- **Mutually Exclusive**: $A \cap B = \emptyset$
- **Union**: $A \cup B = \{x : x \in A \vee x \in B\}$
- **Intersection**: $A \cap B = \{x : x \in A \wedge x \in B\}$
- **Complement**: $A' = \{x : x \in S \wedge x \notin A\}$
- $(A \cap B)' = (A' \cup B')$
- **Multiplication**: R experiments performed sequentially. Then $n_i \cdot ... \cdot n_r$ possible outcomes for r experiments
- **Addition**: e can be performed k ways, and k ways do not overlap : total ways: $n_1 + ... + n_k$
- **Permutation**: Arrangement of r objects out of n, *ordered*. $P_r^n = \frac{n!}{(n-r)!}, P_n^n = n!$
- **Combination**: Selection of r objects out of n, *unordered* $\binom{n}{r} = \frac{n!}{r!(n-r)!}, \binom{n}{r} \times P_r^r = P_r^n$

Probability

- Axioms:
  1. $0 \leq P(A) \leq 1$
  2. $P(S) = 1$
- Propositions:
  1. $P(\emptyset) = 0$
  2. $A_1 ... A_n$ are mutually exclusive,$P(A_1 \cup ... \cup A_n) = P(A_1) + ... + P(A_n)$
  3. $P(A') = 1 - P(A)$
  4. $P(A) = P(A \cap B) + P(A \cap B')$
  5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  6. If $A \subset B, P(A) \leq P(B)$

Conditional Probability

- $P(B|A)$ is probability of B given that A has occured
- $P(B|A) = \frac{P(A \cap B)}{P(A)}$
- $P(A \cap B) = P(B|A)P(A)$
- $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$
- $P(A \cap B \cap C) = P(A)P(B|A)P(C|B \cap A)$
- **Independent**: $P(A \cap B) = P(A)P(B), A \perp B$
  - If $P(A) \neq 0, A \perp B \leftrightarrow P(B|A) = P(B)$ (Knowledge of A does not change B)
- **Independence vs Mutually exclusive**
  - $P(A) > 0 \wedge P(B) > 0, A \perp B \Rightarrow$ not mutually exclusive
- Partition: $A_i ... A_n$ is mutually exclusive and $\bigcup_i^n = 1 A_i = S, A_i ... A_n$ is partition of S
  - $P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i)P(B|A_i)$
  - $n = 2, P(B) = P(A)P(B|A) + P(A')P(B|A')$
- **Bayes Theorem**: $P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$
- $n = 2, P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A')P(B|A')}$

Random Variables

- Notations:
  - $\{X = x\} = \{s \in S : X(s) = x\} \in S$
  - $\{X \in A\} = \{s \in S : X(s) \in A\} \in S$

Probability Distributions

- PMF(*Discrete*) of $X - f(x) = \begin{cases} P(X=x) \text{ if } x \in R_X \\ 0 \text{ otherwise} \end{cases}$
- Properties (**must** satisfy)
  1. $f(x_i) \geq 0, x_i \in R_X$
  2. $f(x_i) = 0, x_i \notin R_X$
  3. $\sum_{i=1}^{\infty} f(x_i) = 1$
- PDF(*Continuous*) of X is function that satisfies the following
  1. $f(x) \geq 0, x \in R_X$ and $f(x) = 0, x \notin R_X$
  2. $\int_{R_X} f(x) \, \mathrm{d}x = 1$
  3. $a \leq b, P(a \leq X \leq b) = \int_a^b f(x) \, \mathrm{d}x$
  - To validate, check (1) and (2)
- CDF (Discrete): $F(X) = P(X \leq x)$
  - $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a-), a-$ (is largest value in R_X smaller than a)
- CDF(Continuous): $F(X) = \int_{-\infty}^x f(t) \, \mathrm{d}t$
  - $f(x) = \frac{\mathrm{d}}{\mathrm{d}t} F(x)$
  - $P(a \leq X \leq b) = F(b) - F(a)$
  - $F(x)$ is non-decreasing.
  - PDF/PMF and CDF have 1 to 1 correspondence.
  - Ranges of $F(x)$ and $f(x)$ satisfy:
    - $0 \leq F(X) \leq 1$
    - for discrete: $0 \leq f(X) \leq 1$
    - for continuous: $f(x) \geq 0$, not necessary $f(x) \leq 1$

Expectation

- Expectation(Discrete):

$$E(X) = \mu_X = \sum_{x_i \in R_X} x_i f(x_i)$$

- Expectation(Continuous):

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x_i f(x_i)$$

- **Properties**
  1. $E(aX + b) = aE(X) + b$
  2. $E(X + Y) = E(X) + E(Y)$
  3. Let $g(\cdot)$ be arbitrary function.
     - $$E[g(X)] = \sum g(x)f(x)$$ or $$E[g(X)] = \int_{R_X} g(x)f(x)$$
     - example: $E(X^2) = \sum x^2 f(x)$
- Variance:

$$\sigma_x^2 = V(X) = E(X - \mu)^2 = E(X^2) - E(X)^2$$

  - Discrete: $V(X) = \sum (x - \mu_x)^2 f(x)$
  - Continuous: $V(X) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x)$
  - **Properties**
    1. $V(aX + b) = a^2 V(X)$
    2. $V(X) = E(X^2) - E(X)^2$
    3. Standard Deviation = $\sigma_x = \sqrt{V(X)}$

Joint Probability Function

- Discrete

$$f_{X,Y}(x, y) = P(X = x, Y = y)$$

- **Properties**
  1. $f(X, Y)(x, y) \geq 0, (x, y) \in R_{X,Y}$
  2. $f(X, Y)(x, y) = 0, (x, y) \notin R_{X,Y}$
  3. $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (x_i, y_i) = 1$
- Continuous

$$P((X, Y) \in D) = \iint_{(x,y) \in D} f(x, y) \, \mathrm{d}y \, \mathrm{d}x$$

  - $P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) \, \mathrm{d}y \, \mathrm{d}x$
- **Properties**
  1. $f_{X,Y}(x, y) \geq 0$, for any $(x, y) \in R_{X,Y}$
  2. $f_{X,Y}(x, y) = 0$, for any $(x, y) \notin R_{X,Y}$
  3. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, \mathrm{d}x \, \mathrm{d}y = 1$

Marginal Probability Distribution

- Discrete: $f_X(x) = \sum_y f_{X,Y}(x, y)$
- Continuous: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, \mathrm{d}y$
- Conditional Distribution:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

  - If $f_X(x) > 0, f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y|x)$
  - $P(Y \leq y \mid X = x) = \int_{-\infty}^y f_{Y|X}(y|x) \, \mathrm{d}y$
  - $E(Y \mid X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) \, \mathrm{d}y$

Independent Random Variable

- **Independent**:

$$f_{X,Y}(x, y) = f_{X(x)} f_{Y(y)}$$

- **Properties**
  1. If X, Y are independent random variables, $$P(X \leq x; Y \leq y) = P(X \leq x)P(Y \leq y)$$
  2. $g_1(X)$ and $g_2(Y)$ are independent. (E.g. X^2 and $\log(Y)$ are independent)
  3. if $F_{X(x)} > 0$, then $f_{Y|X}(y|x) = f_Y(y)$

Expectation and Covariance

- **Expectation**:

$$E(g(X,Y)) = \sum_x \sum_y g(x,y) f_{X,Y}(x,y)$$

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) \,\mathrm{d}y \,\mathrm{d}x$$

- **Covariance**:

$$\operatorname{cov}(X,Y) = \sum_x \sum_y (x - \mu_x)(y - \mu_y) f_{X,Y}(x,y)$$

- **Properties**
  1. $\operatorname{cov}(X,Y) = E(XY) - E(X)E(Y)$
     - $E(XY) = \int \int xy f(x,y) \,\mathrm{d}y \,\mathrm{d}x$
  2. If X and Y are independent, $\operatorname{cov}(X,Y) = 0$
     - $X \perp Y \Rightarrow \operatorname{cov}(X,Y) = 0$
     - $\operatorname{cov}(X,Y) = 0 \nRightarrow X \perp Y$
     - $E(XY) = E(X)E(Y)$
  3. $\operatorname{cov}(aX+b, cY+d) = ac \cdot \operatorname{cov}(X,Y)$
  4. $V(aX+bY) = a^2 V(X) + b^2 V(Y) + 2ab \cdot \operatorname{cov}(X,Y)$

Probability Distributions

Discrete Distributions

Discrete Uniform Distribution

If random variable X assumes values $x_1, \ldots$ with *equal* probability, then X follows discrete uniform distribution. PMF of X is

$$f_{X(x)} = \begin{cases} \frac{1}{k}, x = x_1, \ldots, x_k \\ 0 \;\; \text{otherwise} \end{cases}$$

$$E(X) = \sum_{i=1}^{k} x_i f_X(x_i) = \frac{1}{k} \sum_{i=1}^{k} x_i$$

$$V(X) = E(X^2) - E(X)^2 = \frac{1}{k} \sum_{i=1}^{k} x_i^2 - \mu_x^2$$

- **Bernoulli Trial**: experiment with only 2 outcomes (1/0)
- **Bernoulli Random Variable**: X be no of sucess in Bernoulli trial. X has only 2 values. p is probability of success. PMF:

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

- $X \sim \text{Bernoulli}(p)$
  - $q = 1 - p$
  - $f_X(1) = p, f_X(0) = q$
  - $E(X) = p$
  - $V(X) = pq$
- **Bernoulli Process** - repeated independent and identical bernoulli trials
- **Binomial Random Variable** - No of successes in n trials of bernoulli process.
  - P of x successes in n trials
- $X \sim \text{Bin}(n,p)$
  - $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
  - $E(X) = np$
  - $V(X) = np(1-p)$
- **Negative Binomial Distribution** - No of trials needed for k successes
  - P of x trials needed for k successes
- $X \sim \text{NB}(k,p)$
  - $f_x(x) = P(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$
  - $E(X) = \frac{k}{p}$
  - $V(X) = \frac{(1-p)k}{p^2}$
- **Geometric Distribution**: No of trials needed until first success occurs.
- $X \sim \text{Geom}(p) = \text{NB}(1,p)$
  - $f_x(x) = P(X = x) = (1-p)^{x-1} p$
  - $E(X) = \frac{1}{p}$
  - $V(X) = \frac{1-p}{p^2}$