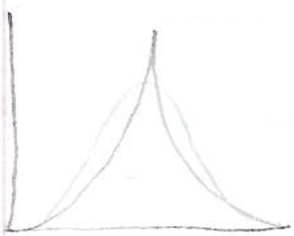
Log Normal: If $\log(x) \sim N(\mu, \sigma^2)$ * Good for data that is skewed right



- $f(x|\mu,\sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)\left(\frac{1}{x}\right) e^{-(\log(x)-\mu)^2/2\sigma^2}$ $0 < x < \infty$ $-\infty < \mu < \infty$ $\sigma > 0$
- $E(x) = e^{\mu + (\sigma^2/2)}$
- $var(x) = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$

$M_x(t) = E(x^n) = e^{n\mu + \frac{n^2\sigma^2}{2}}$

Double Exponential · Given by reflecting the exponential distribution around its mean



- $f_x(x|\mu,\sigma) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}}$; $\sigma > 0$
- $E(x) = \mu$
- $var(x) = 2\sigma^2$
- $M_x(t) = \frac{e^{\mu t}}{1-(\sigma t)^2}$

Exponential Families · distributions that can be expressed as

- $f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right)$; · $h(x) \geq 0$
  - $t_1(x), \dots, t_k(x)$ are real valued funct.
  - $\ni$ no $t_i(x)$ depends on θ
- normal, gamma, beta, binomial, poisson and negative binomial are exponential families.
- to show this we must identify the functions $h(x)$, $c(\theta)$, $w_i(\theta)$, $t_i(x)$ that can create the previous form

Theorem 3.4.2 If X is a random variable with pdf or pmf of an exponential family then

i) $E\left(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial\theta_j} t_i(X)\right) = \frac{\partial}{\partial\theta_j}\log c(\theta)$

ii) $Var\left(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial\theta_j} t_i(X)\right) = -\frac{\partial^2}{\partial\theta_j^2}\log c(\theta) - E\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\theta)}{\partial\theta_j^2} t_i(X)\right)$

Def 3.4.7 • A <u>curved exponential family</u> is an exponential family for which the dimension of vector θ is $d < K$

• A <u>full exponential family</u> is an exponential family for which the dimension of vector θ is $d = K$

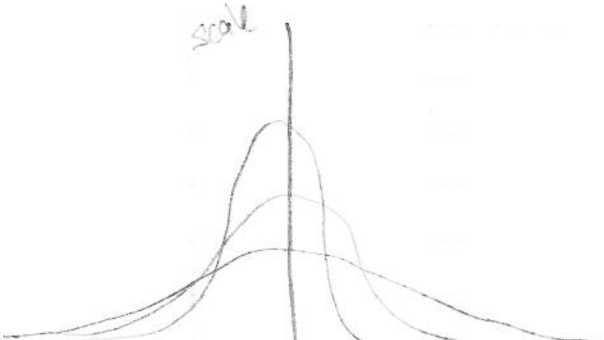
Theorem 3.5.1 Let $f(x)$ be a pdf and let $\mu, \sigma > 0$ be given constants then

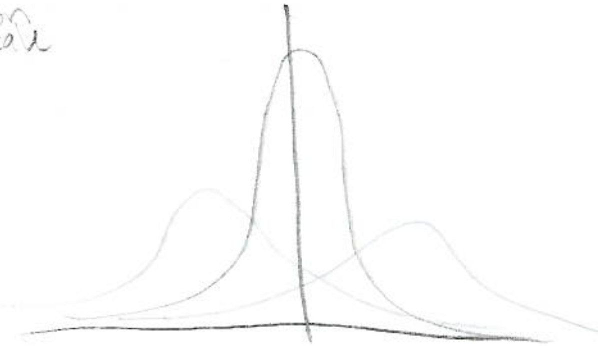
$$g(x|\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) \quad \text{is a pdf}$$

Def 3.5.2 Let $f(x)$ be any pdf Then the family of pdfs $f(x-\mu)$; $-\infty < \mu < \infty$ is called the <u>location family</u> w/ <u>standard pdf</u> $f(x)$ and <u>location parameter</u> μ

Def 3.5.4 Let $f(x)$ be any pdf then $\forall\ \sigma > 0$ the family of pdfs $(1/\sigma) f(x/\sigma)$, indexed by σ is called the <u>scale family</u> w/ <u>standard pdf</u> $f(x)$ and σ is called the <u>scale parameter</u> of the family

Def 3.5.5 Let $f(x)$ be any pdf. Then for any μ, $-\infty < \mu < \infty$ and any $\sigma > 0$ the family of pdfs $(1/\sigma) f((x-\mu)/\sigma)$ indexed by (σ, μ) is called the <u>location-scale family</u> with <u>standard pdf</u> $f(x)$ and scale parameter, σ, and location parameter μ.





Theorem 3.6.1 <u>Chebychev's Inequality</u> Let X be a random variable and $g(x)$ be a nonnegative function then $\forall\, r > 0$

$$P(g(x) \geq r) \leq \frac{E(g(x))}{r}$$

Theorem 3.6.4 Let $X_{\alpha,\beta}$ denote a gamma (α,β) random variable w/ pdf $f(x|\alpha,\beta)$ where $\alpha > 1$. Then for constants a and b

$$P(a < X_{\alpha,\beta} < b) = \beta\left(f(a|\alpha,\beta) - f(b|\alpha,\beta)\right) + P(a < X_{\alpha-1,\beta} < b)$$

Lemma 3.6.5 <u>Stein's Lemma</u> Let $X \sim N(\theta,\sigma^2)$ and let g be a differentiable function satisfying $E(|g'(x)|) < \infty$ then

$$E[g(x)(x-\theta)] = \sigma^2 E(g'(x))$$

Theorem 3.6.7 Let χ^2_p denote a chi squared RV w/ p degrees of freedom for any function $h(x)$

$$E(h(\chi^2_p)) = p\, E\left(\frac{h(\chi^2_{p+2})}{\chi^2_{p+2}}\right)$$

Theorem 3.6.8 <u>Hwang</u> let $g(x)$ be a function w/ $-\infty < E(g(x)) < \infty$ and $-\infty < g(-1) < \infty$ then

a) $X \sim \text{Poisson}(\lambda)$ $\quad E(\lambda g(x)) = E(X g(x-1))$

b) $X \sim \text{Negative Binomial}(r,p)$ $\quad E((1-p)g(x)) = E\left(\frac{X}{r+X-1}\, g(x-1)\right)$

Theorem 3.8.1 Poisson Postulates For each $t \geq 0$ let N_t be an integer value random variable with the following properties. (Think of N_t as # of arrivals)

i) $N_0 = 0$ (no arrivals at start)

ii) $s < t \Rightarrow N_s$ and $N_t - N_s$ are independent (arrivals in disjoint time periods are independent)

iii) N_s and $N_{t+s} - N_t$ are identically distributed (depends only on period length)

iv) $\lim_{t \to 0} \frac{P(N_t = 1)}{t} = \lambda$ (arrival probability proportional to period length if length is small)

continued on next page

v) $\lim_{t \to 0} \frac{P(N_t > 1)}{t} = 0$ (no simultaneous arrivals

If i - v hold then for any integer n

$$P(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

Theorem 3.8.2 If $0 < \sigma < \infty$ then

a) If $n=1$ the inequality is attainable for $K \geq 1$ and not attainable for $0 < K < 1$

b) If $n=2$ the inequality is attainable iff $K=1$

c) if $n \geq 3$ the inequality is not attainable

Lemma 3.8.3 Markovs inequality: If $P(Y \geq 0) = 1$ and $P(Y=0) < 1$ then for any $r > 0$ $P(Y \geq r) \leq \frac{E(Y)}{r}$ with equality iff $P(Y=r) = p = 1 - P(Y=0)$ $0 < p \leq 1$

Theorem 3.8.4 Gauss inequality Let $X \sim f$ where f is unimodal with mode v and define $\tau^2 = E(X-v)^2$ then

$$P(|X-v| > \epsilon) \leq \begin{cases} \frac{4\tau^2}{9\epsilon^2} & \forall\ \epsilon \geq \sqrt{4/3}(\tau) \\ 1 - \frac{\epsilon}{\sqrt{3}\tau} & \forall\ \epsilon \leq \sqrt{4/3}(\tau) \end{cases}$$

Theorem 3.8.5 Vysochanskii-Petunin Let $X \sim f$ where f is unimodal define $\xi^2 = E(X-\alpha)^2$ for an arbitrary point α then

$$P(|X-\alpha| > \epsilon) \leq \begin{cases} \frac{4\xi^2}{9\epsilon^2} & \forall\ \epsilon \geq \sqrt{8/3}(\xi) \\ \frac{4\xi^2}{9\epsilon^2} - \frac{1}{3} & \forall\ \epsilon \leq \sqrt{8/3}(\xi) \end{cases}$$

$\boxed{4}$

Definition 4.1.1 A <u>n-dimensional random vector</u> is a function from a sample space into $\mathbb{R}^n$, n-dimensional Euclidean Space

Definition 4.1.3 Let (X,Y) be a discrete bivariate random vector. Then the function $f(x,y)$ from $\mathbb{R}^2$ into $\mathbb{R}$ defined by $f(x,y) = P(X=x, Y=y)$ is called the <u>joint pmf</u> of $(X,Y) \to f_{X,Y}(x,y)$

Theorem 4.1.6 Let (X,Y) be a discrete bivariate random vector w/ joint pmf $f_{X,Y}(x,y)$. Then the <u>marginal pmfs of Y an X</u> are
$f_X(x) = \sum_{y \in \mathbb{R}} f_{X,Y}(x,y)$ $\quad f_Y(y) = \sum_{x \in \mathbb{R}} f_{X,Y}(x,y)$

Definition 4.1.10 A function $f(x,y)$ from $\mathbb{R}^2$ into $\mathbb{R}$ is called a <u>joint pdf</u> of the continuous bivariate RV (X,Y) if $\forall A \subset \mathbb{R}^2$
$P((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$

Definition 4.2.1 $\boxed{f(y|x)} = P(Y=y \mid X=x) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$

$$\ni \sum_y f(y|x) = 1 \quad \text{or} \quad \int_y f(y|x) = 1 \quad \text{\&} \quad f_X(x) > 0$$

Definition 4.2.5 X, Y are called <u>independent random variables</u> if $\forall\, x \in \mathbb{R},\ y \in \mathbb{R}$ $\quad f_{X,Y}(x,y) = f_X(x) f_Y(y)$ and $f(y|x) = f_Y(y)$

Lemma 4.2.7 Let (X,Y) be a bivariate random vector w/ joint pdf or pmf $f_{X,Y}(x,y)$ then X and Y are independent iff $\exists\ g(x)$ & $h(x)$ $\ni f(x,y) = g(x)h(y)$

proof: $\Rightarrow$ $\int_{-\infty}^{\infty} g(x)\,dx = c$ & $\int_{-\infty}^{\infty} h(y)\,dy = d$

$$\begin{aligned} cd &= \left(\int_{-\infty}^{\infty} g(x)\,dx\right)\left(\int_{-\infty}^{\infty} h(y)\,dy\right) \\ &= \iint_{\mathbb{R}^2} g(x)h(x)\,dx\,dy \\ &= \iint_{\mathbb{R}^2} f_{X,Y}(x,y)\,dx\,dy \\ &= 1 \end{aligned}$$

$f_X(x) = g(x)d \quad f_Y(y) = h(y)c$
$f(x,y) = g(x)h(y) = g(x)h(y)cd = f_X(x)f_Y(y)$

$\Leftarrow$ simply choose $g(x) = f_X(x)$ and $h(y) = f_Y(y)$

//

Theorem 4.2.10 Let X and Y be independent random variables

a) $\forall$ $A \subset \mathbb{R}$ and $B \in \mathbb{R}$ $P(X \in A, Y \in B) = P(X \in A) P(X \in B)$

the events $\{X \in A\}$ and $\{Y \in B\}$ are independent events

b) Let $g(x)$ be a function only of x and $h(y)$ a function only of y then $E(g(X)h(Y)) = (E(g(X)))(E(h(Y)))$

proof $E(g(X)h(Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(x) f(x,y)\,dx\,dy$

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)\, h(x)\, f_x(x) f_y(y)\, dx\, dy$ (independance)

$= \int_{-\infty}^{\infty} h(y) f_y(y) \int_{-\infty}^{\infty} g(x) f_x(x)\, dx\, dy$

$= \left(\int_{-\infty}^{\infty} h(y) f_y(y)\, dy\right)\left(\int_{-\infty}^{\infty} g(x) f_x(x)\, dx\right)$

$= E(g(X))E(h(Y))$

Theorem 4.2.12 Let X and Y be independent w/ MGFs $M_x(t)$ and $M_y(t)$ then the MGF of $Z = X+Y$ is

$$M_Z(t) = M_x(t)\, M_y(t)$$

proof $M_Z(t) = E(e^{tZ}) = E(e^{t(X+Y)}) = E(e^{tX} e^{tY}) = E(e^{tX}) E(e^{tY}) = M_x(t) M_y(t)$

Theorem 4.2.14 Let $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\gamma, \tau^2)$ be independent then $Z = X+Y \sim N(\mu+\gamma, \sigma^2+\tau^2)$

Theorem 4.3.2 If $X \sim Poisson(\theta)$ and $Y \sim Poisson(\lambda)$ are independent then $X+Y \sim Poisson(\theta+\lambda)$

Bivariate Transformation $A = \{(x,y) \mid f_x(x,y) > 0\}$

$\mathcal{B} = \{(u,v) \mid u = g_1(x,y) \;\&\; v = g_2(x,y)\}$

$u = g_1(x,y)$ $\quad$ $v = g_2(x,y)$

$x = h_1(u,v)$ $\quad$ $y = h_2(u,v)$

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

Then $f_{u,v}(u,v) = f_{x,y}(h_1(u,v), h_2(u,v))\,|J|$

Theorem 4.3.5 Let X and Y be independent. Let $g(x)$ be a function only of x and $h(y)$ be a function of only y. Then $U = g(X)$ and $V = h(Y)$ are independent

Theorem 4.4.3 If X and Y are any two random variables then $E(X) = E(E(X|Y))$ provided the expectations exist

Definition 4.4.4 A random variable X is said to have a mixture distribution if the distribution of X depends on a quantity that also has a distribution.

ie $X|Y \sim$ binomial

$Y|\Lambda \sim$ Poisson(Λ)

$\Lambda \sim$ exponential(β)

Theorem 4.4.7 For any RV X and Y

$$Var(X) = E(Var(X|Y)) + Var(E(X|Y))$$

Definition 4.5.1 The covariance of X and Y

$$Cov(X,Y) = E((X-\mu_x)(Y-\mu_y))$$

Definition 4.5.2 The correlation of X and Y

$$\rho_{xy} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

Theorem 4.5.5 If X and Y are independent then $cov(x,y) = 0$ and $\rho_{xy} = 0$

proof
$$\begin{aligned} Cov(X,Y) = E((X-\mu_x)(Y-\mu_y)) &= E(XY) - \mu_y E(X) - \mu_x E(Y) + \mu_x \mu_y \\ &= E(XY) - \mu_y \mu_x - \mu_x \mu_y + \mu_x \mu_y \\ &= E(XY) - \mu_y \mu_x \\ &\overset{\perp}{=} E(X)E(Y) - \mu_y \mu_x \\ &= \mu_x \mu_y - \mu_y \mu_x \\ &= 0 \end{aligned}$$

Theorem 4.5.3 $\quad Cov(X,Y) = E(XY) - \mu_X \mu_Y$

Theorem 4.5.6 If X and Y are any RV and $a, b \in \mathbb{R}$

$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\, cov(X,Y)$

Note: $2ab\, cov(X,Y) = 0$ if $X \perp Y$

Theorem 4.5.7 For any RV X and Y

a) $-1 \le \rho_{XY} \le 1$

b) $|\rho_{XY}| = 1$ iff $\exists\ a \neq 0$ and $b \ni P(Y = aX + b) = 1$

Note $\quad a > 0$ if $\rho_{XY} = 1$

$\quad a < 0$ if $\rho_{XY} = -1$

Definition 4.5.10 Let $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, $0 < \sigma_X$, $0 < \sigma_Y$ and $-1 < \rho < 1$ the bivariate normal pdf is

$$f(x,y) = \frac{1}{2\sigma_x \sigma_y \sqrt{1-\rho^2}}\, e^{\frac{-1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)}$$

Note
a) the marginal distribution of X is $N(\mu_x, \sigma_x^2)$
b) " " " of Y is $N(\mu_Y, \sigma_Y^2)$
c) $\rho_{XY} = \rho$

* This can be expanded to more than 2 variables *

ie

- Random Vector $(X_1, X_2, \dots X_n)$
- Joint PMF/PDF $\quad f_{\underline{X}}(x_1, x_2, \dots x_n)$
- $P(\underline{X} \in A) = \int\int_A\int f_{\underline{X}}(x_1, x_2 \dots x_n)\, dx_1 dx_2 \cdots dx_n$
- $E(g(\underline{X})) = \int\int_{\mathbb{R}^n}\cdots\int g(\underline{X}) f_{\underline{X}}(x_1, x_2, \dots x_n)\, dx_1 dx_2 \cdots dx_n$
- Marginals: $f(x_1, x_2 \dots x_k) = \int\int\int f_{\underline{X}}(x_1, x_2, \dots, x_n)\, dx_{k+1} dx_{k+2} \cdots dx_n$ $\quad$ continuous

$\qquad = \sum_{x_{k+1}} \sum_{x_{k+2}} \cdots \sum_{x_n} f_{\underline{X}}(x_1, x_2, \dots x_n)$ $\quad$ discrete

- $f_{\underline{X}}(x_{k+1}, x_{k+2}, \dots x_n \mid x_1, x_2 \dots x_k) = \dfrac{f(x_1, x_2 \dots x_n)}{f(x_1, x_2 \dots x_k)}$

Definition 4.6.2 Let n and m be positive integers and let $p_1, p_2, \dots p_n$ be numbers satisfying $0 \le p_i \le 1 \;\forall i = 1,2,\dots n$ and $\sum_i^n p_i = 1$ then the random vector $(X_1, X_2, \dots X_n)$ has a multinomial distribution with m trials and cell probabilities $p_1, p_2, \dots p_n$ and

$$f_{\underline{X}}(x_1, x_2, \dots x_n) = \frac{m!}{x_1! \cdot x_2! \cdots x_n!} \, p_1^{x_1} \cdot p_2^{x_2} \cdots p_n^{x_n} = m! \prod_i^n \left[\frac{p_i^{x_i}}{x_i!}\right]$$

Theorem 4.6.4 Multinomial Theorem Let $m, n \in \mathbb{Z}^+$, A be the set of vectors $\underline{x} = (x_1, x_2, \dots x_n) \ni$ each x_i is a non negative integer and $\sum_i^n x_i = m$ then for any $p_1, p_2 \dots p_n \in \mathbb{R}$

$$(p_1 + p_2 + \dots + p_n)^m = \sum_{x \in A} \left[\frac{m!}{x_1! \, x_2! \cdots x_n!} \, p_1^{x_1} \cdot p_2^{x_2} \cdots p_n^{x_n}\right]$$

Definition 4.6.5 Let $\underline{X}_1, \underline{X}_2, \dots \underline{X}_n$ be random vectors w/ joint pmf or pdf $f(\underline{x}_1, \underline{x}_2, \dots \underline{x}_n)$. These are considered mutually independent if $\forall (\underline{x}_1, \underline{x}_2, \dots \underline{x}_n)$

$$f(\underline{x}_1, \underline{x}_2 \dots \underline{x}_n) = f_{\underline{X}_1}(\underline{x}_1) \cdot f_{\underline{X}_2}(\underline{x}_2) \cdots f_{\underline{X}_n}(\underline{x}_n) = \prod_i^n f_{\underline{X}_i}(\underline{x}_i)$$

Theorem 4.6.6 Let $X_1, X_2, \dots X_n$ be mutally independent. Let $g_1, g_2 \dots g_n$ be real valued functions $\ni g_i(x_i)$ is a function only of x_i $i = 1, 2, \dots n$ then

$$E(g_1(\underline{X}_1) \cdot g_2(\underline{X}_2) \cdots g_n(\underline{X}_n)) = (E(g_1(\underline{X}_1)))(E(g_2(\underline{X}_2))) \cdots (E(g_n(\underline{X}_n)))$$

Theorem 4.6.7 Let $X_1, X_2, \dots X_n$ be mutually independent w/ MGFs $M_1(t), M_2(t) \dots M_n(t)$. Let $Z = X_1 + X_2 + \dots X_n$ then the MGF of Z is

$$M_Z(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdots M_{X_n}(t)$$
$$= (M_X(t))^n \quad \text{(if } x \text{ are iid)}$$

Corollary 4.6.9 Let $X_1, X_2, \ldots X_n$ be mutually independent w MGFs $M_{x_1}(t), M_{x_2}(t), \ldots M_{x_n}(t)$. Let $a_1, a_2, \ldots a_n$ and $b_1, b_2, \ldots b_n$ be fixed constants Let $Z = (a_1X_1 + b_1) + \cdots + (a_nX_n + b_n)$ then the MGF of Z is

$$M_Z(t) = \left(e^{t(\Sigma b_i)}\right) M_{x_1}(a_1 t) \cdot M_{x_2}(a_2 t) \cdots M_{x_n}(a_n t)$$

Corollary 4.6.10 Let $X_1, X_2, \ldots X_n$ be mutually independent w/ $X_i \sim N(\mu_i, \sigma_i^2)$ then, for fixed constants $a_1, a_2, \ldots a_n$ and $b_1, b_2, \ldots b_n$

$$Z = \sum_i^n (a_i X_i + b_i) \sim N\left(\sum_i^n (a_i \mu_i + b_i), \sum_i^n a_i^2 \sigma_i^2\right)$$

Theorem 4.6.11 Random Vec $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ are mutually independent iff $\exists$ functions $g_i(x_i)$ for $i = 1, 2, \ldots n \ni$
$f(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n) = g(\underline{x}_1) g(\underline{x}_2) \cdots g(\underline{x}_n)$

Theorem 4.6.12 Let $\underline{X}_1, \underline{X}_2, \ldots \underline{X}_n$ be independent. Let $g_i(x_i)$ be a function only of x_i, $i = 1, 2, \ldots n$ then the random variables $U = g_i(x_i)$ are mutually independant!

Multivariate Transformations

- $A = \{\underline{x} \mid f_{\underline{x}}(\underline{x}) > 0\}$
- $U_1 = g_1(X_1, X_2, \ldots X_n),\ U_2 = g_2(X_1, X_2, \ldots X_n), \ldots\ U_n = g_n(X_1, X_2, \ldots X_n)$
- Suppose $\exists$ $A_0, A_1, \ldots A_k$ forms a partition of $A \ni$
  1. A_0 can be empty and $P(\underline{X} \in A_0) = 0$
  2. The transformation $U_1, U_2 \ldots U_n = g_1, g_2, \ldots g_n$ is 1:1 thus the inverses can be found
- $X_1 = h_1(\underline{x}),\ X_2 = h_2(\underline{x}), \ldots\ X_n = h_n(\underline{x})$
- $J = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \cdots & \frac{\partial x_1}{\partial u_n} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \cdots & \frac{\partial x_2}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \cdots & \frac{\partial x_n}{\partial u_n} \end{vmatrix}$
- $f_{\underline{U}}(u_1, u_2, \ldots u_n) = \sum_i f_X(h_{1i}(u), h_{2i}(u), \ldots h_{ni}(u)) |J_i|$

<u>Lemma 4.7.1</u> Let a and b be any positive numbers and let p and q be any positive numbers greater than 1 satisfying $\frac{1}{p}+\frac{1}{q}=1$ then

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab \quad \text{(equality when } a^p = b^q$$

<u>Theorem 4.7.2 Hölder's Inequality</u> Let X and Y be any two RV and let p and q $\ni$ $1/p + 1/q = 1$ then

$$|E(XY)| \leq E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}$$

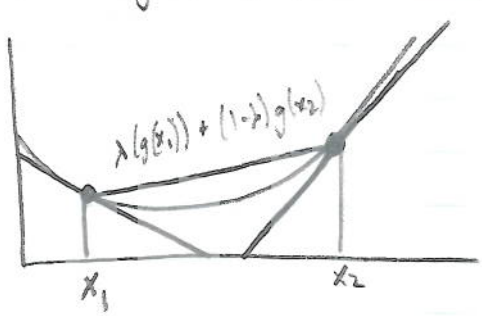
<u>Theorem 4.7.3 Cauchy-Schwarz</u> For any two RV X, Y

$$|E(XY)| \leq E|XY| \leq (E|X|^2)^{1/2}(E|Y|^2)^{1/2}$$

<u>Theorem 4.7.5 Minkowski's Inequality</u> Let X and Y be any two RV then

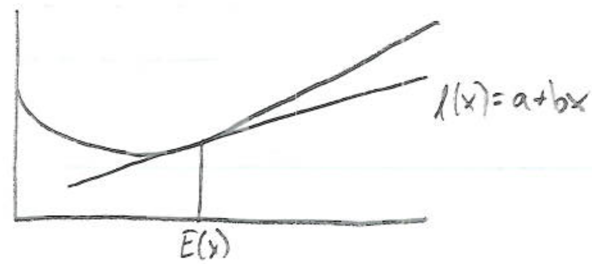
$$[E|X+Y|^p]^{1/p} \leq [E|X|^p]^{1/p} + [E|Y|^p]^{1/p}$$

<u>Definition 4.7.6</u>
- A function $g(x)$ is <u>convex</u> if $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$ $\forall$ x and y and $0<\lambda<1$
- A function $g(x)$ is concave if $-g(x)$ is convex



Theorem 4.7.7 Jensen's For RV X & g(x) is convex then

$$E(g(X)) \geq g(E(X))$$



(Equality holds when $\forall$ line $a+bx$ tangent to $g(x)$ at at $x = E(X)$ $P(g(X) = a + bX) = 1$

Theorem 4.7.9
- Let X be a RV
- Let $g(x)$ and $h(x)$ be any functions $\ni$ $E(g(x))$, $E(h(x))$, $E(g(x)h(x))$ exist

a) If $g(x)$ is a nondecreasing function and $h(x)$ is a non increasing function then

$$E(g(X)h(X)) \leq E(g(X))\,E(h(X))$$

b) If $g(x)$ and $h(x)$ are both non increasing or non decreasing then

$$E(g(x)h(x)) \geq E(g(x))\,E(h(x))$$

Chapter 5

Definition 5.1.1
- The random variables $X_1, X_2, \dots X_n$ are called a random sample of size n from the population $f(x)$ if $X_1, X_2, \dots X_n$ are mutally independent and the marginal pdf or pmf of each x_i is the same function.
- Also denoted iid (independent and identically distributed

Definition 5.2.1 Let $X_1, X_2, \dots X_n$ be a random sample of size n from a population and let $T(x_1, x_2, \dots x_n)$ be a real valued or vector valued function whose domain includes the sample space of $(X_1, X_2, \dots X_n)$ then $Y = T(X_1, X_2, \dots X_n)$ is called a statistic. The probability distribution of y is the sampling distribution.

Definition 5.2.2 The sample mean, $\bar{x} = \frac{1}{n} \sum_1^n x_i$

Definition 5.2.3 The sample variance, $S^2 = \frac{1}{n-1} \sum_1^n (x_i - \bar{x})^2$
The sample standard deviation $\sqrt{S^2}$

Theorem 5.2.4 Let $x_1, x_2, \dots x_n$ be any numbers w/ $\bar{x} = \frac{1}{n} \sum_1^n x_i$ then
a) $\min_a \sum_1^n (x_i - a_i)^2 = \sum_1^n (x_i - \bar{x})^2$
b) $(n-1) S^2 = \sum_1^n (x_i - \bar{x})^2 = \sum_1^n x_i^2 - n\bar{x}^2$

Lemma 5.2.5 $X_1, X_2, \dots X_n$, a random sample from a population and let $g(x)$ be a function $\ni E(g(X_1))$ and $Var(g(X_1))$ exist then

1) $E\left(\sum_1^n g(x_i)\right) = n\left(E(g(x_1))\right)$
2) $var\left(\sum_1^n g(x_i)\right) = n(Var\, g(X_1))$

Theorem 5.2.6 Let $X_1, X_2, \dots X_n$ be a random sample from a pop w/ mean μ and variance $\sigma^2 < \infty$ then
a) $E(\bar{x}) = \mu$
b) $Var(\bar{x}) = \sigma^2/n$
c) $E(s^2) = \sigma^2$

<u>Theorem 5.2.7</u> Let $X_1, X_2, \dots X_n$ be a random sample from a population w/ MGF $M_X(t)$ then the MGF of the sample mean is

$$M_{\bar{X}}(t) = [M_X(t/n)]^n$$

<u>Theorem 5.2.9</u> If X and Y are independent continuous random variables w/ pdfs $f_X(x)$ and $f_Y(y)$ then the pdf of $Z = X+Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\,dw$$

<u>Theorem 5.2.11</u> · Suppose $X_1, X_2, \dots X_n$ is a random sample from a pdf or pmf $f_X(x|\theta) = h(x)c(\theta)e^{\sum w_i(\theta)t_i(x)}$

· Define statistics $T_1, T_2, \dots T_k$ by
$T_i(X_1, X_2, \dots X_n) = \sum_1^n t_i(X_j), \quad i = 1, 2, \dots k$

· If the set $\{(w_1(\theta), w_2(\theta), \dots, w_k(\theta)), \theta \in \Theta\}$ contains an open subset of $\mathbb{R}^k$, then the distribution of $(T_1, T_2, \dots T_k)$ is an exponential family of the form
$f_T(u_1, u_2, \dots, u_k|\theta) = H(u_1, u_2, \dots u_k)[c(\theta)]^n e^{\sum w_i(\theta)u_i}$

<u>Theorem 5.3.1</u> Let $X_1, X_2, \dots X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution and let $\bar{X} = \frac{1}{n}\sum_1^n X_i$ and $S^2 = \frac{1}{n-1}\sum_1^n (X_i - \bar{X})^2$ Then

a) $\bar{X}$ and S^2 are independent
b) $\bar{X} \sim N(\mu, \sigma^2/n)$
c) $(n-1)S^2/\sigma^2 \sim \chi^2_{(n-1)}$

<u>Lemma 5.3.2</u> a) If $Z \sim N(0,1)$, then $Z^2 \sim \chi^2_{(1)}$

b) If $X_1, X_2, \dots X_n$ are independent and $X_i \sim \chi^2_{p_i}$ then
$X_1 + X_2 + \dots + X_n \sim \chi^2_{(p_1+p_2+\dots p_n)}$

Lemma 5.3.3 • let $X_j \sim N(\mu_j, \sigma_j^2)$ $j=1,2,\dots n$ be independent.
for constants a_{ij} and b_{rj} $(j=1,2,\dots n,\ i=1,\dots k,\ r=1,\dots m)$
• where $k+m \le n$
• Define $U_i = \sum_{j=1}^{n} a_{ij} X_j \quad i=1,2,\dots k$
$V_r = \sum_{j=1}^{n} b_{rj} X_j \quad r=1,2,\dots m$

a) The random variables U_i and V_r are independent iff $Cov(U_i, V_r) = \sum_{j=1}^{n} a_{ij} b_{rj} \sigma_j^2 = 0$

b) The random vectors $(U_1, U_2, \dots U_k)$ and $(V_1, V_2, \dots V_m)$ are independent iff $U_i \perp\!\!\!\perp V_r$ $\forall$ pairs i,r $(i=1,2,\dots k;\ r=1,2,\dots m)$

Student T $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$ $\quad f_T(t) = \dfrac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \dfrac{1}{(p\pi)^{1/2}} \dfrac{1}{(1+t^2/p)^{(p+1)/2}} \quad -\infty < t < \infty$

Definition 5.3.4 Let $X_1, X_2, \dots X_n$ be a random sample from a $N(\mu, \sigma^2)$ then $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$

Definition 5.3.6 • For $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu_X, \sigma_X^2)$
$Y_1, Y_2, \dots Y_n \overset{iid}{\sim} N(\mu_Y, \sigma_Y^2)$ $\perp\!\!\!\perp$

• $F = \dfrac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(n-1, m-1)$

• $f_F(x) = \dfrac{\Gamma\frac{p+q}{2}}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\dfrac{p}{q}\right)^{p/2} \dfrac{x^{(p/2)-1}}{\left[1+(p/q)x\right]^{\frac{p+q}{2}}} \quad 0 < x < \infty$

Theorem 5.3.8 a) If $X \sim F_{p,q}$ then $\frac{1}{X} \sim F_{q,p}$
b) If $X \sim t_q$ then $X^2 \sim F_{1,q}$
c) If $X \sim F_{p,q}$ then $\dfrac{\frac{p}{q}X}{\left(1+\left(\frac{p}{q}\right)X\right)} \sim beta(p/2, q/2)$

Definition 5.4.1 The order statistics of $X_1, X_2, \dots X_n$ is the sample placed in ascending order $X_{(1)}, X_{(2)}, \dots X_{(n)}$

Theorem 5.43 · Let $X_1, X_2, \dots X_n$ be a random sample from a discrete distribution w/ pmf $f_x(x_i) = p_i$ where $x_1 < x_2 < \cdots$ are the possible values of X in ascending order

· Define $P_0 = 0$

$P_1 = p_1$

$P_2 = p_1 + p_2$

$\vdots$

· Let $X_{(1)}, X_{(2)}, \dots X_{(n)}$ be the ordered statistics then

Then

$$P(X_{(j)} \le x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^{\,k} (1-P_i)^{n-k}$$

and

$$P(X_{(j)} = x_i) = \sum_{k=j}^{n} \binom{n}{k} \left[P_i^{\,k}(1-P_i)^{n-k} - P_{i-1}^{k}(1-P_{i-1})^{n-k}\right]$$

Theorem 5.4.4 · Let $X_1, X_2, \dots X_n$ be a random sample from a continuous population w/ CDF $F_x(x)$ and pdf $f_x(x)$

Then

$$f_{X(j)}(x) = \frac{n!}{(j-1)!(n-j)!} f_x(x) \left[F_x(x)\right]^{j-1} \left[1 - F_x(x)\right]^{n-j}$$

Theorem 5.4.6. Let $X_{(1)}, X_{(2)}, \dots X_{(n)}$ denote the order statistics of a random sample $X_1, X_2, \dots X_n$ from a continuous population w/ CDF $F_x(x)$ and pdf $f_x(x)$. Then the joint pdf of $X_{(i)}$ & $X_{(j)}$ $1 \le i < j \le n$ is:

$$f_{X_{(i)}, X_{(j)}}(u,v) = \frac{n!}{(i-1)!\,(j-1-i)!\,(n-j)!} f_x(u) f_x(v) \left[F_x(u)\right]^{i-1}$$

$$* \left[F_x(v) - F_x(u)\right]^{j-1-i} \left[1 - F_x(v)\right]^{n-j}$$

for $-\infty < u < v < \infty$

Definition 5.5.1 A sequence of random variables $X_1, X_2, \dots X_n$ converges in probability to a random variable X if, $\forall \epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0 \quad \text{or, equivalently, } \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$$

Theorem 5.5.2 Weak Law of Large Numbers (WLLN)

- Let $X_1, X_2, \dots$ be iid random variables w/ $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$
- Define $\bar{X}_n = \frac{1}{n}\sum_{i}^{n} X_i$ Then $\forall \epsilon > 0$

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$$

"$\bar{X}_n$ converges in probability to μ"

Theorem 5.5.4 Suppose that $X_1, X_2, \dots$ converges in probability to random variable X and that h is a continuous function then $h(X_1), h(X_2) \dots$ converges in probability to $h(X)$.

Definition 5.5.6 A sequence of random variables $X_1, X_2, \dots$ converges almost surely to a random variable X if $\forall \epsilon > 0$

$$P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1$$

Theorem 5.5.9 Strong Law of Large Numbers (SLLN)

- Let $X_1, X_2, \dots$ be iid random variables w/ $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$ and define $\bar{X}_n = \frac{1}{n}\sum_{i}^{n} X_i$ then $\forall \epsilon > 0$

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1$$

"$\bar{X}_n$ converges almost surely to μ"

Definition 5.5.10 A sequence of random variables $X_1, X_2, \dots$ converges in distribution to a random variable X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all points x where $F_X(x)$ is continuous

<u>Theorem 5.5.12</u> If the sequence of random variables $X_1, X_2 \dots$ converges in probability to a random variable X, the sequence also converges in distribution to X

<u>Theorem 5.5.13</u> The sequence of random variables $X_1, X_2 \dots$ converges in probability to a constant μ iff the sequence also converges in distribution to μ
That is

$$P(|X_n - \mu| > \epsilon) \to 0 \quad \forall \epsilon > 0$$

$$\equiv$$

$$P(X_n \le x) \to \begin{cases} 0 & \text{if } x < \mu \\ 1 & \text{if } x > \mu \end{cases}$$

<u>Theorem 5.5.14 (CLT)</u>
- let $X_1, X_2, \dots$ be a sequence of iid random variables w/ $E(X_i) = \mu$ and $0 < Var(X_i) = \sigma^2 < \infty$
- Define $\bar{X}_n = \frac{1}{n} \sum_1^n X_i$
- Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$
- <u>Then</u> $\forall x \in \mathbb{R}$

$$\lim_{n \to \infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dy$$

that is, $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ has a limiting standard Normal dist

$$\equiv \frac{(\bar{X}_n - \mu)}{\sigma / \sqrt{n}}$$

<u>Theorem 5.5.17</u> If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability then

$$\left.\begin{aligned} &a)\ Y_n X_n \to aX \\ &b)\ X_n + Y_n \to X + a \end{aligned}\right\} \text{in distribution}$$

<u>Definition 5.5.20</u> If a function $g(x)$ has derivatives of order r, that is $g^{(r)}(x) = \frac{\partial^r}{\partial x^r} g(x)$ exists. Then for any constant a, the Taylor polynomial of order r about a is

$$T_r(x) = \sum_{0}^{r} \frac{g^{(i)}(a)}{i!} (x-a)^i$$

<u>Theorem 5.5.21</u> If $g^{(r)}(a) = \frac{d^r}{dx^r} g(x)\big|_{x=a}$ exists then

$$\lim_{x \to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0$$

$$\equiv \quad g(x) - T_r(x) = \int_a^x \frac{g^{(r+1)}(t)}{r!} (x-t)^r \, dt$$

<u>Theorem 5.5.24 Delta Method</u>

- Let Y_n be a sequence of random variables that satisfies
  $\sqrt{n}(Y_n - \theta) \to n(0, \sigma^2)$ in dist
- For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0
- Then
  $\sqrt{n}[g(Y_n) - g(\theta)] \to N(0, \sigma^2[g'(\theta)]^2$ in dist

<u>Theorem 5.5.26 2nd Order Delta Method</u>

- Let Y_n be a sequence of RV that satisfies $\sqrt{n}(Y_n - \theta) \to N(0, \sigma^2)$ in dist
- For a given function g and a specific value of θ, suppose that $g'(\theta) = 0$ exists and $g''(\theta)$ exists and is not 0 then

$$n[g(Y_n) - g(\theta)] \to \sigma^2 \frac{g''(\theta)}{2} \chi_{(1)}^2 \text{ in dist}$$

Theorem 5.5.28 Let $\underline{X}_1, \underline{X}_2, \ldots \underline{X}_n$ be a random sample w/ $E(X_{ij}) = \mu_i$ and $Cov(X_{ik}, X_{jk}) = \sigma_{ij}$. For a given function g with continuous 1st partial derivatives and a specific value of $\mu = (\mu_1, \mu_2, \ldots, \mu_p)$ for which

$$\tau^2 = \sum\sum \sigma_{ij} \frac{\partial g(\mu)}{\partial \mu_i} \frac{\partial g(\mu)}{\partial \mu_j} > 0$$

Then

$$\sqrt{n}\left[g(\bar{X}_1, \bar{X}_2, \ldots \bar{X}_s) - g(\mu_1, \mu_2, \ldots \mu_p)\right] \to N(0, \tau^2) \text{ in dist}$$

Chapter 6

Definition 6.2.1 A statistic $T(\underline{X})$ is a sufficient statistic for θ if the conditional distribution of the sample $\underline{X}$ given the value of $T(\underline{X})$ does not depend on θ

Theorem 6.2.2. If $p(\underline{x}|\theta)$ is the joint pdf or pmf of $\underline{X}$ and $q(t|\theta)$ is the pdf or pmf of $T(\underline{x})$ then $T(\underline{X})$ is a sufficient for θ if $\forall$ x in the sample space, the ratio $\frac{p(\underline{x}|\theta)}{q(T(\underline{x})|\theta)}$ is constant as a function of θ.

Theorem 6.2.6 Factorization Theorem - Let $f(x|\theta)$ denote the joint pdf or pmf of a sample $\underline{X}$. A statistic $T(\underline{X})$ is a sufficient statistic for θ iff there exist functions $g(t|\theta)$ and $h(x)$ $\ni$ $\forall$ sample points x and all parameter points θ

$$f(\underline{x}|\theta) = g(T(\underline{x})|\theta)\,h(\underline{x})$$

Theorem 6.2.10 Let $X_1, X_2, \dots X_n$ be iid observations from a pdf or pmf $f(x|\theta)$ that belongs to an exponential family

$$f(x|\theta) = h(x)\,c(\theta)\,e^{\sum_{i}^{k} w_i(\theta) t_i(x)} \quad \text{where } \theta = (\theta_1, \theta_2, \dots \theta_d),\ d \le k$$

Then

$$T(\underline{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \sum_{j=1}^{n} t_2(X_j), \dots \sum_{j=1}^{n} t_k(X_j)\right) \text{ is sufficient for } \theta$$

Definition 6.2.11 A sufficient statistic $T(\underline{X})$ is called a minimal sufficient statistic if for any other sufficient statistic $T'(\underline{x})$, $T(x)$ is a function of $T'(x)$

Theorem 6.2.13 Let $f(\underline{x}|\theta)$ be the pmf or pdf of a sample $\underline{X}$. Suppose there exists a function $T(\underline{x})$ $\ni$ $\forall$ two sample points $\underline{x}$ and $\underline{y}$, the ratio $f(\underline{x}|\theta)/f(\underline{y}|\theta)$ is constant as a function of θ iff $T(\underline{x}) = T(\underline{y})$. Then $T(\underline{X})$ is a minimal sufficient for θ

Definition 6.2.16 A statistic $S(\underset{\sim}{X})$ who distribution does not depend on the parameter θ is called an ancillary statistic

Definition 6.2.21 Let $f(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(\underset{\sim}{X})$. The family of probability distributions is called complete if $E_\theta(g(T)) = 0 \ \forall \ \theta$ implies $P_\theta(g(T)=0) = 1 \ \forall \ \theta$. Equivalently, $T(\underset{\sim}{X})$ is called a complete statistic

Theorem 6.2.24 (Basu's Theorem) If $T(\underset{\sim}{X})$ is a complete and minimal sufficient statistic, then $T(\underset{\sim}{X})$ is independent of every acillary statistic

Theorem 6.2.25 The statistic defined in 6.2.10 is also complete if $\{(w_1(\underset{\sim}{\theta}), w_2(\underset{\sim}{\theta}), \dots w_k(\underset{\sim}{\theta})) \mid \underset{\sim}{\theta} \in \Theta\}$ contains an open set in $\mathbb{R}^k$

Theorem 6.2.28 If a minimal sufficient statistic exists, then any complete statistic is also a minimal statistics

Skipped 6.3.1 & 6.4

Chapter 7

Definition 7.1.1 A point estimator is any function $W(X_1, X_2, \dots X_n)$ of a sample, that is, any statistic is a point estimator

Method of Moments Define:

$$m_1 = \frac{1}{n}\sum_1^n x_i^1 \qquad \mu_1' = E(X)$$
$$m_2 = \frac{1}{n}\sum_1^n x_i^2 \qquad \mu_2' = E(X^2)$$
$$\vdots \qquad\qquad \vdots$$
$$m_k = \frac{1}{n}\sum_1^n x_i^k \qquad \mu_k' = E(X^k)$$

Solve the system

$$m_1 = \mu_1'$$
$$m_2 = \mu_2'$$
$$\vdots$$
$$m_k = \mu_k'$$

Maximum Likelihood Estimators $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta_1, \theta_2, \dots \theta_k)$

$$L(\theta|\underline{x}) = \prod_1^n f(x_i|\underline{\theta})$$

Def 7.2.4 MLE $= \hat{\theta} =$ parameter which $L(\theta|\underline{x})$ attains it maximum found using the calculus

Theorem 7.2.10 Invariance property of MLEs if $\hat{\theta}$ is the MLE of θ, then for any function $\tau(\theta)$ the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$

Bayesian Estimators
- $X_1, X_2, \dots X_n$ is drawn from a population indexed by θ
- Prior distribution, $\pi(\theta)$, is the distribution of θ based on 'prior' beliefs or intuition
- Posterior, $\pi(\theta|\underline{x})$, is the updated prior when we see $X_1, X_2, \dots X_n$

$$\pi(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\,\pi(\theta)}{\int f(x|\theta)\,\pi(\theta)\,d\theta}$$

Definition 7.2.15 Let $\mathcal{F}$ denote the class of pdfs or pmfs $f(x|\theta)$ indexed by θ. A class Π of prior distributions is a conjugate family for $\mathcal{F}$ if the posterior distribution is in the class Π $\forall f \in \mathcal{F}$, all priors in Π and all $x \in \mathcal{X}$

ie: If we start with a beta prior we will end up w/ a beta posterior

7.2.4 omitted

Definition 7.3.1 The mean square error (MSE) of an estimator W of a parameter θ is $E_\theta(W-\theta)^2$

Definition 7.3.2 The bias of a point estimator W of a parameter θ is the difference between the expected value of W and θ

$$\text{Bias} = E(W) - \theta$$

Note: if Bias = 0 → the estimator W is unbiased

$$\rightarrow \text{MSE} = E_\theta(W-\theta)^2 = Var_\theta(W)$$

Definition 7.3.7 · An estimator W^* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $E_\theta(W^*) = \tau(\theta)$ $\forall \theta$ and, for any other estimator W' w/ $E_\theta(W')$ we have $Var_\theta(W^*) \leq Var_\theta(W')$

· Also called the uniform minimum variance unbiased estimator of $\tau(\theta)$ (UMVUE)

Theorem 7.3.9 Let $X_1, X_2, \ldots X_n$ be a sample w/ pdf $f(\underline{x}|\theta)$ and let $W(\underline{x}) = W(X_1, X_2, \ldots X_n)$ be any estimator satisfying

$$\frac{d}{d\theta} E_\theta(W(\underline{X})) = \int_{\underline{x}} \frac{\partial}{\partial\theta}\left[W(\underline{x}) f(\underline{x}|\theta)\right] dx \quad \text{and } var(W(\underline{x})) < \infty$$

Then

$$Var_\theta(W(\underline{X})) \geq \underbrace{\frac{\left(\frac{d}{d\theta} E_\theta(W(\underline{X}))\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta} \ln f(\underline{X}|\theta)\right)^2\right)}}_{\text{Cramér Rao Lower Bound}}$$

Corollary 7.3.10 When $X_1, X_2, \ldots X_n$ are iid theorem 7.3.9 becomes

$$\mathrm{Var}_\theta(W(\underline{X})) \geq \frac{\left(\frac{d}{d\theta} E_\theta W(\underline{X})\right)^2}{n E_\theta\left(\left(\frac{\partial}{\partial\theta} \ln f(x|\theta)\right)^2\right)}$$

Lemma 7.3.11 If $f(x|\theta)$ satisfies $\frac{d}{d\theta} E_\theta\left(\frac{\partial}{\partial\theta} \ln f(x|\theta)\right) = \int \frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta} \ln f(x|\theta)\, f(x|\theta)\right] dx$
(Always true for exponential Families

then

$$E_\theta\left(\left(\frac{\partial}{\partial\theta} \log f(x|\theta)\right)^2\right) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2} \ln(f(x|\theta))\right)$$

* This result can, very usefully, be applied to Theorem 7.3.9 or Corollary 7.3.10

Corollary 7.3.15 Let $X_1, X_2, \ldots, X_n$ be iid $f(x|\theta)$ where $f(x|\theta)$ satisfies the conditions of theorem 7.3.9. If W is an unbiased estimator of $\tau(\theta)$ then $W(\underline{X})$ attains the CRLB iff

$$\frac{\partial}{\partial\theta} \ln(L(\theta|\underline{x})) = a(\theta)\,[W(\underline{x}) - \tau(\theta)] \quad \text{for some function } a(\theta)$$

Theorem 7.3.17 Rao-Blackwell Let W be any unbiased estimator of $\tau(\theta)$ and let T be a sufficient statistic for θ.
- Define $\phi(T) = E(W|T)$
- Then $E_\theta(\phi(T)) = \tau(\theta)$ and $\mathrm{Var}_\theta(\phi(T)) \leq \mathrm{var}_\theta(W)\ \forall \theta$

$\rightarrow \phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$

Theorem 7.3.19 If W is a best unbiased estimator of $\tau(\theta)$ then W is unique

Theorem 7.3.20 If $E_\theta(W) = \tau(\theta)$, W is the best unbiased estimator of $\tau(\theta)$ iff W is uncorrelated with all unbiased estimators of 0

Theorem 7.3.23 Let T be a complete sufficient statistic for a parameter θ and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value

7.3.4 omitted

Chapter 8

Definition 8.1.1 A hypothesis is a statement about a population parameter

Definition 8.1.2 The two complementary hypotheses in a hypothesis test are called the null hypothesis and the alternative hypothesis

Definition 8.1.3 A hypothesis testing procedure or hypothesis test is a rule that specifies

i: For which sample values the decision is made to accept H_0 as true

The subset of the sample space for which H_0 will be rejected is called the rejection region. The complement is called the acceptance region

Definition 8.2.1 The likelihood ratio test statistic for testing $H_0: \theta \in \Theta_0$, $H_1: \theta \in \Theta_0^c$ is

$$\lambda(\underline{x}) = \frac{\sup_{\Theta_0} L(\theta|\underline{x})}{\sup_{\Theta} L(\theta|\underline{x})} = \frac{L(\hat{\theta}_0|\underline{x})}{L(\hat{\theta}|\underline{x})}$$

$\ni$ $\hat{\theta}_0$ is the Restricted MLE
& $\hat{\theta}$ is the unrestricted MLE

A likelihood ratio test gives the rejection region $R = \{\underline{x} \mid \lambda(\underline{x}) \le c\}$ where $0 \le c \le 1$

Theorem 8.2.4 If $T(\underline{X})$ is a sufficient statistic for θ and $\lambda^*(t)$ and $\lambda(\underline{x})$ are the LRT statistics based on T and X respectively then $\lambda^*(T(\underline{x})) = \lambda(\underline{x})$ $\forall$ x in the sample space

Bayesian Tests 1) $P(\theta \in \Theta_0 | \underline{x}) = P(H_0 \text{ is true})$
2) $P(\theta \in \Theta_0^c | \underline{x}) = P(H_1 \text{ is true})$

8.2.3 omitted

Table 8.3.1

		Decision	
		Fail to reject H_0	Reject H_0
Truth	H_0	Correct	Type I error
	H_1	Type II error	Correct

Definition 8.3.1 Power function of a hypothesis test w/ rejection region R is the function of θ defined by $\beta(\theta) = P_\theta(X \in R)$

* we want $\beta(\theta)$ close to zero $\forall\ \theta \in \Theta_0$
close to one $\forall\ \theta \in \Theta_0^c$

Definition 8.3.5 For $0 \le \alpha \le 1$, a test w/ power function $\beta(\theta)$ is a size α test if $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$

Definition 8.3.6 For $0 \le \alpha \le 1$, a test w/ power function $\beta(\theta)$ is a level α test if $\sup_{\theta \in \Theta_0} \le \alpha$

Definition 8.3.9 A test w/ power function $\beta(\theta)$ is unbiased if $\beta(\theta') \ge \beta(\theta'')$ for every $\theta' \in \Theta_0^c$ and $\theta'' \in \Theta_0$

Definition 8.3.11 Let $\mathcal{C}$ be a class of tests for testing $H_0: \theta \in \Theta_0$ vs $H_1: \theta \in \Theta_0^c$. A test in class $\mathcal{C}$, w/ power function $\beta(\theta)$ is a uniformly most powerful test if $\beta(\theta) \ge \beta'(\theta)$ $\forall\ \theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class $\mathcal{C}$

Theorem 8.3.12 Neyman-Pearson Lemma Consider testing $H_0: \theta = \theta_0$ versus $H_1: \theta = \theta_1$ where the pdf or pmf corresponding to θ_i is $f(x|\theta_i)$ $i = 0, 1$ using a test w/ rejection region R that satisfies

$$x \in R \text{ if } f(x|\theta_1) > k f(x|\theta_0)$$

① and

$$x \in R^c \text{ if } f(x|\theta_1) < k f(x|\theta_0)$$

② for some $k \geq 0$ and

$$\alpha = P_{\theta_0}(\underline{X} \in R)$$

Then

a.) Any test that satisfies ① and ② is a UMP level α test

b.) If $\exists$ a test satisfying ① and ② w/ $k > 0$, then every UMP level α test is a size α test and every UMP level α test satisfies ① except perhaps on a set A satisfying $P_{\theta_0}(\underline{X} \in A) = P_{\theta_1}(\underline{X} \in A) = 0$

Corollary 8.3.13
- Consider testing $H_0: \theta = \theta_0$ versus $H_1: \theta = \theta_1$
- Suppose $T(\underline{X})$ is a sufficient statistic for θ and $g(t|\theta_i)$ is the pdf or pmf of T corresponding to θ_i $i = 0, 1$.
- Then any test based on T w/ rejection region S is a UMP level α test if it satisfies

$$t \in S \text{ if } g(t|\theta_1) > k g(t|\theta_0)$$

and

$$t \in S^c \text{ if } g(t|\theta_1) < k g(t|\theta_0)$$

for some $k \geq 0$ where

$$\alpha = P_{\theta_0}(T \in S)$$

Definition 8.3.16 A family of pdfs or pmfs $\{g(t|\theta) \mid \theta \in \Theta\}$ for a univariate random variable T w/ real-valued parameter θ has a monotone likelihood ratio (MLR) if for every $\theta_2 > \theta_1$ $g(t|\theta_2)/g(t|\theta_1)$ is a monotone (nonincreasing or nondecreasing) function of t on $\{t \mid g(t|\theta_1) > 0 \text{ or } g(t|\theta_2) > 0\}$

Theorem 8.3.17 Karlin Rubin Consider testing $H_0: \theta \le \theta_0$
$H_1: \theta > \theta_0$

Suppose that T is a sufficient statistic for θ and the family of pdfs or pmfs $\{g(t|\theta) | \theta \in \Theta\}$ of T has a MLR. Then for any t_0, the test that rejects H_0 iff $T > t_0$ is a UMP level α test where $\alpha = P_{\theta_0}(T > t_0)$

Chapter 9

Definition 9.1.1 An *interval Estimate* of θ is any *interval estimator* $[L(\underline{X}), U(\underline{X})]$ making the inference $L(\underline{X}) < \theta < U(\underline{X})$

Definition 9.1.4 *Coverage Probability* $= P_\theta(\theta \in [L(\underline{X}), U(\underline{X})])$

Definition 9.1.5 *Confident Coefficient* $= \inf_\theta P_\theta(\theta \in [L(\underline{X}), U(\underline{X})])$

Theorem 9.2.2 i) $\forall\ \theta_0 \in \Theta$, let $A(\theta_0)$ = the acceptance region of a level α test of $H_0: \theta = \theta_0$
$\forall\ \underline{x} \in \underline{X}$ define $C(\underline{x}) = \{\theta_0 \mid x \in A(\theta_0)\}$
then $C(\underline{X})$ is a $1-\alpha$ confidence set

ii) If $C(\underline{X})$ is a $1-\alpha$ confidence set
$\forall\ \theta_0 \in \Theta$ $A(\theta_0) = \{x \mid \theta_0 \in C(\underline{x})\}$ is the acceptance region of a level α test of $H_0: \theta = \theta_0$

proof 422

Definition 9.2.6 A random Variable $Q(\underline{X}, \theta)$ is a *pivotal quantity* or *pivot* if $Q(\underline{X}, \theta)$ is independent of all parameters

PIVOT RULE If $T \sim f(t|\theta)$ can be expressed as

$$f(t|\theta) = g(Q(t,\theta))\left|\frac{\partial}{\partial t} Q(t,\theta)\right|$$

then $Q(t,\theta)$ is a pivot

thus $C(\underline{x}) = \{\theta_0 \mid a \leq Q(\underline{x}, \theta_0) \leq b\}$

① If $Q(t,\theta)$ is an inc. function of θ
$L(\underline{x}, a) \leq \theta \leq U(\underline{x}, b)$

② If $Q(t,\theta)$ is a dec. function of θ
$L(\underline{x}, b) \leq \theta \leq U(\underline{x}, a)$

<u>Theorem 9.2.12</u> · Let T be a statistic w/ continuous $F_T(t|\theta)$

· Let $\alpha_1 + \alpha_2 = \alpha$

· $\forall\, t \in \mathcal{T}$

i) If $F_T(t|\theta)$ is a dec function of θ

$F_T(t|\theta_u(t)) = \alpha_1$, $F_T(t|\theta_L(t)) = 1-\alpha_2$

ii) If $F_T(t|\theta)$ is an inc. function of θ

$F_T(t|\theta_u(t)) = 1-\alpha_2$, $F_T(t|\theta_L(t)) = \alpha_1$

Then $(\theta_L(t), \theta_U(t))$ is a $1-\alpha$ CI

<u>Proof</u> 432

<u>Theorem 9.2.14</u> · Let T be a discrete statistic w/ CDF $F_T(t|\theta)$

· Let $\alpha_1 + \alpha_2 = \alpha$

· $\forall\, t \in \mathcal{T}$

i) If $F_T(t|\theta)$ is a dec function of θ

$P(T \le t|\theta_u(t)) = \alpha_1$, $P(T \ge t|\theta_L(t)) = \alpha_2$

ii) If $F_T(t|\theta)$ is an inc. function of θ

$P(T \ge t|\theta_u(t)) = \alpha_1$, $P(T \le t|\theta_L(t)) = \alpha_2$

Then $(\theta_L(t), \theta_u(t))$ is a $1-\alpha$ CI

<u>Proof</u> 434

<u>Credible Set</u> With $\pi(\theta|\underline{x})$, the posterior of θ given $\underline{x}$ then for any set $A \subset \Theta$ the Credible Prb. of A is

$\int_A \pi(\theta|\underline{x})\,d\theta = P(\theta \in A|x)$

<u>Theorem 9.3.2</u> Let $f(x)$ be a unimodal pdf. If $[a,b]$ satisfies

i) $\int_a^b f(x)\,dx = 1-\alpha$

ii) $f(a) = f(b) > 0$ and

iii) $a \le x^* \le b$ where x^* is a mode of $f(x)$

Then $[a,b]$ is the shortest interval among all that satisfy i

proof: p442