1 Definitions and Theorems

1.1 Combinatorial Analysis

Basic Principle of Counting: A: m outcomes, B: n outcomes, $\Rightarrow A \cap B$: mn outcomes.

- **Generalized**: A_i: i outcomes $\Rightarrow \prod n_i$ outcomes.

Permutations n distinct objs $\Rightarrow n!$ permutations.

- **Generalized**: $n = \sum n_i$, n_i distinct $\Rightarrow n!/(\prod n_i!)$ permutations.

Combinations n distinct choose $r \Rightarrow \binom{n}{r} = n!/r!(n-r)!$.

- $r < 0$ or $r > n \Rightarrow \binom{n}{r} = 0$;
- $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$.

Binomial Thm $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

- $\sum_{k=0}^{n} \binom{n}{k} = 2^n$;
- $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$, $n \geq 1$;
- $\binom{n}{0} + \binom{n}{2} + \cdots = \binom{n}{1} + \binom{n}{3} + \cdots$.

Multinomial Coef $n = \sum n_i$, n_i distinct, choose r groups $\Rightarrow \binom{n}{n_1, n_2, \cdots, n_r} = n!/(\prod n_i!)$.

- $\left(\sum_{i=1}^{r} x_i\right) = \sum_{n_1+n_2+\cdots+n_r=n} \binom{n}{n_1, n_2, \cdots, n_r} \prod_{i=1}^{r} x_i^{n_i}$

Integer Sols Solve $x_1 + x_2 + \cdots + x_r = n$, $(x_i > 0, \forall i)$

- $\binom{n-1}{r-1}$ positive int sols;
- $\binom{n+r-1}{r-1}$ non-negative int sols.

1.2 Axioms of Probability

Set Ops Laws:

- **Commutative** $EF = FE$, $E \cup F = F \cup E$;
- **Associative** $(EF)G = E(FG)$, $(E \cup F) \cup G = E \cup (F \cup G)$;
- **Distributive** $E(F \cup G) = EF \cup EG$, $E \cap (F \cup G) = (E \cap F) \cup (E \cap G)$;
- **DeMorgan** $(\bigcup_{i=1}^{n} E_i)^c = \bigcap_{i=1}^{n} E_i^c$, $(\bigcap_{i=1}^{n} E_i)^c = \bigcup_{i=1}^{n} E_i^c$.

Axioms of Probability

- $0 \leq P(E) \leq 1$;
- $S :=$ sample space $\Rightarrow P(S) = 1$;
- $E_1, E_2, \cdots$ disjoint $\Rightarrow P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$.

Properties of Probability

- $P(\emptyset) = 0$;
- $P(A) \leq P(B)$ if $A \subseteq B$;
- $P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(E^c) = 1 - P(E)$

Inclusion Exclusion

$$\begin{aligned} &P(E_1 \cup E_2 \cup \cdots \cup E_n) \\ =& \sum_{i=1}^{n} P(E_i) - \sum_{i \leq i_1 < i_2 \leq n} P(E_{i_1} \cap E_{i_2}) + \cdots + (-1)^{r+1} \\ &+ (-1)^{n-1} P(E_1 \cap E_2 \cap \cdots \cap E_n) \end{aligned}$$

$\nearrow$, $\searrow$ **Sequences** $\{E_n\}$ is **increasing** if $E_1 \subseteq E_2 \subseteq \cdots$, **decreasing** if $E_1 \supseteq E_2 \supseteq \cdots$.

- $\{E_n\} \nearrow \Rightarrow \lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$;
- $\{E_n\} \searrow \Rightarrow \lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$.

1.3 Cond Probability and Independence

Conditional Probability B happens given A happens:

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

- $P(AB) := P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$;
- $P(A_1 \cdots A_2) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1A_2 \cdots A_{n-1})$.

Bayes Formula: $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$.

- **Generalized**: A_i's **partition** $S \Rightarrow P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$.

Partition A_i partition S iff

- **Exclusive**: $A_i \cap A_j = \emptyset$;
- **Exhaustive**: $\bigcup_{i=1}^{\infty} A_i = S$.

Bayes' 2$^{\text{nd}}$ Formula: $A_i > 0$'s partition $S \Rightarrow$

$$P(A_j|B) = P(B|A_j)P(A_j) \bigg/ \left(\sum_{i=1}^{n} P(B|A_i)P(A_i)\right)$$

Odds of $A := P(A)/P(A^c) = P(A)/[1 - P(A)]$.

Independent A and B are **independent** if $P(AB) = P(A)P(B)$, **dependent** otherwise.

- A, B independent $\Rightarrow A$ and B^c, A^c and B, A^c and B^c independent;
- A independent of B, C **DOES NOT IMPLY** A independent of BC.

Independence of 3 events: satisfy the **ALL 4** conditions:

- $P(ABC) = P(A)P(B)P(C)$;
- $P(AC) = P(A)P(C)$;
- $P(AB) = P(A)P(B)$;
- $P(BC) = P(B)P(C)$.
- A, B, C independent $\Rightarrow A$ independent of any events formed by B and C.

Algebra of Cond Probability: let $P(A) > 0$, then

- $\forall B \Rightarrow 0 \leq P(B|A) \leq 1$;
- $P(S|A) = 1$;
- B_i's **mut exclusive** $\Rightarrow P(\bigcup_{i=1}^{\infty} B_i|A) = \sum_{i=1}^{\infty} P(B_i|A)$.

1.4 Random Variables

Discrete: **finite** or **countably infinite**

Probability Mass Function (p.m.f.) $p_X(x)$:

$$p_X(x) = \begin{cases} P(X = x), & \text{if } x = x_1, x_2, \cdots \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\sum_{i=1}^{\infty} p_X(x_i) = 1$.

Cumulative Distribution Function (c.d.f.) $F_X(x)$:

$$F_X(x) = P(X \leq x)$$

Expectation: X **discrete** with **p.m.f.** $p_X \Rightarrow$ **expectation**

$$E(X) := \sum_x x p_X(x).$$

Tail Sum Formula: $X \in \mathbb{Z}^+ \Rightarrow$

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k).$$

Expectation of a Function: X **discrete** with p_X, g:

$$E[g(X)] = \sum_i g(x_i)p_X(x_i) = \sum_x g(x)p_X(x).$$

$$E[aX + b] = aE(X) + b.$$

Variance Let $\mu := E(X)$, i.e. **mean** of X,

$$\mathrm{Var}(X) = E[(X - \mu)^2].$$

Zero Variance If $\mathrm{Var}(X) = 0$, then the random variable X is a constant. **Standard Deviation** σ_X or $\mathrm{SD}(X)$: $\sigma_X := \sqrt{\mathrm{Var}(X)}$.

Scaling and shifting:

- $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$;
- $\mathrm{SD}(aX + b) = |a| \mathrm{SD}(X)$.

Properties of d.f.:

- F_X **non-decreasing** $\Rightarrow F_X(a) \leq F_X(b)$;
- $\lim_{b\to\infty} F_X(b) = 1$, and $\lim_{b\to-\infty} F_X(b) = 0$;
- F_X has **left limits**, i.e. $\lim_{x\to b^-} F_X(x)$ exists for all $b \in \mathbb{R}$;
- F_X is **right continuous**, i.e. $\forall b \in \mathbb{R}$, $\lim_{x\to b^+} F_X(x) = F_X(b)$.

Useful Calculations:

1. P from **d.f.**:
   - $P(a < X \leq b) = F_X(b) - F_X(a)$;
   - $P(X = a) = F_X(a) - F_X(a^-)$, where $F_X(a^-) = \lim_{x\to a^-} F_X(x)$.
   - $\begin{aligned} P(a \leq X \leq b) &= P(X = a) + P(a < X \leq b) \\ &= F_X(b) - F_X(a^-). \end{aligned}$
2. P from **PMF**: $P(A) = \sum_{x\in A} p_X(x)$;
3. **p.m.f.** from **d.f.**: $p_X(x) = F_X(x) - F_X(x^-)$;
4. **d.f.** from **p.m.f.**: $F_X(x) = \sum_{y\leq x} p_X(y)$.

1.5 Continuous Random Variables

Continuous Variable & p.d.f X is **continuous** if $\exists$ nonnegative f_X s.t. $\forall x \in \mathbb{R}$, $P(a < X \leq b) = \int_a^b f_X(x)dx$, for $-\infty < a < b < +\infty$.

- $F_X(x) = \int_{-\infty}^{x} f_X(t)\ \mathrm{d}t$;
- $f_X(x) = \dfrac{\partial d}{\partial dx} F_X(x)$.
- $P(X = x) = 0$;
- **d.f.** $F_X(x)$ is **continuous**.
- $\forall a, b \in (-\infty, \infty)$, $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$.

Determing constant: $\int_{-\infty}^{\infty} f_X(x)\ \mathrm{d}x = 1$.

Expectation: $E(X) = \int_{-\infty}^{\infty} x f_X(x)\ \mathrm{d}x$,.

Variance: $\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x)\ \mathrm{d}x$.

Functions of Expectations:

- $E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)\mathrm{d}x$
- $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$
- $E(aX + b) = aE(X) + b$

Tail Sum Formula given X **nonnegative** $\Rightarrow$

$$E(X) = \int_0^{\infty} P(X > x)\ \mathrm{d}x.$$

Y **monotonic, differentiable on** X Let $g(x)$ be **strictly monotonic, differentiable** function of X. $\Rightarrow$ The **p.d.f.** of $Y := g(X)$:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \dfrac{d}{dy} g^{-1}(y) \right|, & \text{if } y \in g(\mathcal{R}_X) \\ 0, & \text{otherwise.} \end{cases}$$

where g^{-1} is the inverse function of g.

1.6 Jointly Distributed Random Variables

Joint d.f.: X, Y defined on **SAME** sample space S, **joint distribution function** of X and Y, denoted by $F_{XY}(x,y)$,is

$$F_{XY}(x,y) = P(X \leq x, Y \leq y), \quad \text{for } x, y \in \mathbb{R},$$

where $\{X \leq x, Y \leq y\} := \{X \leq x\} \cap \{Y \leq y\}$. **Marginal Distribution Function**: (of X and Y)

$$F_X(x) = P(X \leq x) = P(X \leq x, Y < \infty) = F_{XY}(x, \infty)$$

$$F_Y(y) = P(Y \leq y) = P(X < \infty, Y \leq y) = F_{XY}(\infty, y)$$

Useful Calculations

- $P(X > a, Y > b) = 1 - F_X(a) - F_Y(b) + F_{XY}(a,b)$;
- $\begin{aligned} &P(a_1 < X \leq a_2, b_1 < Y \leq b_2) \\ =& F_{XY}(a_2,b_2) - F_{XY}(a_1,b_2) - F_{XY}(a_2,b_1) + F_{XY}(a_1,b_1). \end{aligned}$

Joint p.m.f. $p_{X,Y}(x,y) = P(X = x, Y = y)$, for $x, y \in \mathbb{R}$.

marginal p.m.f (of X and Y respectively)

- $p_X(x) = \sum_y p_{X,Y}(x,y)$;
- $p_Y(y) = \sum_x p_{X,Y}(x,y)$.

Some useful formulas

- $P(a_q < X \leq a_2, b_1 < Y \leq b_2) = \sum_{a_1 < x \leq a_2} \sum_{b_1 < y \leq b_2} p_{X,Y}(x,y)$;
- $F_{X,Y}(a,b) = P(X \leq a, Y \leq b) = \sum_{x\leq a} \sum_{y\leq b} p_{(X,Y)}(x,y)$;
- $P(X > a, Y > b) = \sum_{x>a} \sum_{y>b} p_{(X,Y)}(x,y)$.

Jointly Continuous Random Variables X, Y are **JCRM** if $\forall x, y \in \mathbb{R}$, $\exists$ nonnegative **joint p.m.f.** $f_{X,Y}(x,y)$ of X and Y, s.t. $\forall C \subset \mathbb{R}^2$,

$$P((X,Y) \in C) := \int\int_{(x,y)\in C} f_{X,Y}(x,y)\ \mathrm{d}x\mathrm{d}y.$$

Marginal p.d.f. of X **and** Y:

- $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ \mathrm{d}y$;
- $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ \mathrm{d}x$.

Some useful formulas

- Let $A, B \subset \mathbb{R}$, take $C = A \times B$:

$$P(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x,y)\ \mathrm{d}y\mathrm{d}x.$$

- Let $a_1, a_2, b_1, b_2 \in \mathbb{R}$ where $a_1 < a_2$ and $b_1 < b_2$:

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x,y)\ \mathrm{d}y\mathrm{d}x.$$

- Let $a, b \in \mathbb{R}$:

$$F_{X,Y}(a,b) = P(X \leq a, Y \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x,y)\ \mathrm{d}y\mathrm{d}x.$$

- $f_{X,Y}(x,y) \dfrac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$.

Independent X, Y are **independent** if, for any $A, B \subset \mathbb{R}$, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$, or equivalently $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, or equivalently iff $(\forall x, y)[(\exists g, h : \mathbb{R} \to \mathbb{R}) \Rightarrow f_{X,Y}(x,y) = g(x)h(y)]$.

Convolution X, Y **continuous and independent**, hence $f_{X,Y}(x,y) = f_X(x)f_Y(y) \Rightarrow$ **convolution** of f_X, f_Y:

$$\begin{aligned} F_{X+Y}(a) &= P(X + Y \leq a) \\ &= \int_{-\infty}^{\infty} F_X(a - y)f_Y(y)\ \mathrm{d}y \\ &= \int_{-\infty}^{\infty} F_Y(a - x)f_X(x)\ \mathrm{d}x \end{aligned}$$

Conditional P.M/D.F. of X given $Y = y$:

- **D**: $p_{X|Y}(x,y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)}$;
- **C**: $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$,.

for all values y such that $f_Y(y) > 0$ ($p_Y(y) > 0$).

Conditional D.F. of X given $Y = y$:

- **D**: $F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{a\leq x} p_{X|Y}(a|y)$.
- **C**: $F_{(X|Y)}(x|y) = P(X \leq x|Y = y) = \int_{-\infty}^{x} f_{X|Y}(t|y)\ \mathrm{d}t$.

Independent $\Rightarrow$ **Cond = Marginal p.m.f.** X, Y **independent** $\Rightarrow$ the following two are **SAME**:

- **conditional p.m.f.** of X given $Y = y$;
- **marginal p.m.f.** of X for every Y such that $p_Y(y) > 0$.

Joint p.d.f. of functions Following conditions are satisfied:

1. Let X and Y be **jointly continuously** distribtued random variables with known joint density function.
2. Let U and V be given functions of X and Y in the form

   $$U = g(X,Y), \quad V = h(X,Y).$$

   And we can **uniquely** solvve for X and Y in terms of U and V, i.e. say $x = a(u,v)$ and $y = b(u,v)$.
3. The functions g and h have **continuous partial derivatives** at all points (x,y) and

   $$J(x,y) := \begin{vmatrix} \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \\ \dfrac{\partial h}{\partial x} & \dfrac{\partial h}{\partial y} \end{vmatrix} = \frac{\partial g}{\partial x}\frac{\partial h}{\partial y} - \frac{\partial g}{\partial y}\frac{\partial h}{\partial x} \neq 0$$

   for all (x,y).

Then, the **joint probability density function** of U and V is given by

$$f_{U,V}(u,v) = f_{X,Y}(x,y)|J(x,y)|^{-1}$$

where $x = a(u,v)$ and $y = b(u,v)$, (check point 2).

Joint p.d.f. Generalized joint p.d.f. of the n random variables $X_1, X_2, \cdots, X_n$ given, want to compute the joint density function of $Y_1, Y_2, \cdots, Y_n$ where $Y_i = g_i(X_1, X_2, \cdots, X_n)$. Assume that the function g_j have **continuous partial derivatives** and the **Jacobian determinant** $J(x_1, x_2, \cdots, x_n) \neq 0$ at **all points** $(x_1, x_2, \cdots, x_n)$, where

$$J(x_1, x_2, \cdots, x_n) = \begin{vmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} & \cdots & \dfrac{\partial g_1}{\partial x_n} \\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2} & \cdots & \dfrac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial g_n}{\partial x_1} & \dfrac{\partial g_n}{\partial x_2} & \cdots & \dfrac{\partial g_n}{\partial x_n} \end{vmatrix}$$

Further, we suppose that the equations $y_i = g_i(x_1, x_2, \cdots, x_n)$ have a **unique** solution for $x_1, x_2, \cdots, x_n$, say $x_i = h_i(y_1, y_2, \cdots, y_n)$. Then, the **joint density function** of $Y_1, Y_2, \cdots, Y_n$ is given by.

$$\begin{aligned} &f_{Y_1,Y_2,\cdots,Y_n}(y_1, y_2, \cdots, y_n) \\ =& f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n)|J(x_1, x_2, \cdots, x_n)|^{-1} \end{aligned}$$

where $x_i = h_i(y_1, y_2, \cdots, y_n)$

Joint of 3: $F_{X,Y,Z}(x,y,z) = P(X \leq x, Y \leq y, Z \leq z)$.

Equivalent Cond for Joint of 3:

- Random variables X, Y, Z are **independent**.
- For all $x, y, z \in \mathbb{R}$, we have $f_{X,Y,Z}(x,y,z) = f_X(x)f_Y(y)f_Z(z)$.
- For all $x, y, z \in \mathbb{R}$, we have $F_{X,Y,Z}(x,y,z) = F_X(x)F_Y(y)F_Z(z)$.

Independence Random variables X, Y, and Z are independent if and only if there exist functions $g_1, g_2, g_3 : \mathbb{R} \to \mathbb{R}$ such that for all $x, y, z \in \mathbb{R}$, we have

$$f_{X,Y,Z}(x,y,z) = g_1(x)g_2(y)g_3(z).$$

1.7 Properties of Expectation

Expectation of Func of Joint:

- **D**: $E[g(X,Y)] = \sum_y \sum_x g(x,y)p_{X,Y}(x,y)$;
- **C**: $E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)\ \mathrm{d}x\mathrm{d}y$.

Properties:

- If $a \leq X \leq b$, then $a \leq E(X) \leq b$.
- If $g(x,y) \geq 0$ whenever $p_{X,Y}(x,y) > 0$, then $E[g(X,Y)] \geq 0$.
- $E[g(X,Y) + h(X,Y)] = E[g(X,Y)] + E[h(X,Y)]$.
- $E[g(X) + h(Y)] = E[g(X)] + E[h(Y)]$.
- **Monotone property**: if jointly distributed random variables X and Y satisfy $X \leq Y$, then $E[X] \leq E[Y]$.
- **Special case**: mean of sum is sum of means: $E[X + Y] = E[X] + E[Y]$.

Covariance: let $\mu_X = E(X)$ and $\mu_Y = E(Y)$, then

$$\mathrm{Cov}(X,Y) := E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y).$$

Properties of Covariance:

- $\mathrm{Cov}(X,Y) \neq 0 \Rightarrow X$ and Y are **correlated**;
- $\mathrm{Cov}(X,Y) = 0 \Rightarrow X$ and Y are **uncorrelated**.
- $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$
- $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
- $\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathrm{Cov}(X_i, Y_j)$.

Product of Independent: If X and Y are **independent**, then for any functions $g, h : \mathbb{R} \to \mathbb{R}$, we have

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

Covariance of Independent If X and Y are **independent**, then $\mathrm{Cov}(X,Y) = 0$.

Variance of a Sum

$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}\,(X_k) + 2\sum_{1\le i<j\le n} \mathrm{Cov}\,(X_i, X_j)$$

When $X_1, X_2, \ldots, X_n$ are **independent**, we have

$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}\,(X_k)\,.$$

Correlation Coefficient: $\rho_{X,Y} = \dfrac{\mathrm{Cov}\,(X,Y)}{\sqrt{\mathrm{Var}\,(X)\,\mathrm{Var}\,(Y)}}$.

(Sometimes denoted by Corr)

- $-1 \le \rho_{X,Y} \le 1$;
- $\rho\,(X,Y) = \pm 1 \Rightarrow y = ax + b$.

Conditional Expectation:

- **D**: $E\,[X \mid Y = y] = \sum_x x \cdot p_{X|Y}\,(x \mid y)$;
- **C**: $E\,[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}\,(x \mid y)\,\mathrm{d}x$.

Conditional Expectation of a Function:

- **D**: $E[g(X)|Y = y] = \sum_x g(x) p_{X|Y}(x|y)$;
- **C**: $E[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\,\mathrm{d}x$.

Therefore, $E\left[\sum_{k=1}^{n} X_k | Y = y\right] = \sum_{k=1}^{n} E[X_k|Y = y]$.

Law of Total Variance:

$$\mathrm{Var}\,(X) = E\,[\mathrm{Var}\,(X \mid Y)] + \mathrm{Var}\,(E\,[X \mid Y])\,.$$

Moment Generating Function (MGF):

- **D**: $M_X(t) := E[e^{tX}] = \sum_x e^{tx} p_X(x)$;
- **C**: $M_X(t) := E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,\mathrm{d}x$.

Multiplicative Property of MGF X, Y are **independent**:

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Uniqueness Property of MGF Suppose that there exists an $h < 0$ such that $M_X(t) = M_Y(t), \quad \forall t \in (-h, h)$. Then X and Y have the same distribution, i.e. $F_X = F_Y$ and $f_X = f_Y$.

Joint Moment Generating Function:

$$M_{X_1,\cdots,X_n}(t_1, \cdots, t_n) = E\,[\exp(t_1 X_1 + \cdots + t_n X_n)]\,.$$

Recover Individual MGF from Joint:

$$M_{X_i}(t_i) = M_{X_1,\cdots,X_n}(0, \cdots, t_i, \cdots, 0).$$

Independence of Mean and Variance from Normal Sample Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed normal random variables with mean μ and variance σ^2. Then the sample mean $\overline{X}$ and the sample variance S^2 are independent, and $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$ and $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.

1.8 Limit Theorems

Markov's Inequality Let X be a **nonnegative** random variable. For $a > 0$, we have

$$P(X \ge a) \le \frac{E[X]}{a}.$$

Chebyshev's Inequality Let X be a random variable with mean μ and variance σ^2. For $a > 0$, we have

$$P(|X - \mu| \ge a) \le \frac{\sigma^2}{a^2}.$$

The Weak Law of Large Numbers Let $X_1, X_2, \cdots$ be a sequence of **independent** and **identically** distributed random variables with common mean μ. Then, for any $\epsilon > 0$, we have

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \ge \epsilon\right) = 0.$$

Central Limit theorem Let $X_1, X_2, \cdots$ be a sequence of **independently** and **identically** distributed random variables, each having mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard norm as $n \to \infty$. That is

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \int_x^{-\infty} \exp(-t^2/2)\;\mathrm{d}t$$

The Strong Law of Large Numbers Let $X_1, X_2, \cdots$ be a sequence of **independent** and **identically distributed** random variables with common mean μ. Then, with probability 1, we have

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu.$$

One-sided Checbychev's Inequality Let X be a random variable with mean 0 and finite variance σ^2. Then, for $a > 0$, we have

$$P(X \ge a) \le \frac{\sigma^2}{\sigma^2 + a^2}.$$

Jensen's Inequality If $g(x)$ is a **convex** function, then

$$g(E[X]) \le E[g(X)]$$

provided that expectations exist and are finite.

Convex Functions A function is convex if either of the following equivalent conditions hold:

1. for all $0 \le p \le 1$ and for all $x_1, x_2 \in R_X$,

$$g(px_1 + (1-p)x_2) \le pg(x_1) + (1-p)g(x_2).$$

2. differentiable function: convex of interval if and only if

$$g(x) \ge g(y) + g'(y)(x - y)$$

   for all x, y in the interval.

3. A twice differentiable function is convex if and only if its second derivative is nonnegative.

2 Distributions

2.1 Bernoulli $\mathrm{Be}\,(p)$

$$X = \begin{cases} 1, & \text{if it is a success;} \\ 0 & \text{if it is a failure.} \end{cases}$$

- $P(X = 1) = p$;
- $P(X = 0) = 1 - p$;
- $E(X) = p$;
- $\mathrm{Var}\,(X) = p(1-p)$.

2.2 Binomial $\mathrm{Bin}\,(n,p)$

X as **number of successes** in n $\mathrm{Bernoulli}\,(p)$ trials.

- $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$;
- $E(X) = np$;
- $\mathrm{Var}\,(X) = np(1-p)$;

Sum of Binomial (same prob) $X \sim \mathrm{Bin}\,(n,p)$, $Y \sim \mathrm{Bin}\,(m,p)$, where X, Y are independent. **p.m.f.** of $X + Y$:

$$P(X + Y = k) = \binom{n+m}{k} p^k (1-p)^{n+m-k}$$

Normal Approximation: De Moivre-Laplace Theorem: for n satisfying n satisfying $npq \ge 10$:

$$\mathrm{Bin}\,(n,p) \approx \mathrm{N}\,(np, npq)\,.$$

Equivalently, $\dfrac{X - np}{\sqrt{npq}} \approx Z$ where $Z \sim \mathrm{N}\,(0,1)$.

Continuity Correction (Normal) If $X \sim \mathrm{Bin}\,(n,p)$, then

$$P(X = k) = P(k - 0.5 < X < k + 0.5)$$
$$P(X \ge k) = P(X \ge k - 0.5)$$
$$P(X \le k) = P(X \le k + 0.5)$$

Poisson Approximation to Binomial Works when n is large and p is small. **As a working rule,** $p < 0.1$ **and set** $\lambda = np$.

2.3 Geometric $\mathrm{Geom}\,(p)$

X as **number of trials until the first success** in a sequence of independent $\mathrm{Bernoulli}\,(p)$ trials.

- $P(X = k) = pq^{k-1}$;
- $E(X) = \dfrac{1}{p}$;
- $\mathrm{Var}\,(X) = \dfrac{1-p}{p^2}$;

Alternative Geometric Distribution: Let X' be the **number of failures** in the $\mathrm{Bernoulli}\,(p)$ trials in order to obtain the first success. Here, $X = X' + 1$.

- $P(X' = k) = pq^k$;
- $E(X') = \dfrac{1-p}{p}$;
- $\mathrm{Var}\,(X') = \dfrac{1-p}{p^2}$;

Negative Binomial Random Variable $\mathrm{NB}\,(r,p)$: define X to be **number of** $\mathrm{Bernoulli}\,(p)$ **trials** required to obtain r **successes**. For $k \ge r$, we have:

- $P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$;
- $E(X) = \dfrac{r}{p}$;
- $\mathrm{Var}\,(X) = \dfrac{r(1-p)}{p^2}$;

Remark: $\mathrm{Geom}\,(p) = \mathrm{NB}\,(1,p)\,.$

2.4 Poisson $\mathrm{Poisson}\,(\lambda)$

Poisson: $P(X = k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$, for $k = 0, 1, 2\cdots$.

- $E(X) = \lambda$;
- $\mathrm{Var}\,(X) = \lambda$.

Sum of Possion If $X \equiv \mathrm{Poisson}\,(\lambda)$, $Y \sim \mathrm{Poisson}\,(\mu)$, and X and Y are **independent**, then the **p.m.f.** of $X + Y$:

$$P(X + Y = n) = \sum_{k=0}^{n} P(X = k, Y = n - k)$$
$$= \frac{\exp[-(\lambda + \mu)](\lambda + \mu)^n}{n!}$$

The sum of two independent Poisson random variables is **still Poisson**, the mean is the **sum of the means**.

2.5 Hypergeometric $\mathrm{Hypergeom}\,(N, m, n)$

$$P(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

for $x = 0, 1, \cdots, \min(m, n)$.

- $\mathrm{E}\,(X) = \dfrac{nm}{N}$;
- $\mathrm{Var}\,(X) = \dfrac{nm}{N}\left[\dfrac{(n-1)(m-1)}{N-1} + 1 - \dfrac{nm}{N}\right]$.

Example: Suppose that we have N balls, of which m are red and $N - m$ are blue. We choose n of these balls, **without replacement**, and define X to be the number of red balls chosen. Then X is a **hypergeometric** random variable, with $P(X = x) = \binom{m}{x}\binom{N-m}{n-x}/\binom{N}{n}$.

2.6 Uniform Distribution

A continuous random variable X is said to have a **uniform distribution** on the interval (a, b) if its p.d.f. is given by

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise.} \end{cases}$$

Then,

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

We have:

- $E(X) = \dfrac{a+b}{2}$;
- $\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$.

2.7 Normal Distribution $N(\mu, \sigma^2)$

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

Standard Normal: $f_Z(z) = \dfrac{1}{\sqrt{2\pi}} e^{-z^2/2}$.

If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} \sim N(0,1)$, and

$$P(a < Y \le b) = P\left(\frac{a-\mu}{\sigma} < Z \le \frac{b-\mu}{\sigma}\right)$$
$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Properties of Normal Distribution

- $P(Z \ge 0) = P(Z \le 0) = \dfrac{1}{2}$;
- $-Z \sim N(0,1)$;
- $P(Z \le x) = 1 - P(Z > x)$ for $-\infty < x < \infty$;
- $P(Z \le -x) = P(Z \ge x)$ for $-\infty < x < \infty$;
- if $Y \sim N(\mu, \sigma^2)$, then $X := (Y - \mu)/\sigma \sim N(0,1)$;
- if $X \sim N(0,1)$, then $Y := aX + b \sim N(b, a^2)$.

Expectation and Variance

- if $Y \sim N(\mu, \sigma^2)$, then $E(Y) = \mu$ and $\mathrm{Var}(Y) = \sigma^2$;
- if $Z \sim N(0,1)$, then $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$.

Sum of Normal If $X_i, i = 1, \cdots, n$ are independent random variables that are normally distributed with respective parameters μ_i, σ_i^2, $i = 1, \cdots, n$ then

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right).$$

2.8 Expontential Distribution: $\mathrm{Exp}\,(\lambda)$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

- **Memoryless Property**:

$$P(X > s + t | X > s) = P(X > t), \quad s, t > 0.$$

- $E(X) = \dfrac{1}{\lambda}$;
- $\mathrm{Var}(X) = \dfrac{1}{\lambda^2}$.

2.9 Other Distributions

Gamma $\mathrm{Gamma}\,(\alpha, \lambda)$:

$$f_X(x) = \begin{cases} \dfrac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & x > 0 \\ 0, & x \le 0, \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$.

$$F_X(x) = \begin{cases} \dfrac{\gamma(\alpha, \lambda x)}{\Gamma(\alpha)}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

- $E(X) = \dfrac{\alpha}{\lambda}$;
- $\mathrm{Var}(X) = \dfrac{\alpha}{\lambda^2}$.
- $\Gamma(1) = \int_0^\infty e^{-y}\;\mathrm{d}y = 1$.
- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
- $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$.

Sum of Gamma with Same λ: $X \sim \mathrm{Gamma}\,(\alpha, \lambda)$, $Y \sim \mathrm{Gamma}\,(\lambda)$, $\Rightarrow X + Y \sim \mathrm{Gamma}\,(\alpha + \beta, \lambda)\,.$

Weibull $W(v, \alpha, \beta)$: $\alpha > 0$ and $\lambda > 0$, if its **probability density function** is given by $f_X(x) =$

$$\begin{cases} \dfrac{\beta}{\alpha}\left(\dfrac{x-v}{\alpha}\right)^{\beta-1} \exp\left[-\left(\dfrac{x-v}{\alpha}\right)^\beta\right], & x > v \\ 0, & x \le v, \end{cases}$$

- $\mathrm{E}(X) = v + \alpha\Gamma\left(1 + \dfrac{1}{\beta}\right)$;
- $\mathrm{Var}(X) = \alpha^2\left[\Gamma\left(1 + \dfrac{2}{\beta}\right) - \left(\Gamma\left(1 + \dfrac{1}{\beta}\right)\right)^2\right]$.

Cauchy $C(\alpha, \theta)$:

$$f_X(x) = \frac{1}{\pi\alpha\left[1 + \left(\dfrac{x-\theta}{\alpha}\right)^2\right]},$$

- $E(X)$ does not exist;
- $\mathrm{Var}(X)$ does not exist.

Beta $\mathrm{Beta}\,(\alpha, \beta)$: $\alpha > 0$ and $\beta > 0$

$$f_X(x) = \begin{cases} \dfrac{1}{B(a,b)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

- $E(X) = \dfrac{\alpha}{\alpha + \beta}$;
- $\mathrm{Var}(X) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Beta Function

$$\mathrm{B}\,(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 (1-u)^{\alpha-1} u^{\beta-1}\mathrm{d}u.$$

χ^2**-Distribution** $\chi^2(n)$: Let $X_1, X_2, \ldots, X_n$ be independent standard normal random variables. $Q = \sum_{i=1}^n X_i^2$ has a **chi-squared distribution** with n degrees of freedom:

$$f_Q(q) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} q^{n/2-1} e^{-q/2}, & q > 0 \\ 0, & q \le 0. \end{cases}$$

$$F_Q(q) = \begin{cases} \dfrac{\gamma(n/2, q/2)}{\Gamma(n/2)}, & q > 0 \\ 0, & q \le 0, \end{cases}$$

where $\gamma = \int_0^{q/2} t^{n/2-1} e^{-t}\,dt$ and $\Gamma = \int_0^\infty t^{n/2-1} e^{-t}\,dt$.

- $E(Q) = n$;
- $\mathrm{Var}(Q) = 2n$.

Log-Normal Distribution $\mathrm{LN}(\mu, \sigma)$ Let Z be a standard normal random variable. Then, the random variable $X = e^{\mu + \sigma Z}$ has a **log-normal distribution**.

$$f_X(x) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\ln x - \mu)^2\right], & x > 0 \\ 0, & x \le 0. \end{cases}$$

$$F_X(x) = \begin{cases} \dfrac{1}{2} + \dfrac{1}{2}\,\mathrm{erf}\left(\dfrac{\ln x - \mu}{\sigma\sqrt{2}}\right), & x > 0 \\ 0, & x \le 0. \end{cases}$$

where $\mathrm{erf}(z) = \dfrac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt$ is the **error function**.

Bivariate Normal Distribution Random variables X, Y have a **bivariate normal distribution** if, for constants $\mu_x, \mu_y, \sigma_x, \sigma_y > 0$, $-1 < \rho < 1$, there **joint density function** is given, for all $-\infty < x, y < \infty$, by

$$f_{X,Y}(x,y) := \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times e^{\text{exponent}}$$

where exponent is:

$$-\frac{\left[\left(\dfrac{x-\mu_x}{\sigma_x}\right)^2 + \left(\dfrac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\dfrac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}.\right]}{2(1-p^2)}.$$