# Contents

# 1 Sets

## 1.1 Set Identities

Union: $A \cup B : \{x\epsilon\mathbb{S} : x\epsilon A \text{ OR } x\epsilon B\}$ Intersection: $A \cap B : \{x\epsilon\mathbb{S} : x\epsilon A \text{ AND } x\epsilon B\}$ Complement: $A^c : \{x\epsilon\mathbb{S} : x\not\epsilon A\}$ Difference: $A - B : \{x\epsilon\mathbb{S} : x\epsilon A,\ x\not\epsilon B\}$ Infinite Union: $\bigcup_{i=1}^{\infty} A_i : \{x\epsilon\mathbb{S}, x\epsilon A_i \ni A_i\}$ Infinite Intersection: $\bigcap_{i=1}^{\infty} A_i : \{x\epsilon\mathbb{S}, x\epsilon A_i \forall A_i\}$

## 1.2 Set Relationships

Containment: $A \subseteq B$ (A is a subset of B): $x\epsilon A$ means $x\epsilon B$ Equality: Two sets are equal if they contain each other: $A = B \therefore A \subseteq B, B \subseteq A$ Disjoint: $A \cap B = \{\}$

## 1.3 Set Properties

Commutativity: $A \cup B = B \cup A$, $A \cap B = B \cap A$ Associativity: $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$ Distributive: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ DeMorgan's Law: $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$

# 2 Sigma Algebras

## 2.1 Identity

A collection of subsets of $S$ is a $\sigma$-algebra ($\mathbb{B}$) iff: a. $\emptyset \epsilon \mathbb{B}$ b. $A \epsilon \mathbb{B} \implies A^c \epsilon \mathbb{B}$ c. $A_1, A_2, ... \epsilon \mathbb{B} \implies \bigcup_{n=1}^{\infty} A_n \epsilon \mathbb{B}$

## 2.2 Construction

$S$ is finite/countable: $\mathbb{B} = \mathbb{P}(\mathbb{S})$ (Power Set of $\mathbb{S}$, all possible subsets of $\mathbb{S}$)

$S$ is infinite/uncountable: Use Borel sets: $\mathbb{B} = \{(a, b), [a, b), [a, b]\}$ for $a < b$ and all countable $\cup$ and $\cap$ of those

# 3 Probability Functions

## 3.1 Axioms

Given $\mathbb{S}$ and $\sigma$-algebra, a probability function with domain $\mathbb{B}$ satisfies: a. $P(A) \geq 0$ for all $A \epsilon \mathbb{B}$ b. $P(\mathbb{S}) = 1$ c. If $A_1, A_2, ...$ are pairwise disjoint, then $P(\bigcup\limits_{n=1}^{\infty} A_n) = \sum\limits_{n=1}^{\infty} P(A_n)$

## 3.2 Properties

1) $P(\emptyset) = 0$ 2) $A \subseteq \mathbb{S} \implies P(A) \leq 1$ 3) $P(A^c) = 1 - P(A)$ 4) $P(B \cap A^c) = P(B) - P(A \cap B)$ 5) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ 6) $A \subseteq B \implies P(A) \leq P(B)$ 7) Let $c_1, c_2, ...$ be a partition of $\mathbb{S}$ (ie. $c_i \cap c_j = \emptyset$ for $i \neq j$, $\bigcup\limits_{i=1}^{\infty} c_i = \mathbb{S}$) - $P(A) = \sum\limits_{i=1}^{\infty} P(A \cap c_i)$ 8) For any $A_1, A_2, ...$; $P(\bigcup\limits_{i=1}^{\infty} A_i) \leq \sum\limits_{i=1}^{\infty} P(A_i)$

# 4 Counting

## 4.1 Sampling

| | **w/o Repl.** | **w/ Repl.** |
|---|---|---|
| **Ordered Perm.** | $\boxed{\frac{n!}{(n-1)!}}$ | $\boxed{n^r}$ |
| **Unordered Comb.** | $\boxed{\frac{n!}{(n-r)!r!} : \binom{n}{r}}$ | $\boxed{\binom{n+r-1}{r}}$ |

## 4.2 Axioms

Enumerating equally likely outcomes (assume large but finite $\mathbb{S}, |\mathbb{S}| = N$). Want $P(A)$ where $A \subset \mathbb{S}, A \epsilon \mathbb{B}$ - $P(A) = \frac{\text{\# things in A}}{N}$

Product Rule: - If a job consists of $k$ separate experiments, the $i^{th}$ of which can be done in $n_i$ ways, then the job can be done in $n_1 * n_2 * ... * n_k$ ways

Sum Rule: - If there are $k$ events, the $i^{th}$ of which can occur in $n_i$ ways, then there are $n_1 + n_2 + ... + n_k$ to complete exactly 1 event Inclusion/Exclusion: want to enumerate elements in $A : N_A = |A|$, sometimes easier to find: - $N_{A^c} = |A^c| \therefore N_A = N - N_{A^c}$

## 4.3 Continous

Consider $\mathbb{S} \subset \mathbb{R}^d$ with uniform probability Then for $A \subseteq \mathbb{S}, P(A) = \frac{\int\limits_A ds}{\int\limits_{\mathbb{S}} ds}$

## 4.4 Conditional Probability

If $A, B \subseteq \mathbb{S}$ and $P(B) > 0$; $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$ Often use the law of total probability ($c_i \cap c_j = \emptyset$ for $i \neq j, \bigcup\limits_{i=1}^{\infty} c_i = \mathbb{S}$): $P(B) = \sum\limits_{i=1}^{n} P(B|c_i)P(c_i)$

## 4.5 Independence

$A \perp\!\!\!\perp B$ iff $P(A|B) = P(A)$ $A \perp\!\!\!\perp B \implies A \perp\!\!\!\perp B^c, A^c \perp\!\!\!\perp B, A^c \perp\!\!\!\perp B^c$ Mutual Indepedence: A collection of events $A_1, ..., A_n$ are mut. ind. if, for any subcollection of $A_{i_1}, ..., A_{i_k}$ we have: - $P(\bigcap\limits_{j=1}^{k} A_{i_j}) = \prod\limits_{j=1}^{k} P(A_{i_j})$

## 4.6 Conditional Independence

$A$ and $B$ are conditionally independent given $C$ if: $P([A \cap B]|C) = P(A|C)P(B|C)$

# 5 Random Variables

## 5.1 Definition

A random variable (vector) is a function that maps from the sample space $\mathbb{S}$ to the real numbers $\mathbb{R}$ Formally: $X : \mathbb{S} \Rightarrow \mathbb{R}, \underset{\sim}{\boldsymbol{X}} : \mathbb{S} \Rightarrow \mathbb{R}$

## 5.2 Cumulative Distribution Function

The CDF of a random variable ($F_X(x)$) is defined as: $P(X \leq x)$ for all $x \epsilon \mathbb{R}$ a. $\lim\limits_{x \to -\infty} F_X(x) = 0, \lim\limits_{x \to \infty} F_X(x) = 1$ b. $F_X(x)$ is non-decreasing ie. for $x_i \leq x_2, F(x_1) \leq F(x_2)$ c. $F_X(x)$ is right-conitnuous ie. $\lim\limits_{x \downarrow x_0} F_X(x) = F_X(x_0)$

## 5.3 Probability Density/Mass Function

A PMF is given by $f_X(x) = P(X = x)$ A PDFof a continuous random variable satisfies the following: - $\int\limits_{-\infty}^{x} f_x(t)dt$ for all $x \therefore f(X) = \frac{dF_x}{dx}$ - $P(a \leq x \leq b) = \int\limits_{a}^{b} f_X(x)dx = P(a < x < b) = F(b) - F(a)$ A function is a valid PMF/PDF iff: a) $f_X(x) \geq 0, \forall x$ b) $\sum\limits_{x \epsilon X} f_X(x) = 1$ -OR- $\int\limits_{x} f_X(x)dx = 1$

## 5.4 Kernel

Any non-negative function with a finite integral or sum can be made into a PDF or PMF - $h(x) \geq 0 \forall x$ - $\int\limits_{x \epsilon X} h(x)dx = k, 0 < k < \infty$ - $f_X(x) = \frac{1}{k} h(x) I_X(x)$

## 5.5 Common PDFs

Besides those given in the book: - Survival Function: $S_X(x) = P(X > x) = 1 - F_X(x)$ - Hazard Function: $H_X(x) = \frac{f_X(x)}{S_X(x)}$ - Gamma Function: $\Gamma(\alpha) = \int\limits_{0}^{\infty} t^{\alpha-1} e^{-t} dt$ - - If $\alpha$ is an integer: $\Gamma(\alpha) = (\alpha - 1)!$ - - For general $\alpha$: $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ - - Also: $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

# 6 Expected Value

## 6.1 Definition

Given a random variable $g(x)$:

$$\mathbb{E}[g(x)] = \begin{cases} \int\limits_{-\infty}^{\infty} g(x) f_X(x) dx, & \text{Continuous} \\ \sum\limits_{x \epsilon X} g(x) f_X(x), & \text{Discrete} \end{cases}$$

Law of Unconscious Statistician: Let $Y = g(x)$ - $\mathbb{E}[g(x)] = \int\limits_{-\infty}^{\infty} g(x) f_X(x) dx = \int\limits_{-\infty}^{\infty} Y f_Y(y) dy = \mathbb{E}[Y]$

Probability as an Expectation: $P(x \epsilon A) = \int\limits_{A} f_X(x) dx = \int\limits_{-\infty}^{\infty} I_A(x) f_X(x) dx = \mathbb{E}[I_A(x)]$

## 6.2 Properties of Expected Values

1) $\mathbb{E}[ax + b] = a\,\mathbb{E}[x] + b, \mathbb{E}[ag_1(x) + bg_2(x)] = a\,\mathbb{E}[g_1(x)] + b\,\mathbb{E}[g_2(x)]$ 2) If $g(x) \geq 0, \forall x \epsilon X$, then $\mathbb{E}[g(x)] \geq 0$ 3) If $g_1(x) \geq g_2(x), \forall x \epsilon X$, then $\mathbb{E}[g_1(x)] \geq \mathbb{E}[g_2(x)]$ 4) If $a \leq g(x) \leq, \forall x \epsilon X$, then $a \leq \mathbb{E}[g(x)] \leq b$

# 7 Moments

## 7.1 Definition

For each interger $n$, the $n^{th}$ moment of $X$ is $\mathbb{E}[X^n]$ The $n^{th}$ central moment is: $\mathbb{E}[X - \mathbb{E}[X]]^n$

Expected value is the first moment, Variance is the second central moment

Properties of Variance: - $Var(aX + b) = a^2 Var(X)$ - $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

## 7.2 Jensen's Inequality

Want to compare $\mathbb{E}[X]$ vs. $\mathbb{E}[Y]$ where $Y = g(X)$. Often can't directly compare.

$$JE : \begin{cases} \mathbb{E}[g(X)] \geq g(\mathbb{E}[X]), & g(x) \text{is convex} \\ \mathbb{E}[g(X)] \leq g(\mathbb{E}[X]), & g(x) \text{is concave} \end{cases}$$

How to tell if $g(x)$ is convex: - Draw it (convex is bowl-shaped) - Second Derivative: $g''(x) > 0 \implies$ convex

## 7.3 Moment Generating Function (MGF)

$$M_X(t) = \mathbb{E}[e^{tx}] = \begin{cases} \int\limits_{-\infty}^{\infty} e^{tx} f_X(x)dx, & X \text{is continuous} \\ \sum\limits_{x \epsilon X} e^{tx} f_X(x), & X \text{is discrete} \end{cases}$$

This holds if the expectation exists for $t$ in the neighborhood of 0. That is, there exists an $h > 0$ such that $\mathbb{E}(e^{tx})$ exists for all $-h < t < h$

$\mathbb{E}[X^n] = M_X^{(n)}(0) = \frac{dn}{dt^n} M_x(t)|_{t=0}$

## 7.4 Characterizing Distributions

a) If X and Y have bounded support, $F_X(u) = F_Y(u)$ for all $u$ iff $\mathbb{E}[X^r] = \mathbb{E}[Y^r]$, $r = 0, 1, 2, 3, ...$ (all moments are equal) b) If MGF exists $M_X(t) = M_Y(t)$ for some $t$ in neighborhood of 0, then $F_X(u) = F_Y(u)$ for all $u$

A sequence of random variables, $X_i, i = 1, 2, 3, ...$ each with an MGF $M_{X_i}(t)$. Further suppose $\lim\limits_{i->\infty} M_{X_i}(t) = M_x(t)$ for t in neighborhood of 0, and $M_x(t)$ is also an MGF Then: there is a unique CDF $F_X(x)$ whose moments are determined by $M_x(t)$ and $\lim\limits_{(} i- > \infty) F_{X_i}(x) = F_x(x)$ Basically, if the MGFs of RVs converge to an MGF, then the RVs themselves converge to the RV of the converged MGF

Lemma: Let $a_1, a_2, a_3 ...$ be a sequence of numbers such that $\lim\limits_{n \to \infty} a_n = a$ Then: $\lim\limits_{n \to \infty} (1 + \frac{a_n}{n})^n = e^a$

Theorem: Let $Y = aX = b \therefore M_Y(t) = e^{at} M_X(t)$

# 8 Transformations

## 8.1 Definition

$X$ is a random variable, then $Y = g(X)$ is also a random variable. To find $P(Y)$ we need either $F_Y(y)$ or $f_Y(y)$ - $g(X)$ maps from $\mathbb{X}$ to $\mathbb{Y}$, basically $\mathbb{S} \to \mathbb{X} \to \mathbb{Y}$ - $\forall A, P(Y \epsilon A) = P(g(X) \epsilon A) = P(\{x \epsilon \mathbb{X} : g(x) = A\}) = P(X \epsilon g^{-1}(A))$

## 8.2 Discrete

$$f_Y(y) = \begin{cases} \sum\limits_{X \epsilon g^{-1}(y)} P(X = x), & Y \epsilon \mathbb{Y} \\ 0, & \text{otherwise} \end{cases}$$

Steps: 1) Find $\mathbb{Y}$ 2) Identify $g^{-1}(y)$ 3) Sum over appropriate $x$ (if $g^{-1}(y)$ is a set with one element, $f_Y(y) = f_X(g^{-1}(y)))$

## Continous

$$F_Y(y) = P(Y = y) = P(g(x) \leq y) = \int\limits_{x \epsilon \mathbb{X} : g(x) \leq y} f_X(x) dx$$

If $Y = g(X)$ is monotone, $g^{-1}$ exists. If it's increasing, the inverse is as well (vise versa for decreasing)

If $g(X)$ is increasing, $F_Y(y) = F_X(g^{-1}(y))$. If $g(X)$ is decreasing, $F_Y(y) = -F_X(g^{-1}(y))$. In both:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) |\frac{d}{dy} g^{-1}(y)|, & y \epsilon \mathbb{Y} \\ 0, & \text{otherwise} \end{cases}$$

Steps: 1) Find $\mathbb{Y}$ 2) Find $g^{-1}(y)$ 3) Find $\frac{d}{dy} g^{-1}(y)$ 4) Plug into $f_X(g^{-1}) |\frac{d}{dy} g^{-1}(y)$

If the transformation is non-monotonic, all you need to do is find the points of inflection and partition the transformation within ach region of monotinicty

## 8.3 Probability Integral Transform

NOT SURE IF REALLY NEED (7 Oct 2025)

## 8.4 Location Scale Family

Let $f_X(x)$ be a PDF and $\mu \epsilon \mathbb{R}, \sigma > 0$, then $g(x) = \frac{1}{\sigma} f_X(\frac{x - \mu}{\sigma})$ This is the case when there exists a $Z$ such that $X = \mu + \sigma Z$

## 8.5 MonteCarlo Integration

Write an integral as an expectation: $I = \int\limits_a^b h(x) dx = \int\limits_a^b \frac{h(x)}{f_X(x)} f_X(x) dx = \mathbb{E}[\frac{h(x)}{f_X(x)} I_{(a,b)}(x)]$

Steps: 1) Simulate $x_1, ..., x_n$ from $f_X(x)$ 2) Calculate $g(x_j) = \frac{h(x_j)}{f_X(x_j)} I_{(a,b)}^{(x_j)}, \forall j$ 3) $\mathbb{E}[g(x)] \approx \frac{1}{n} \sum\limits_{j=1}^{n} g(x_j) \equiv \bar{g}$

$SE(\bar{g}_n) \approx \frac{1}{\sqrt{n}} s.d.(g(x_1), ..., g(x_n))$

## 8.6 Importance Sampling

FILL IN STUFF

## 8.7 Oct 30

Ex: $X|Z \sim N(Z, \sigma^2), Y|Z \sim N(Z, \sigma^2), (X \perp\!\!\!\perp Y)|Z$ We can say: $X = Z + \epsilon_X, Y = Z + \epsilon_Y : \epsilon_x, \epsilon_Y \sim N(0, \sigma^2), Z \sim N(\mu, \tau^2)$ $Cov(X,Y) = Cov(Z = \epsilon_X, Z + \epsilon_Y) = Cov(Z,Z) = Var(Z) = \tau^2$ For the correlation we need: $Var(Y) = Var(Z) + Var(\epsilon_Y) = \tau^2 + \sigma^2$ $Var(X) = Var(Z) + Var(\epsilon_X) = \tau^2 + \sigma^2$ $Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\tau^2}{\tau^2+\sigma^2}$

## 8.8 Law of Total Covariance

For random variables X, Y, Z, with hierarchy as $(X|Y,Z), (Y|Z),$ and $Z$; $Cov(X,Y) = \mathbb{E}[Cov(X,Y|Z)] + Cov(\mathbb{E}[X|Z], \mathbb{E}[Y|Z])$

# 9 Random Samples and Sums of Random Variables

## 9.1 Definition

The random variables $X_1, ..., X_n$ are a random sample of size n from population $f_X(x)$ if $X_i \overset{\text{iid}}{\sim} f_X(\cdot), i = 1, ..., n$

## 9.2 Joint PDF/PMF

$X_1, ..., X_n$ is a random sample. Since they are *iid*, $f(x_1, ..., x_n) = \prod\limits_{i=1}^{n} f_X(x_i)$

Ex. Let $X_1, ..., X_n$ be the failure times in years of the $i^{th}$ identical circiut components. Assume $X_i \overset{\text{iid}}{\sim} Exp(\beta)$ Thus: $f(x_1, ..., x_n) = \prod\limits_{i=1}^{n} \frac{1}{\beta} e^{-x_i/\beta} = \frac{1}{\beta^n} e^{\frac{1}{\beta} \sum X_i}$ Use this to find: $P(X_1 > 2, X_2 > 2, ..., X_n > 2) = [P(X_1 > 2)]^n = [1 - P(X_1 < 2)]^n = [1 - 1 + e^{\frac{-2}{\beta}}]^n = e^{\frac{-2n}{\beta}}$

Definition: Sampling Distribution - Let $X_1, ..., X_n$ be a random sample of size n. Let $T(X_1, ..., X_n)$ be a real-valued, real-vector function whose domain includes $\mathbb{X}$. Then $T(X_1, ..., X_n)$ is a statistic and its distribution is a sampling distribution.

Common Statistics $\bar{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ $s^2 = \frac{1}{n-1} \sum\limits_{i=1}^{n} (X_i - \bar{X})^2$ Order statistics: media, range, etc.

Theorem: Let $x_1, ..., x_n$ be any numbers and $\bar{x} = \frac{1}{n} \sum\limits_{i=1}^{n} x_i$ then: a. $\underset{a}{min} \sum\limits_{i=1}^{n} (x_i - a)^2 = \sum\limits_{i=1}^{n} (x_i - \bar{x})^2$ -or- $\bar{x} = \underset{a}{argmin} \sum i = 1^n (x_i - a)^2$ b. $(n-1)s^2 = \sum\limits_{i=1}^{n} (x_i - \bar{x})^2 = \sum\limits_{i=1}^{n} x_i^2 - n\bar{x}^2$

Theorem: Let $Z_1, ..., X_n$ be a random sample with population mean and variance $\mu, \sigma^2$, then: 1) $\mathbb{E}(\bar{X}) = \mu$ 2) $Var(\bar{X}) = \frac{\sigma^2}{n}$ 3) $\mathbb{E}(s^2) = \sigma^2$

# 10 Convergence

## 10.1 Convergence of a Sequence

To show a sequence converges to some value $a$, need to show that for every $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that $n \geq N, |a_n - a| < \epsilon$

Proof Framework:

1. Let $\epsilon > 0$ be arbitrary
2. Find $N \in \mathbb{N}$. Typically, $N$ is a function of $\epsilon$
3. Show that for $n \geq N, |a_n - a| < \epsilon$

## 10.2 Divergence of a Sequence

A sequence is said to diverge if it does not converge

Proof Framework:

1. Assume $a_n$ converges to some $L$
2. Show that for some $\epsilon > 0$, there are no positive $N$ which satisfy the convergence criteria

## 10.3 Convergence of a Series

A series is said to converge if its partial sums converge to some limit $L$: $S_n = \sum_{i=1}^{n} a_i$

## 10.4 Convergence of Functions

Pointwise Convergence: For $n \in \mathbb{N}, f_n \to f$ pointwise on A if $\forall x \in A, f_n(x) \to f(x)$

- Need to consider $|f_n(x) - f_n| < \epsilon$ for some $n \leq N$

Uniform Convergence: For $n \in \mathbb{N}, f_n \to f$ uniformly on $A$ if $\forall \epsilon > 0$, there is an $N \in \mathbb{N}$ such that $|f_n(x) = f(x)| < \epsilon, \forall x \in A$

## 10.5 Convergence in Probability

$X_n \xrightarrow{P} X$

$$\forall \epsilon > 0, \lim_{n \to \infty} P(|X - n - X| \geq \epsilon) = 0$$

-OR-

$$\forall \epsilon > 0, \lim_{n \to \infty} P(|X - n - X| < \epsilon) = 1$$

## 10.6 Chebychev's Inequality

Let $X$ be a random variable and $g(x)$ be a non-negative function. Then, for $r > 0$

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

## 10.7 Moments

Let $X$ be a random variable and $g(x) = e^{tX}$. Then:

$$P(X > r) = P(g(x) > g(r)) \leq \frac{\mathbb{E}[g(x)]}{g(r)} = e^{-rt} M_x(t)$$

## 10.8 Weak Law of Large Numbers

Let $X_1, X_2...$ be iid with $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2$.

Define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then: $\overline{X}_n \xrightarrow{P} \mu$

## 10.9 Theorem 5.5.4

Suppose $X_1, X_2... \xrightarrow{P} X$. Let $h(\cdot)$ be a continuous function. Then:

$$h(X_1), h(X_2)... \xrightarrow{P} h(X)$$

## 10.10 Almost Sure Convergence

$X_n \xrightarrow{a.s.} X$

$$\forall \epsilon > 0, P(\lim_{n\to\infty} |X_n - X| < \epsilon) < 1$$

Borel-Cantelli Lemma: Let $E_1, E_2, ...$ be a seuquence of events in some probability space. If the sum of the probability of $E_n$ is finite $\left(\sum_{n=1}^{\infty} P(E_n) < \infty\right)$, then the probability of infinitely many $E_n = 0$

- Useful when we want to show Almost Sure Convergence when $E_n = \{s \in \mathbb{S} : |X_n(s) - X(s)| \geq \epsilon\}$

## 10.11 Strong Law of Large Numbers

Let $X_1, X_2...$ be iid with $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2$. Let $\overline{X}_n = \frac{1}{n}\sum_{n=1}^{n} X_i$.

Then, for $\epsilon > 0, P(\lim_{n\to\infty} |\overline{X}_n - \mu| < \epsilon) = 1$, i.e., $\overline{X}_n \xrightarrow{a.s.} \mu$

## 10.12 Convergence in $r^{th}$ mean

A sequence of r.v. $X_1, X_2...$ converges to $X$ in $r^{th}$ mean if $\lim_{n\to\infty} \mathbb{E}(|X_n - X|^r) = 0$ Written as: $X_n \xrightarrow{L^r} X$

Proof: Assume $X_n \xrightarrow{L^r} X$. We know then that $\lim_{n\to\infty} \mathbb{E}(|X_n - X|^r) = 0$. By Chebychev's, $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = \lim_{n\to\infty} P(|X_n - X|^r > \epsilon^r) \leq \lim_{n\to\infty} \frac{\mathbb{E}(|X_n - X|^r)}{\epsilon^r} = 0$

Ex. Let $U \sim Unif(0,1)$ $X_1 = I^{(u)}_{[0,1]}$ $X_2 = I^{(u)}_{[0,1/2]}, X_3 = I^{(u)}_{[1/2,1]}$ $X_4 = I^{(u)}_{[0,1/3]}, X_5 = I^{(u)}_{[1/3,2/3]}, X_6 = I^{(u)}_{[2/3,1]}$... We showed previously that $X_n \xrightarrow{P} 0$ but $X_n \not\xrightarrow{a.s.} 0$ What about $X_n \xrightarrow{L^r} 0$? Well, $\lim_{n\to\infty} \mathbb{E}(|X_n - 0|^r) = \lim_{n\to\infty} \mathbb{E}((X_n)^r)$ $X_n$ will either be 0 or 1, depending on the $Unif()$. Thus, $= \lim_{n\to\infty} \mathbb{E}((X_n))$. We know $\mathbb{E}(X_2) = P(0 < u < 1/2) = 1/2)$. This can be applied to all $X_n$'s. Thus $= \lim_{n\to\infty} h(n) = 0$. Thus the limit goes to 0, which then means that $X_n \xrightarrow{L^r} 0$. This also implies $X_n \xrightarrow{P} 0$

If $s > r \geq 1$ then $X_n \xrightarrow{L^r} X \implies X_n \xrightarrow{L^r} X$

## 10.13 Convergence in Distribution

A sequence of r.v. $X_1, X_2...$ converges in distribution to r.v. $X$ if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all points $x$ where $F_X(x)$ is continuous. $X_n \xrightarrow{D} X$ is implied by all the other convergences (probability, almost surely, and $r^{th}$ means).

Ex. $X_1, X_2, ... \overset{\text{iid}}{\sim} Unif(0,1)$. Let's examine $X_{(n)}$ as $n \to \infty$. Recall that:

$$F_{X_{(n)}}(x) = \begin{cases} 0, & x < 0 \\ x^n, & x \in [0,1) \\ 1, & x \geq 1 \end{cases}$$

$lim_{n\to\infty} F_{X_{(n)}}(x) = \begin{cases} 0, & x < 0 \\ lim_{n\to\infty} x^n, & x \in [0,1) \\ 1, & x \geq 1 \end{cases} = \begin{cases} 0, & x < 1 \\ 1, & x \geq 1 \end{cases}$ Note that this is the CDF for a point mass (step function) at 1, so $X_{(n)} \overset{D}{\to} 1$.

Theorem: For a sequence of random variables $X_{(n)} \overset{P}{\to} C$ iff $X_{(n)} \overset{D}{\to} C$ In words, $P(|X_n - C| > \epsilon) \to 0 \forall \epsilon$ is equivalent to $lim_{n\to\infty} F_{X_{(n)}}(x) = \begin{cases} 0, & x < 1 \\ 1, & x \geq 1 \end{cases}$

## 10.14 Slowing Down Convergence

Sometimes it is helpful to 'slow down' convergence so the limiting distribution isn't a constant. As an example, consider $Y_n = n(1 - X_{(n)})$ $F(Y_n) = P(Y_N \leq y) = P(n(1 - X_{(n)})) = P(1 - \frac{y}{n} \leq X_{(n)}) = 1 - P(X_{(n)} \leq 1 - \frac{y}{n}) = 1 - (1 - \frac{y}{n})^n$ Thus, $lim_{n\to\infty} F(Y_n) = lim_{n\to\infty} 1 - (1 - \frac{y}{n})^n = 1 - lim_{n\to\infty}(1 - \frac{y}{n})^n = 1 - e^{-y}$ Thus $Y_n \overset{D}{\to} Y \sim Exp(1)$

## 10.15 Central Limit Theorem

Consider $X_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ for a sequence $X_1, X_2, ....$ We have shown: $\bar{X_n} \overset{P}{\to} \mu$, $\bar{X_n} \overset{a.s.}{\to} \mu$ Let $X_1, X_2, ...$ be a seuqnce of iid r.v. whose MGFs exist in a neighborhood of 0. Let $\mathbb{E}(X_i) = \mu$, $Var(X_i) = \sigma^2 > 0$.

Let $X_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Let $G_n(x)$ be the CDF of $\frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma}$ or $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ Then for any $x \in \mathbb{R} : lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}y^2} dy =$ CDF of Standard Normal. Thus, $\frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} \overset{D}{\to} N(0,1)$, or $Z_n = \frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma}$ and $Z \sim N(0,1), Z_n \overset{D}{\to} Z$

Proof: Let $Y_i = \frac{X_i - \mu}{\sigma}$, $\mathbb{E}(Y_i) = 0$, $Var(Y_i) = 1$. $\frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^{n} X_i - n\mu\right) = \frac{1}{\sigma\sqrt{n}}(n\bar{X_n} - n\mu)$ $= \frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} = Z_n$ By Thm. 2.3.15, Thm 4.6.7, $M_{Z_n}(t) = \left[M_Y(\frac{t}{\sqrt{n}})\right]^n$

Can also show using a Taylor Series of $M_Y(\frac{t}{\sqrt{n}})$ about 0:

$$M_Y(\frac{t}{\sqrt{n}}) = M_Y(0) + M_Y'(0)(\frac{t}{sqrtn} - 0) + \frac{M_y''(0)}{2!}(\frac{t}{\sqrt{n}} - 0)^2 + R_n 1 + 0 + \frac{t^2}{2n} + R_n$$

By Thm. 5.5.14, $R_n \to 0$ faster than $\frac{t^2}{2n}$ so $M_{Z_n}(t) = \left[1 + \frac{t^2}{2n} + R_n\right]^n$

Thus: $lim_{n\to\infty} M_{Z_n}(t) = lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + R_n\right]^n = e^{t^2/2} = M_Z(t) =$ MGF of N(0,1)

## 10.16 Slutsky's Theorem

If $X_n \overset{D}{\to} X$ and $Y_n \overset{P}{\to} a$ then:

1. $Y_n X_n \overset{D}{\to} aX$
2. $X_n + Y_n \overset{D}{\to} X + a$

## 10.17 Delta Method

Let $Y_1, Y_2, ...$ be random variables where $\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2)$. Then, for any $g(\theta)$ where $g'(\theta)$ exists and is not 0:

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{D} N(o, \sigma^2(g'(\theta))^2)$$

## 10.18 2nd Order Delta Method

Let $Y_1, Y_2, ...$ be random variables where $\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2)$. Then, for any $g(\theta)$ where $g'(\theta) = 0$ and $g''(\theta) \neq 0$:

$$n(g(Y_n) - g(\theta)) \xrightarrow{D} g''(\theta)\frac{\sigma^2}{2}\chi_1^2$$

# 11 Exponential Families

A family of PDFs/PMFs with parameter vector *statvec*$\theta$ is called an Exponential Family if it can be expressed as:

$$f(x|\underset{\sim}{\boldsymbol{\theta}}) = h(x)c(\underset{\sim}{\boldsymbol{\theta}})exp\left[\sum_{i=1}^{K} w_i(\underset{\sim}{\boldsymbol{\theta}})t_i(x)\right]$$

In its canonical (natural) form:

$$f(x|\underset{\sim}{\boldsymbol{\theta}}) = h(x)c^*(\underset{\sim}{\boldsymbol{\eta}})exp\left[\sum_{i=1}^{K} \eta_i t_i(x)\right]$$

## 11.1 Expected Value Theorem

Where $X$ is a random vairable belonging to the Exponential Family:

$$\mathbb{E}\left[\sum_{i=1}^{K} \frac{\partial w_i(\underset{\sim}{\boldsymbol{\theta}})}{\partial \theta_j} t_i(x)\right] = \frac{-\partial ln(c(\underset{\sim}{\boldsymbol{\theta}}))}{\partial \theta_j}$$

$$Var\left(\sum_{i=1}^{K} \frac{\partial w_i(\underset{\sim}{\boldsymbol{\theta}})}{\partial \theta_j} t_i(x)\right) = \frac{-\partial^2 ln(c(\underset{\sim}{\boldsymbol{\theta}}))}{\partial^2 \theta_j} - \mathbb{E}\left[\sum_{i=1}^{K} \frac{\partial^2 w_i(\underset{\sim}{\boldsymbol{\theta}})}{\partial^2 \theta_j} t_i(x)\right]$$

In canonical form:

$$\mathbb{E}[t_j(x)] = \frac{-\partial ln(c^*(\eta))}{\partial \eta_j}$$

$$Var[t_j(x)] = \frac{-\partial^2 ln(c^*(\eta))}{\partial^2 \eta_j}$$