ST2334 AY23/24 Sem 2 Finals Cheat Sheet

Chapter 1: Probability

Sample Space	• The sample space, denoted by S, is the set of ALL possible outcomes of a statistical experiment. • The sample space depends on the problem of interest. • An **event** is a subset of a sample space.
Notation	For a finite set A, $\lvert A\rvert$ denotes the number of elements in A.
Equally Likely Probability	If S is a finite sample space in which all outcomes are equally likely and E is an event in S, then the **probability** of E, denoted $P(E)$, is $P(E) = \frac{\text{The number of outcomes in } E}{\text{The total number of outcomes in } S} = \frac{\lvert E\rvert}{\lvert S\rvert}$
Statistical Experiment	A Statistical Experiment is any procedure that produces data/ observations.
Sample Point	A sample point is an outcome (element) in the sample space
Event	An event is a subset of the sample space.

- The sample space is itself an event, and is called a sure event
- An event that contains NO ELEMENTS is the empty set, denoted by $\emptyset$, aka null event.

Event Operation & Relationship Laws

Basic		Distributive Law
$A\cap A' = \emptyset$	$A\cap\emptyset = \emptyset$	$A\cup(B\cap C) = (A\cup B)\cap(A\cup C)$
$A\cup A' = S$	$(A')' = A$	$A\cap(B\cup C) = (A\cap B)\cup(A\cap C)$
Set Union Law with Complement		**Absorption Law**
$A\cup B = A\cup(B\cap A')$		$A = (A\cap B)\cup(A\cap B')$
De Morgan's Law		
$(A_1\cup A_2\cup\ldots\cup A_N)' = A'_1\cap A'_2\cap\ldots\cap A'_N$ Note: $(A\cup B)' = A'\cap B$		$(A_1\cap A_2\cap\ldots\cap A_N)' = A'_1\cup A'_2\cup\ldots\cup A'_N$ Note: $(A\cap B)' = A'\cup B$

$$P(n,r) = \frac{n!}{(n-r)!} = n(n-1)(n-2)\ldots(n-r+1)$$

	Order Matters	*Order Don't Matter*
Repetition is Allowed	n^k	$\binom{k+n-1}{k}$
Repetition is NOT allowed	$P(n,k)$	$\binom{n}{k}$

Probability Axioms

Let S be a sample space. A probability function P from the set of all events in S to the set of real numbers satisfies the following axioms: For all events A and B in S,
1. $0 \le P(A) \le 1$
2. $P(\emptyset) = 0$ and $P(S) = 1$
3. If A and B are disjoint events ($A\cap B = \emptyset$), then (i.e A & B are mutually exclusive events) $P(A\cup B) = P(A) + P(B)$

Basic Properties of Probabilities

Proposition 1: The probability of the empty set $\emptyset$ is $P(\emptyset) = 0$	
Proposition 2: If $A_1, A_2, \ldots, A_N$ are mutually exclusive events, that is $A_i\cap A_j = \emptyset$ for any $i \ne j$, then $P(A_1\cup A_2\cup\ldots\cup A_N) = P(A_1) + P(A_2) + \ldots + P(A_N)$	
Proposition 3: Complement Rule For any event A, we have: $P(A') = 1 - P(A)$	**Proposition 4:** For any 2 events A & B, $P(A) = P(A\cap B) + P(A\cap B')$
Proposition 5: General Union 2 Events For any events A & B, $P(A\cup B) = P(A) + P(B) - P(A\cap B)$	**Proposition 6:** If $A\subset B$, then $P(A) \le P(B)$ $P(A) = P(A\mid B)\cdot P(B) + P(A\mid B')\cdot P(B')$

Independence, Mutual Exclusivity

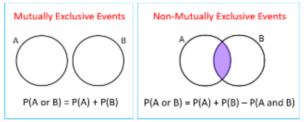
ME	• 2 events CANNOT occur at the same time • A, B *mutually exclusive* $\Leftrightarrow P(A\cap B) = \emptyset$
Independent Indep → $\perp$ dep → $\not\perp$	$P(A) = \frac{P(A\cap B)}{P(B)} \Rightarrow P(A)\times P(B) = P(A\cap B)$ • A, B *independent* $\Leftrightarrow P(A\cap B) = P(A)P(B)$ • If independent & $P(A) \ne 0$ → $P(B\mid A) = P(B)$ • If independent & $P(B) \ne 0$ → $P(A\mid B) = P(A)$
Complement	$P(A') = 1 - P(A)$
Expected Value	$\sum_{k=1}^{n} a_k p_k = a_1p_1 + a_2p_2 + a_3p_3 + \cdots + a_np_n$

Conditional Probability

$$P(B\mid A) = \frac{P(A\cap B)}{P(A)} \quad -(1)$$

Multiplying both sides of (1) by P(A)	Dividing both sides of (2) by P(B|A)
$P(A\cap B) = P(B\mid A)\cdot P(A) \quad -(2)$	$P(A) = \frac{P(A\cap B)}{P(B\mid A)} \quad -(3)$
Multiplication Rule	**Inverse Probability Formula**
$P(A\cap B) = P(B\mid A)\cdot P(A), \text{if } P(A) \ne 0$ $P(A\cap B) = P(A\mid B)\cdot P(B), \text{if } P(B) \ne 0$	$P(B\mid A) = \frac{P(A\cap B)}{P(A)}$ Then inverse: $P(A\mid B) = \frac{P(A)P(B\mid A)}{P(B)}$

False Positive	False Negative	Sensitivity	Specificity
P(+'ve | D')	P(-'ve | D)	P(+'ve | D)	P(-'ve | D')



Partition, Law of Total Probability

$$P(B) = \sum_{i=1}^{N} P(B\cap A_i) = \sum_{i=1}^{N} P(A_i)\,P(B\mid A_i)$$

For any events A & B, we have: $P(B) = P(A)P(B\mid A) + P(A')P(B\mid A')$

Bayes Theorem:

K variables	$P(B_k\mid A) = \frac{P(A\mid B_k)\cdot P(A_k)}{\sum_{i=1}^{n} P(B_i)P(A\mid B_i)}$ $P(B_k\mid A) = \frac{P(A\mid B_k)\cdot P(B_k)}{P(A\mid B_1)\cdot P(B_1) + P(A\mid B_2)\cdot P(B_2) + \cdots + P(A\mid B_n)\cdot P(B_n)}$
2 variables	$P(B\mid A) = \frac{P(A\mid B)\cdot P(B)}{P(A)} = \frac{P(A\mid B)\cdot P(B)}{P(A\mid B)\cdot P(B) + P(A\mid B')\cdot P(B')} = \frac{P(A\cap B)}{P(A)}$

Pairwise Independent/ Mutually Independent

- Events are mutually independent IFF 4 conditions are satisfied:

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

- Events can be pairwise independent without satisfying the condition $P(A\cap B\cap C) = P(A)\bullet P(B)\bullet P(C)$
- Conversely, they can satisfy the condition $P(A\cap B\cap C) = P(A)\bullet P(B)\bullet P(C)$ without being pairwise independent.

Mutually Independent:
$P(A_1\cap A_2\cap\ldots\cap A_n) = P(A_1)\bullet P(A_2)\bullet\ldots\bullet P(A_n)$

Chapter 2: Random Variables

Probability Mass Function (PMF)	$f(x) = \begin{cases} P(X=x) & \text{for } x\in R_X \\ 0 & \text{for } x\notin R_X \end{cases}$ Properties of PMF: The pmf, $f(x)$ of a discrete random variable MUST satisfy these conditions: 1) $f(x_i) \ge 0$ for all $x_i\in R_X$ 2) $f(x_i) = 0$ for all $x_i\notin R_X$ 3) $\sum_{i=1}^{\infty} f(x_i) = 1 \;\; OR \sum_{x_i\in R_X} f(x_i) = 1$ For any set $B\subset\mathbb{R}$, we have: $P(X\in B) = \sum_{x_i\in B\cap R_X} f(x_i)$
Probability Density Function (PDF)	1) $f(x) \ge 0$ for all $x\in R_X$; $f(x) = 0$ for $x\notin R_X$ 2) $\int_{R_X} f(x)dx = 1$ This is equivalent to: $\int_{-\infty}^{\infty} f(x)\,dx = 1$ Since $f(x) = 0$ for $x\notin R_X$ 3) For any a and b such that $a \le b$: $P(a\le X\le b) = \int_a^b f(x)\,dx$ Note: $P(a<X<b) = P(a<X\le b) = P(a\le X<b) =$ $P(a\le X\le b) = \int_a^b f(x)\,dx$. They all represent the area under the graph $f(x)$ between $x = a$ and $x = b$ To check if pdf: 1. $f(x) \ge 0$ for all $x\in R_X$; $f(x) = 0$ for $x\notin R_X$ 2. $\int_{R_X} f(x)dx = 1$

Cumulative Distribution Function (CDF) $F(x) = P(X\le x)$

Discrete	$F(x) = \sum_{t\in R_X; t\le x}$, $f(t) = \sum_{t\in R_X; t\le x} P(X=t)$ • The cumulative distribution function of a DRV is a step function. • For any 2 numbers $a < b$, we have: $P(a\le X\le b) = P(X\le b) - P(X<a) = F(b) - F(a-)$ $F(a-) = \lim_{x\uparrow a} F(x)$
Continuous	$F(x) = \int_{-\infty}^{x} f(t)\,dt$, $f(x) = \frac{dF(x)}{dx}$ Further: $P(a\le X\le b) = P(a<X<b) = F(b) - F(a)$

- Discrete → Summation; Continuous → Integrate
- The ranges of $F(x)$ and $f(x)$ satisfy the following conditions:
  1. $0 \le F(x) \le 1$
  2. For discrete distributions, $0 \le f(x) < 1$
  3. For continuous distributions, $0 \le f(x)$, but NOT NECESSARILY that $f(x) \le 1$

Expectation & Variance **expectation = mean

Expectation for DRV	Expectation for CRV
$\mu_X = E(X) = \sum_{x_i\in R_X} x_i f(x_i)$ $= \sum_{x_i\in R_X} x_i P(X=x) = \frac{\sum f(x)}{\sum f}$	$\mu_X = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$ $= \int_{x\in R_X} x f(x)\,dx$
Let g(•) be an arbitrary function	
$E[g(X)] = \sum_{x\in R_X} g(x)f(x)$	$E[g(X)] = \int_{R_X} g(x)f(x)\,dx$
Variance	
$\sigma_X^2 = V(X) = E(X-\mu_X)^2 = E(X^2) - [E(X)]^2$	

Note:
- $V(X) \ge 0$ for any X.
- Equality holds iff $P(X = E(X)) = 1$, that is when X is a constant
- The positive root of the variance = standard deviation of X

$$\sigma_X = \sqrt{V(X)}$$

Variance for DRV	Variance for CRV
$V(X) = \sum_{x\in R_X} (x-\mu_X)^2 f(x)$	$V(X) = \int_{-\infty}^{\infty} (x-\mu_X)^2 f(x)\,dx$

Basic Properties of Expectations & Variance:

Expectation	Variance
a) $E(a) = a$ b) $E(aX) = aE(X)$ c) $E(aX\pm b) = aE(X)\pm b$ d) $E(aX\pm bY) = aE(X)\pm bE(Y)$ e) $E(x_1+x_2+\cdots+x_n) = nE(X)$ $E(a_1X_1+\ldots+a_kX_k) = a_1E(X_1)+\ldots+a_kE(X_k)$	a) $V(a) = 0$ b) $V(aX) = a^2V(X)$ c) $V(aX\pm b) = a^2V(X)$ d) $V(aX\pm bY) = a^2V(X)\pm b^2V(Y)$ e) $V(x_1+x_2+\cdots+x_n) = nV(X)$

Chapter 3: Joint Distributions

2D Random Vector	
Let E be an experiment and S be a corresponding sample space. Suppose X and Y are two functions each assigning a real number to each $s\in S$. We call (X,Y) a **2D random vector**, or a **2Drandom variable**.	
2D Discrete	(X,Y) is a **discrete 2D random variable** if the number of possible values of $(X(s), Y(s))$ are finite/countable. That is, the possible values of $(X(s),Y(s))$ may be represented by: (x_i, y_j), $i = 1,2,3,\ldots$; $j = 1,2,3,\ldots$
2D Continuous	(X,Y) is a **continuous 2D random variable** if the possible values of $(X(s),Y(s))$ can assume any value in some region of the Euclidean space $\mathbb{R}^2$.
We can view X and Y separately to JUDGE whether (X,Y) is discrete or cont. • If both X and Y are **discrete** random variables → (X,Y) is **discrete**. • If both X and Y are **continuous** random variables → (X,Y) is **continuous**.	
n-Dimensional Random Vector	
Let $X_1, X_2, \ldots, X_n$ be n functions each assigning a real number to all outcome $s\in S$. We call $(X_1, X_2, \ldots, X_n)$ a **n-dimensional random vector**, or a **n-dimensional random variable**.	

Discrete Joint Probability Function

$$f_{X,Y}(x,y) = P(X=x, Y=y), \text{for } (x,y)\in R_{X,Y}$$

Properties of Discrete Joint Probability Function
1. $f_{X,Y}(x,y) \ge 0$ for any $(x,y)\in R_{X,Y}$
2. $f_{X,Y}(x,y) = 0$ for any $(x,y)\notin R_{X,Y}$
3. $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f_{X,Y}(x_i,y_j) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P(X=x_i, Y=y_j) = 1$
 $\sum\sum_{(x,y)\in R_{X,Y}} f(x,y) = 1$
4. Let A be any subset of $R_{X,Y}$, then: $P((X,Y)\in A) = \sum\sum_{(x,y)\in A} f_{X,Y}(x,y)$

Continuous Joint Probability Function

$$P((X,Y)\in D) = \iint_{(x,y)\in D} f_{X,Y}(x,y)\,dydx$$

For any $D\subset\mathbb{R}^2$, more specifically:

$$P(a\le X\le b, c\le Y\le d) = \int_a^b\int_c^d f_{X,Y}(x,y)\,dydx$$

Properties of Continuous Joint Probability Function
1. $f_{X,Y}(x,y) \ge 0$ for any $(x,y)\in R_{X,Y}$
2. $f_{X,Y}(x,y) = 0$ for any $(x,y)\notin R_{X,Y}$
3. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dydx = 1$, $\iint_{(x,y)\in D} f_{X,Y}(x,y)\,dydx = 1$

Marginal Probability Distribution

Y is DRV (Discrete)	For any x: $f_X(x) = \sum_y f_{X,Y}(x,y)$
Y is CRV (Continuous)	For any x: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$

- Marginal distribution is like a "projection" of the 2D function $f_{X,Y}(x,y)$ onto the 1D function.
- The marginal distribution of X is the individual distribution of X ignoring the values of Y.
- $f_X(x)$ is a probability function; so it satisfies all the properties of the probability function

Conditional Distribution

Given $f_X(x) > 0$, cond prob fn:	Given $f_Y(y) > 0$, cond prob fn:
$f_{Y\mid X}(y\mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$	$f_{X\mid Y}(x\mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
If $f_X(x) > 0, f_{X,Y}(x,y) = f_X(x)f_{Y\mid X}(y\mid x)$,	*If* $f_Y(y) > 0, f_{X,Y}(x,y) = f_Y(y)f_{X\mid Y}(x\mid y)$
Discrete	$P(Y=y\mid X=x) = \frac{P(X=x, Y=y)}{P(X=x)} = \frac{f_{X,Y}(x,y)}{f_X(x)}$
Continuous	$P(Y\le y\mid X=x) = \int_{-\infty}^{y} f_{Y\mid X}(y\mid x)\,dy$ $E(Y\mid X=x) = \int_{-\infty}^{\infty} y\, f_{Y\mid X}(y\mid x)\,dy$

$$\text{conditional distribution} = \frac{\text{joint density}}{\text{marginal distribution}}$$

Independent Random Variables

- Random variables X and Y are independent IFF for any x and y: $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
- Random variables $X_1, X_2, \ldots, X_n$ are independent IFF for any $x_1, x_2, \ldots, x_n$:

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n)$$

- Just check if their joint probability = product of their individual probabilities
  - If $R_{X,Y}$ is NOT a product space → X and Y are NOT independent

Properties of Independent Random Variables

Suppose X, Y are independent random variables:
1. If A and B are arbitrary subsets of $\mathbb{R}$, the events $X\in A$ and $Y\in B$ are independent events in S. As such: $P(X\in A; Y\in B) = P(X\in A)P(Y\in B)$
 For any real numbers x, y: $P(X\le x; Y\le y) = P(X\le x)P(Y\le y)$
2. For arbitrary functions $g_1(\bullet)$ and $g_2(\bullet)$, $g_1(X)$ and $g_2(Y)$ are independent. For example:
   - X^2 and Y are independent.
   - $sin(X)$ and $cos(Y)$ are independent.
   - e^X and $log(Y)$ are independent.
3. Independence is connected with conditional distribution.
   - If $f_X(x) > 0$, then $f_{Y\mid X}(y\mid x) = f_Y(y)$
   - If $f_Y(y) > 0$, then $f_{X\mid Y}(x\mid y) = f_X(x)$

CHECKING INDEPENDENCE
We have a handy way to check independence.
X and Y are independent if and only if both of the following hold:
(a) $R_{X,Y}$, the range where the probability function is positive, is a product space.
(b) For any $(x,y)\in R_{X,Y}$, we have

$$\boxed{f_{X,Y}(x,y) = C\times g_1(x)\times g_2(y).}$$

That is, $f_{X,Y}(x,y)$ can be "factorized" as the product of two functions g_1 and g_2, where g_1 **depends on** x **only**, g_2 **depends on** y **only**, and C is a constant not depending on both x and y.
Note: $g_1(x)$ and $g_2(y)$ on their own NEED NOT be probability functions.

Expectation & Covariance

Definition 9: Expectation of 2-Dimensional Random Variables	
Consider any 2-variable function $g(x,y)$	
If (x,y) is DRV (Discrete)	$E[g(X,Y)] = \sum_x\sum_y g(x,y)f_{X,Y}(x,y)$
If (x,y) is CRV (Continuous)	$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)\,dydx$
Note: If we let $g(X,Y) = (X-E(X))(Y-E(Y)) = (X-\mu_X)(Y-\mu_Y)$ The expectation $E[g(X,Y)]$ leads to the covariance of X and Y.	
Covariance	
The covariance of X and Y is defined to be: $cov(X,Y) = E[(X-E(X))(Y-E(Y))]$	
If X and Y are DRV (Discrete)	$cov(X,Y) = \sum_x\sum_y (x-\mu_X)(y-\mu_Y)f_{X,Y}(x,y)$
If X and Y are CRV (Continuous)	$cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f_{X,Y}(x,y)\,dydx$

Properties of the Covariance
1. $cov(X,Y) = E(XY) - E(X)E(Y)$
2. If X and Y are independent, then $cov(X,Y) = 0$
 However, $cov(X,Y) = 0$ does not imply independence (1 way relation). i.e:
   1. $X\perp Y \Rightarrow cov(X,Y) = 0$ (X &Y independent → cov = 0)
      - Since $E(XY) = E(X)E(Y) \rightarrow cov(X,Y) = 0$
   2. $cov(X,Y) = 0 \not\Rightarrow X\perp Y$ (cov = 0 does not imply independence)
3. $cov(aX+b, cY+d) = ac\bullet cov(X,Y)$
   1. $cov(X,Y) = cov(Y,X)$
   2. $cov(X+b, Y) = cov(X,Y)$
   3. $cov(aX,Y) = a\bullet cov(X,Y)$
4. $V(aX+bY) = a^2V(X) + b^2V(Y) + 2ab\bullet cov(X,Y)$
   1. $V(aX) = a^2V(X)$
   2. $V(X+Y) = V(X) + V(Y) + 2cov(X,Y)$

Properties of Variance and Covariance

Using $V(X+Y) = V(X) + V(Y) + 2cov(X,Y)$, we can derive the following:
1. For random variables X and Y that are independent, we have: $V(X\pm Y) = V(X)\pm V(Y)$
2. For random variables $X_1, X_2, \ldots, X_n$, we have: (Not independent) $V(X_1+X_2+\ldots+X_n) = V(X_1) + V(X_2) + \ldots + V(X_n) + 2\sum_{j>i} cov(X_i, X_j)$
3. For random variables $X_1, X_2, \ldots, X_n$ that are independent, we have: $V(X_1\pm X_2\pm\ldots\pm X_n) = V(X_1)\pm V(X_2)\pm\ldots\pm V(X_n)$

$cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_1-E(X))(y_1-E(Y))$	$V(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_1-E(X))^2$

$f(x,y)$		$x=2$	$x=4$	$f_Y(y)$
y	1	0.10	0.15	0.25
	3	0.20	0.30	0.50
	5	0.10	0.15	0.25
$f_X(x)$		0.40	0.60	1

To test if X and Y are independent:
-Check $f_{X,Y}(x,y) = f_X(x)\bullet f_Y(y)$ for all possible combinations.

	Baseball	Basketball	Football	Total
Male	13	15	20	48
Female	23	16	13	52
Total	36	31	33	100

Marginal distribution of gender

Marginal distribution of sports

Markov's inequality	$P(X\ge c) \le \frac{E[X]}{c}$
Chebyshev's inequality	$P(\lvert X - E[X]\rvert \ge k) \le \frac{Var[X]}{k^2}$
Jensen's inequality	$E[g(X)] \ge g(E[X])$ if g convex

Chapter 4: Special Probability Distributions

In a Discrete Uniform Distribution:

Expectation of X	$\mu_X = E(X) = \sum_{i=1}^{k} x_i f_X(x_i) = \frac{1}{k}\sum_{i=1}^{k} x_i$
Variance of X	$\sigma_X^2 = V(X) = E(X^2) - (E(X))^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \mu_X^2$

Note the Probability Equivalence:

- $P(X > a) = 1 - P(X \le a)$
- $P(X < a) = P(X \le (a-1))$
- $P(b \le X \le a) = P(X \le a) - P(X \le (b-1))$
- $P(b < X < a) = P(X \le (a-1)) - P(X \le b)$
- $P(X \ge a) = 1 - P(X \le (a-1))$
- $P(X \le a) = P(X \le a)$

Different Probability Distributions:

Bernoulli

$X \sim Bernoulli(p)$

$$f_X(x) = P(X = x) = \begin{cases} p, & x = 1; \\ 1-p, & x = 0; \end{cases}$$
$$= p^x(1-p)^{1-x}, for\ x = 0, 1$$

Expectation	$\mu_X = E(X) = p$
Variance	$\sigma_X^2 = V(X) = p(1-p) = pq$

Binomial

$X \sim Bin(n, p)$

$$(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad for\ x = 0,1,2,3,\ldots,n$$

Expectation	$E(X) = np$
Variance	$V(X) = np(1-p)$

- $E(X) = E(X_1) + \cdots + E(X_n) = p + \cdots + p = np$
- $V(X) = V(X_1 + \ldots + X_n) = V(X_1) + \ldots + V(X_n) = pq + \ldots + pq = npq$

Negative Binomial

$X \sim NB(k, p)$

$$f_X(x) = P(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \quad for\ x = k, k+1, k+2, 3, \ldots$$

Expectation	$E(X) = \frac{k}{p}$
Variance	$V(X) = \frac{(1-p)k}{p^2}$

Geometric

$X \sim Geom(p)$

$$f_X(x) = P(X = x) = (1-p)^{x-1} p$$

Expectation	$E(X) = \frac{1}{p}$
Variance	$V(X) = \frac{1-p}{p^2}$

Poisson

$X \sim Poisson(\lambda)$, where $\lambda > 0$

$$f_X(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!},$$
$where\ k = 0,1,\ldots is\ the\ \#occurence\ of\ such\ events$

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

Note:

- We can approx a binomial distribution using Poisson Approx
  - Given $n \to \infty$ and $p \to 0$ in such a way that $\lambda = np$ remains a constant
    - $n \ge 20$ and $p \le 0.05$ OR $n \ge 100$ and $np \le 10$

Poisson Approximation (~" Binomial)
$\lim_{p \to 0; n \to \infty} P(X = x) = \frac{e^{-np}(np)^x}{x!}$

Continuous Uniform

$X \sim U(a, b)$

PDF	CDF
$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b; \\ 0, & otherwise \end{cases}$	$F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$

Expectation	$E(X) = \frac{a+b}{2}$
Variance	$V(X) = E(X^2) - (E(X))^2 = \frac{(b-a)^2}{12}$

$F_X(x) = 0$ when $x < a$ and $F_X(x) = 1$ when $x > b$

When $a \le x \le b$:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{-\infty}^{a} 0\,dt + \int_{a}^{x} \frac{1}{b-a}\,dt = \frac{1}{b-a}[t]_a^x = \frac{x-a}{b-a}$$

Exponential

$X \sim Exp(\lambda)$

1st form	2nd form
$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & if\ x \ge 0; \\ 0, & if\ x < 0 \end{cases}$	$f_X(x) = \begin{cases} \frac{1}{\mu} e^{-x/\mu}, & x \ge 0; \\ 0, & x < 0 \end{cases}$
$E(X) = \frac{1}{\lambda}$	$E(X) = \mu$
$V(X) = E(X^2) - (E(X))^2 = \frac{1}{\lambda^2}$	$V(X) = E(X^2) - (E(X))^2 = \mu^2$

- Inverse relationship $\mu = \frac{1}{\lambda}$
- Suppose that X has an Exponential Distribution with $\lambda > 0$. Then for any 2 +'ve number s and t, we have:
  $$P(X > s + t \mid X > s) = P(X > t)$$
- Exponential Distribution has "no memory"/ "memoryless"
- Exponential Distribution is the ONLY memoryless CRV.

Proof Memoryless: Given $X \sim Exp(\lambda)$, we check that:

$$P(X > s + t \mid X > s) = \frac{P(\{X > s + t\} \cap \{X > s\})}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)}$$
$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda(s)}}$$
$$= e^{-\lambda(t)} = P(X > t)$$

Normal

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

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

Expectation	$E(X) = \mu$
Variance	$V(X) = E(X^2) - (E(X))^2 = \sigma^2$

Properties of Normal Distributions:

- Total area under curve = 1
- If 2 curves have same V(X), σ^2 → Same shape, maybe diff points.
- The **larger the mean**, the more the curve **shifts right.**
- As σ(s.d) increases → curve flattens (Larger range of values)

Standardized Normal Distribution

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

- $E(Z) = 0$ and $V(Z) = 1$.
  $$pdf\ of\ Z = \phi(z) = f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$
  $$cdf\ of\ Z = \Phi(z) = \int_{-\infty}^{z} \phi(t)\,dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\,dt$$
- For any $X \sim N(\mu, \sigma^2)$ and any real numbers x_1, x_2, where $P(x_1 < X < x_2)$
  $$x_1 < X < x_2 \Leftrightarrow \frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma}$$
  $$P(x_1 < X < x_2) = P(z_1 < Z < z_2), where z_1 = \frac{x_1 - \mu}{\sigma}, z_2 = \frac{x_2 - \mu}{\sigma}$$
  $$P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right)$$

Properties of Standard Normal Distribution

1. $P(Z \ge 0) = P(Z \le 0) = \Phi(0) = 0.5$
2. For any z, $\Phi(z) = P(Z \le z) = P(Z \ge -z) = 1 - \Phi(-z)$
   - Symmetric property!
3. If $Z \sim N(0,1)$, then $-Z \sim N(0,1)$;
4. If $Z \sim N(0,1)$, then $\sigma Z + \mu \sim N(\mu, \sigma^2)$

Approximation to Binomial

Let $X \sim Bin(n, p)$, so that $E(X) = np$ and $V(X) = np(1-p)$. n → ∞:

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{np(1-p)}} is approximately \sim N(0,1)$$

Continuity Correction (IMPT)

$P(X = k) \approx P(k - 1/2 < X < k + 1/2)$
$P(a \le X \le b) \approx P(a - 1/2 < X < b + 1/2)$
$P(a < X \le b) \approx P(a + 1/2 < X < b + 1/2)$
$P(a \le X < b) \approx P(a - 1/2 < X < b - 1/2)$
$P(a < X < b) \approx P(a + 1/2 < X < b - 1/2)$
$P(X \le c) = P(0 \le X \le c) \approx P(-1/2 < X < c + 1/2)$
$P(X > c) = P(c < X \le n) \approx P(c + 1/2 < X < n + 1/2)$

Chapter 5: Sampling & Sampling Distributions

	Statistic	Realization
Sample Mean	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
	Statistic	**Realization**
Sample Variance	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Validity of $\bar{X}$ as an estimator for μ_X

- The expectation of $\bar{X}$ is equal to the population mean μ_X. $(E(\bar{X}) = \mu_X)$
- In the long run, $\bar{X}$ does not introduce any systematic bias as an estimator of μ_X.
  - Hence, $\bar{X}$ can serve as a valid estimator of μ_X.
- For an infinite population, when n gets larger and larger, $\frac{\sigma_X^2}{n}$, the variance of $\bar{X}$, becomes smaller and smaller, that is, the accuracy of $\bar{X}$ as an estimator of μ_X keeps improving. $\quad \mu_{\bar{X}} = E(\bar{X}) = \mu_X\ \&\ \sigma_{\bar{X}}^2 = V(\bar{X}) = \frac{\sigma_X^2}{n}$

Std Error

- Denoted by $\sigma_{\bar{X}}^2$.
- Measures the spread of the sampling distribution (s.d).
- The standard error of $\bar{X}$ describes how much $\bar{x}$ tends to vary from sample to sample of size n.
- As n increases, $\frac{\sigma_X^2}{n}$ decreases → $\bar{X}$ tend to be closer to μ_X as n increases.

Law of Large Nums

$$P(|\bar{X} - \mu| > \varepsilon) \to 0\ as\ n \to \infty$$

- As such, $\bar{X}$ converges to μ_X as n grows indefinitely → This is known as the Law of Large Numbers (LNN).
- As the sample size increases, the probability that the sample mean differs from the population mean goes to zero.
- It is increasingly likely that $\bar{X}$ is close to μ_X, as n gets larger

CLT

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \to Z \sim N(0,1)\ equivalently\ \bar{X} \to N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Large n → random samples follows the normal distribution.
- In the case where $X_1, X_2, \ldots, X_n$ are independent and identically distributed $N(\mu, \sigma^2)$, then:
  $$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad or \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Exactly, regardless of the sample size n. (For Normal Distribution)

Chi Square χ^2 Distribution

A χ^2 **random variable with n degree of freedom(df)** as $\chi^2(n)$.

Properties of χ^2 Distributions

1. If $Y \sim \chi^2(n)$, then $E(Y) = n$ and $V(Y) = 2n$.
2. For large n, $\chi^2(n)$ is approximately $N(n, 2n)$.
3. If Y_1 and Y_2 are independent χ^2 random variables with m and n degrees of freedom respectively, then $Y_1 + Y_2$ is a χ^2 random variable with $m + n$ degrees of freedom.
4. The χ^2 distribution is a family of curves, each determined by the degrees of freedom, n.
   - All the density functions have a **LONG RIGHT TAIL**.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- $E(S^2) = \sigma^2$
- If S^2 is the VAR of a random sample of size n taken from a normal population having the variance σ^2, then the random variable:
  $$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$$
  Has a χ^2 distribution with $n - 1$ degrees of freedom.

t-distribution

$$T = \frac{Z}{\sqrt{\frac{U}{n}}}$$

Follows the **t-distribution with n degrees of freedom.**

Properties of t-Distribution

1. The t-Distribution approached $N(0,1)$ as $n \to \infty$.
   - When $n \ge 30$, we can replace it (approximate) it to be $N(0,1)$.
2. If $T \sim t(n)$, then $E(T) = 0$ and $V(T) = \frac{n}{(n-2)}$ for $n > 2$.
3. The graph of the t-Distribution is symmetric about the vertical axis and resembles the graph of the standard normal distribution. (Graph $t(n)$ similar to graph $N(0,1)$)

** t-distribution appears as a result of random sampling

If $X_1, \ldots, X_n$ are independent and identically distributed normal random variables with mean μ and variance σ^2, then: $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n)$

Follows a t-Distribution with $n - 1$ degrees of freedom.

F-Distribution

$$F = \frac{U/m}{V/n}$$

Follows the **F-distribution with (m, n) degrees of freedom.**

Properties of t-Distribution

1. The F-distribution with (m, n) df is denoted by: $F(m, n)$.
2. If $X \sim F(m, n)$, then:
   $$E(X) = \frac{n}{n-2}, \text{ for } n > 2$$
   $$V(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \text{ for } n > 4$$
3. If $F \sim F(n, m)$, then $\frac{1}{F} \sim F(m, n)$.
   - This follows immediately from the def of the F-Distribution.
   - The values of interest are $F(m, n; \alpha)$ such that:
     $$P(F > F(m, n; \alpha)) = \alpha, \quad where\ F \sim F(m, n)$$
4. $F(m, n; 1 - \alpha) = \frac{1}{F}(n, m; \alpha)$

Chapter 6: Estimation

Let $X_1, X_2, \ldots, X_n$ be a random sample from the same population with mean μ and var σ^2. Then S^2 is an unbiased estimator of σ^2, since $E(S^2) = \sigma^2$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Define z_α to be the number with an upper-tail probability of α for the standard normal distribution Z. That is $P(Z > z_\alpha) = \alpha$.

$$P\left(-z_{\alpha/2} \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1 - \alpha$$

To get Max Error	To get min sample size
$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$n \ge \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$

Point Estimate: Different Cases (Refer to Examplify Formula Sheet)

Confidence Interval:

A confidence interval is a range of values that is likely to contain a population parameter based on a certain degree of confidence.

Population Proportion:	Population Mean
$CI = p^* \pm z^* \times \sqrt{\frac{p^*(1-p^*)}{n}}$	$CI = \bar{x} \pm t^* \times \frac{s}{\sqrt{n}}$
$z = \frac{x - \mu}{\sigma}$	$t = \frac{\bar{x} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$

Independent: (Refer to Examplify Formula Sheet)

Paired Data:

Assumptions for Paired Data:

1. $(X_1, Y_1), \ldots, (X_n, Y_n)$ are matched pairs, where $X_1, \ldots X_n$ is a random sample from population 1, $Y_1, \ldots, Y_n$ is a random sample from population 2.
2. X_i and Y_i are dependent.
3. (X_i, Y_i) and (X_j, Y_j) are independent for any $i \ne j$.
4. For matched pairs, define $D_i = X_i - Y_i$, $\mu_D = \mu_1 - \mu_2$.
5. Now we can treat $D_1, D_2, \ldots, D_n$ as a random sample from a SINGLE population with mean μ_D and variance σ_D^2.

We can employ all the techniques for single population to Paired Data.

	Where:	Where:
$T = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}}$	$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}$	$S_D^2 = \frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1}$

- If n < 30, and the population is normally distributed, then: $T \sim t_{n-1}$
- If $n \ge 30$ is large, then: $T \sim N(0,1)$

For paired data, if n is small & the pop is norm distributed, a $100(1-\alpha)\%$ CI for μ_D:

$$\bar{d} \pm t_{n-1;\alpha/2} \cdot \frac{s_D}{\sqrt{n}}$$

If n is large, a $100(1-\alpha)\%$ CI for μ_D $\bar{d} \pm z_{\alpha/2} \cdot \frac{s_D}{\sqrt{n}}$

Independent Sample vs Paired Data:

- **Independent samples** involve measurements from two completely independent groups
- **Paired Samples**: Paired samples involve the same individuals or units before and after a treatment, intervention, or simply over time.

Chapter 7: Hypothesis Testing

Step 1: Null Hypothesis vs Alternative Hypothesis

The outcome of hypothesis testing is either to REJECT or NOT REJECT H_0

Step 2: Level of Significance

For any test of hypothesis, there are only 2 possible conclusions:

1. Reject H_0 and therefore conclude H_1.
2. DO NOT Reject H_0 and therefore conclude H_0.

Whatever decision is made, there is always a possibility of making an error:

	DO NOT Reject H_0	Reject H_0
H_0 is TRUE	Correct Decision	Type 1 Error
H_0 is FALSE	Type II Error	Correct Decision

The rejection of H_0 when H_0 is TRUE is called a Type I error.

The probability of making a Type I error is called the **level of significance**, α. That is:
$$\alpha = P(Type\ I\ Error) = P(Reject\ H_0 | H_0\ is\ true)$$

Not rejecting H_0 when H_0 is FALSE is called a Type II error.

The probability of making a Type II error, denoted by β. That is:
$$\beta = P(Type\ II\ Error) = P(Do\ not\ reject\ H_0 | H_0\ is\ false)$$
The **power of the test** is defined by:
$$1 - \beta = P(Reject\ H_0 | H_0\ is\ false)$$

Remarks: Type 1 and Type 2 errors are dependent events.

Step 3: Test Statistics, Distribution and Rejection Region

As the significance level is given, a decision rule can be found such that it divides the set of all possible values of the test statistic into two regions, one being the **rejection region (or critical region)** and the other, the **acceptance region.**

Step 4 & 5: Calculation & Conclusion

- We check if the value is within our rejection region.
  - YES → sample **improbable assuming H_0** is true, hence we **reject** H_0.
  - NO → We **failed to reject H_0**

p-value for Hypothesis Testing

Suppose the computed test statistic was z.

2-tail	For a 2-sided test, a "worse" result would be if $\|Z\| > z$ or $Z < -\|z\|$. i.e $\|Z\| > \|z\|$. The p-value is given by: $p - value = P(\|Z\| > \|z\|) = 2P(\|Z\| > z) = 2P(Z < -\|z\|)$
1-tail	For the alternative hypothesis $H_1: \mu < \mu_0$, the p-value is $P(Z < -\|z\|)$, **L** For the alternative hypothesis $H_1: \mu > \mu_0$, the p-value is $P(Z > \|z\|)$, **R**

Hypotest: Known Variance

1-tail test: $H_0: \mu = \mu_0 \quad vs \quad H_1: \mu < \mu_0\ OR\ \mu > \mu_0$

- Reject H_0 when $\bar{X}$ is too large/small compared to μ_0.
  $$P(Z < -z_\alpha) = \alpha\ \ OR\ \ P(Z > z_\alpha) = \alpha$$

$H_1: \mu < \mu_0$ → $z < -z_\alpha$	$H_1: \mu > \mu_0$ → $z > z_\alpha$

2-tail test $H_0: \mu = c \quad vs \quad H_1: \mu \ne \mu_0$

- Reject H_0 when $\bar{X}$ is too large/small compared to μ_0.
  $$P(|Z| > z_{\alpha/2}) = \alpha$$
- Rejection region is defined by: $|Z| > z_{\alpha/2}$, which is:
  $$z < -z_{\alpha/2}\ \ OR\ \ z > z_{\alpha/2}$$

Hypotest: Unknown Variance

1-tail test: $H_0: \mu = \mu_0 \quad vs \quad H_1: \mu < \mu_0\ OR\ \mu > \mu_0$

- Reject H_0 when $\bar{X}$ is too large/small compared to μ_0.
  $$P(t < -t_{n-1;\alpha}) = \alpha\ \ OR\ \ P(t > t_{n-1;\alpha}) = \alpha$$

$H_1: \mu < \mu_0$ → $t < -t_{n-1;\alpha}$	$H_1: \mu > \mu_0$ → $t > t_{n-1;\alpha}$

2-tail test: $H_0: \mu = c \quad vs \quad H_1: \mu \ne \mu_0$

- Reject H_0 when $\bar{X}$ is too large/small compared to μ_0.
  $$P(|T| > t_{n-1;\alpha/2}) = \alpha$$
- Rejection region is defined by: $|Z| > z_{\alpha/2}$, which is:
  $$t < -t_{n-1;\alpha/2}\ \ OR\ \ t > t_{n-1;\alpha/2}$$

Test Comparing Means: Independent Samples (Refer to Examplify Formula Sheet)

Independent Samples: Rejection Regions & p-values

H_1	Rejection Region	p-value
$\mu_1 - \mu_2 > \delta_0$	$z > z_\alpha$	$P(Z > \|z\|)$
$\mu_1 - \mu_2 < \delta_0$	$z < -z_\alpha$	$P(Z < -\|z\|)$
$\mu_1 - \mu_2 \ne \delta_0$	$z > z_{\alpha/2}$ OR $z < -z_{\alpha/2}$	$2P(Z > \|z\|)$ OR $2P(Z < -\|z\|)$

Test Comparing Means: Paired Data

- For paired data, define $D_i = X_i - Y_i$.
- For the null hypothesis $H_0: \mu_D = \mu_{D_0}$, the test statistics is given by: $T = \frac{\bar{D} - \mu_{D_0}}{S_D/\sqrt{n}}$
- If n is small ($n < 30$) and the population if normally distributed, then: $T \sim t_{n-1}$
- If $n \ge 30$ is large, then: $T \sim N(0,1)$
- Equal variance applies when ½ ≤ $S1/S2$ ≤ 2