# Sets

## 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: $A^c : \{x \epsilon \mathbb{S} : x \notin A\}$
Difference: $A - B : \{x \epsilon \mathbb{S} : x \epsilon A, x \notin B\}$
Infinite Union: $\bigcup_{i=1}^{\infty} A_i : \{x \epsilon \mathbb{S}, x \epsilon A_i \ni A_i\}$
Infinite Intersection: $\bigcap_{i=1}^{\infty} A_i : \{x \epsilon \mathbb{S}, x \epsilon A_i \forall A_i\}$

## 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: $A \cap B = \{\}$

## 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: $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$

# Sigma Algebras

## 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. $A_1, A_2, ... \epsilon \mathbb{B} \implies \bigcup_{n=1}^{\infty} A_n \epsilon \mathbb{B}$

## 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

# Probability Functions

## 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 $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$

## 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 $i \neq j$, $\bigcup_{i=1}^{\infty} c_i = \mathbb{S}$)
- $P(A) = \sum_{i=1}^{\infty} P(A \cap c_i)$
8) For any $A_1, A_2, ...$; $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$

# Counting

## Sampling

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

## 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:
- $N_{A^c} = |A^c| \therefore N_A = N - N_{A^c}$

## Continous

Consider $\mathbb{S} \subset \mathbb{R}^d$ with uniform probability

Then for $A \subseteq \mathbb{S}, P(A) = \frac{\int\limits_A ds}{\int\limits_{\mathbb{S}} ds}$

## 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 $i \neq j$, $\bigcup\limits_{i=1}^{\infty} c_i = \mathbb{S}$):
$P(B) = \sum\limits_{i=1}^{n} P(B|c_i)P(c_i)$

## 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:
- $P(\bigcap\limits_{j=1}^{k} A_{i_j}) = \prod\limits_{j=1}^{k} P(A_{i_j})$

## Conditional Independence

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

# Random Variables

## Definition

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

## 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. $\lim\limits_{x \downarrow x_0} F_X(x) = F_X(x_0)$

## 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- $\int\limits_{x} f_X(x)dx = 1$

## Kernel

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

## 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: $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

# Expected Value

## 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 $Y = g(x)$
- $\mathbb{E}[g(x)] = \int\limits_{-\infty}^{\infty} g(x)f_X(x)dx = \int\limits_{-\infty}^{\infty} Yf_Y(y)dy = \mathbb{E}[Y]$

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

## 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 $a \leq \mathbb{E}[g(x)] \leq b$

# Moments

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

## 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

## 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 $-h < t < h$

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

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

# Transformations

## 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 $\mathbb{S} \to \mathbb{X} \to \mathbb{Y}$
- $\forall A, P(Y \epsilon A) = P(g(X) \epsilon A) = P(\{x \epsilon \mathbb{X} : g(x) = A\}) = P(X \epsilon g^{-1}(A))$

## 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

## Probability Integral Transform

NOT SURE IF REALLY NEED (7 Oct 2025)

## 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 $X = \mu + \sigma Z$

## 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) $\mathbb{E}[g(x)] \approx \frac{1}{n}\sum\limits_{j=1}^n g(x_j) \equiv \bar{g}$

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

## Importance Sampling

FILL IN STUFF

## 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:

$Var(Y) = Var(Z) + Var(\epsilon_Y) = \tau^2 + \sigma^2$

$Var(X) = Var(Z) + Var(\epsilon_X) = \tau^2 + \sigma^2$

$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\tau^2}{\tau^2+\sigma^2}$

## Law of Total Covariance

For random variables X, Y, Z, with hierarchy as $(X|Y,Z), (Y|Z)$, and $Z$;

$Cov(X,Y) = \mathbb{E}[Cov(X,Y|Z)] + Cov(\mathbb{E}[X|Z], \mathbb{E}[Y|Z])$

# Random Samples and Sums of Random Variables

## Definition

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

## 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. $\min\limits_a \sum\limits_{i=1}^n (x_i - a)^2 = \sum\limits_{i=1}^n (x_i - \bar{x})^2$ -or- $\bar{x} = \arg\min\limits_a \sum_i = 1^n(x_i - a)^2$

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) $\mathbb{E}(s^2) = \sigma^2$