ST2334 Summary Notes

AY23/24 Sem 1, github.com/gerteck

1. Basic Probability Concepts

- **Sample Space:** S All possible outcomes of stat. expt.
- **Null Event**: Event that contains no element, empty set, $\varnothing$
- **Axioms of Probability**:
  For any event X, $0 \leq P(X) \leq 1$. $P(S) = 1$.
  If $A \cap B = \emptyset$ (Mut Excl), $P(A \cup B) = P(A) + P(B)$.
- Finite sample space with equally likely outcomes: $P(A)$ = ($\frac{\#samplepointsA}{\#totalsamplepointsS}$). (e.g. birthday problem)

Event Operation & Relationships

- **Event Operations:** Union, Intersection, Complement.
- **Event Relationships:** Contained: $A \subset B$
  Equivalence: $A \subset B$ with $A \supset B \to A = B$
  Mutually Exclusive: $A \cap B = \emptyset$.
- **De Morgan's Law:** $(A \cup B)' = A' \cap B'$ and $(A \cap B)' = A' \cup B'$

Counting Methods

- Multiplication Principle: (Sequential Events)
- Addition Principle: (Pairwise Disjoin sets)
- **Permutation**: ${}_nP_r = \frac{n!}{(n-r)!}$
- **Combination**: $\binom{n}{r} = \frac{n!}{(n-r)!r!}$

Conditional Probability

- Understand conditional as reduced sample space.

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$$

Independence

$A \perp B \leftrightarrow P(A \cap B) = P(A)P(B)$
$A \perp B \leftrightarrow P(A|B) = P(A)$

Law of Total Probability

- **Partition:** If $A_1, \cdots, A_n$ mutually exclusive, $\bigcup_{i=1}^{n} A_i = S$, then $A_1, \cdots, A_n$ are partitions.
- If $A_1, \cdots, A_n$ are partitions of S, then for any event B:

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Bayes' Theorem

Let $A_1, \cdots, A_n$ be partitions of S. For any event B:

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_k)P(A_i)}$$

For when $n = 2$, $\{A, A'\}$ becomes a partition of S.

$$P(A|B) = \frac{P(A)P(B|A))}{P(A)P(B|A) + P(A')P(B|A')}$$

2. Random Variables

A function X, which assigns a real number to every s $\in$ S is called a random variable.
- **Range space**: $Rx = \{x | x = X(s), s \in S\}$
- Likewise, the set X $\in$ A, for A being a subset of R, is also a subset of $S : s \in S : X(s) \in A$.

Probability Distribution

Two main types of RV used in practice: discrete and continuous.
- Probability assigned to each possible X
- Given RV X with range of R_x:
  Discrete: Numbers in R_x are finite or countable
  Continuous: R_x is interval

(Discrete) Probability Mass Function $f(x)$:

$$f(x) \begin{cases} P(X = x), & \text{for } x \in R_X \\ 0, & \text{for } x \notin R_X \end{cases}$$

1. $f(x_i) = P(X = x_i) \geq 0$ for $x_i \in R_x$
2. $f(x_i) = 0$ for $x_i \notin R_x$
3. $\sum_{i=1}^{\infty} f(x_i) = 1$ (PSum = 1)
4. $\forall B \subseteq \mathbb{R}, P(X \in B) = \sum_{x_i \in B \cap R_x} f(x_i)$

(Continuous) Probability Density Function $f(x)$:

- Given R_x is interval. Quantifies probability that X is in some range.
- $p.f.$ must satisfy:
  1. $f(x) \geq 0$, $f(x) = 0$ for $x \notin R_x$
  2. No need $f(x) \leq 1$ (Concerned with area)
  3. $\int_{R_x} f(x)dx = 1$ (Integration over R_X = 1)
  4. $\forall a, b$ s.t. $a \leq b, P(a \leq X \leq b) = \int_a^b f(x)dx$
- **Note:** $P(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$
- Hence, to check if a function is a pdf,
  1. $f(x) \geq 0$ for $x \in R_x$, $f(x) = 0$ for $x \notin R_x$
  2. $\int_{R_x} f(x)dx = 1$.

Cumulative Distribution Function

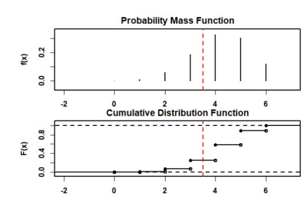
Describes distribution of a RV X: cumulative distribution function (cdf), applicable for discrete or continuous RV.

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

$F(x)$ is non-decreasing and $0 \leq F(x) \leq 1$
- Probability fn & cumulative distribution fn have one-to-one correspondence. For any probability fn given, the cdf is uniquely determined, vice versa.

CDF Discrete RV: Step Function $F(x)$

$$F(x) = \sum_{t \in R_x; t \leq x} f(t)$$

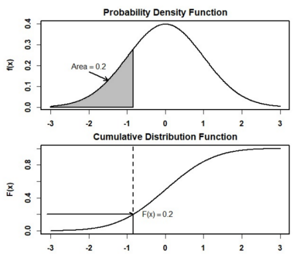


- $P(a \leq X \leq b) = P(X \leq b) - P(X < a)$
- $P(a \leq X \leq b) = F(b) - F(a-)$
- $P(a \leq X \leq b) = F(b) - \lim_{x \to a^-} F(x)$
- $0 \leq f(x) \leq 1$
- c.d.f has to be **right continuous** (• —)

CDF Continuous RV: $F(x)$

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

$$impt: f(x) = \frac{d(F(x))}{dx}$$



- $P(a \le X \le b) = P(a < X < b) = F(b) - F(a)$
- $0 \le f(x)$.
  e.g. $f(x) = 3x^2$ is a valid $p.f.$ since $\int_{R_x} f(x)dx = 1$

Expectation μ & Variance σ

Expectation of Random Variable: μ

- **Mean of discrete RV**:

$$\mu = E(X) = \sum_{x \in R_x} x_i f(x_i)$$

- **E.g.**: X discrete RV with p.m.f. $f(x)$ and range R_X
  $\mu = E(g(x)) = \sum_{x \in R_x} g(x) f(x)$

- **Mean of continuous RV**:

$$\mu = E(X) = \int_{x \in R_x} x f(x) dx$$

- **E.g.**: X continuous RV with p.d.f. $f(x)$ and range R_X
  $\mu = E(g(x)) = \int_{x \in R_x} g(x) f(x) dx$
- **Properties of Expectation:**
- $E(aX + b) = aE(X) + b$
- Linearity of expectation: $E(X + Y) = E(X) + E(Y)$

Variance of Random Variable: σ

$$\sigma_X^2 = V(X) = E[(X - \mu_X)^2]$$

- **Variance of discrete RV**:

$$V(X) = \sum_{x \in R_x} (x - \mu_X)^2 f(x)$$

- **Variance of continuous RV**:

$$V(X) = \int_{x \in R_x} (x - \mu_X)^2 f(x) dx$$

- $V(X) \ge 0$ and $V(X) = 0$ when X is a constant
- $V(aX + b) = a^2 V(X)$
- **alt. form:** $V(X) = E(X^2) - (E(X))^2$
- **Standard Deviation:** $\sigma_X = \sqrt{V(X)}$

3. Joint Distributions

- Consider more than 1 RV simultaneously,
- Given sample space S. Let X and Y be functions mapping $s \in S \to \mathbb{R}$: (X, Y) is 2D random vector.

  Range spc: $R_{X,Y} = \{(x, y) | x = X(s), y = Y(s), s \in S\}$
- **Discrete 2D RV:**
  # of possible values of $(X(s), Y(s))$ finite / countable
- **Continuous 2D RV**:
  # of possible values of $(X(s), Y(s))$ assume any value in some region of the Euclidean space $\mathbb{R}^2$
- If both X and Y are discrete/continuous, then (X, Y) is discrete/continuous respectively.

Joint Probability Function

- **Joint Probability (mass) function, 2D discrete RV**:

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

  - $f_{X,Y}(x, y) \ge 0$ for any $(x, y) \in R_{X,Y}$
  - $f_{X,Y}(x, y) = 0$ for any $(x, y) \notin R_{X,Y}$
  - $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(X = x_i, Y = y_i) = 1$
  - Let $A \subseteq R_{X,Y}$.
    $P((X, Y) \in A) = \sum \sum_{(x,y) \in A} f_{X,Y}(x, y)$

- **Joint Probability (density) function, 2D cont. RV**:

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

  - $f_{X,Y}(x, y) \ge 0$ for any $(x, y) \in R_{X,Y}$
  - $f_{X,Y}(x, y) = 0$ for any $(x, y) \notin R_{X,Y}$
  - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$
    or equivalently:
  - $\int \int_{(x,y) \in R_{X,Y}} f_{X,Y}(x, y) dx dy = 1$

Marginal Probability Function

Marginal distribution of X is individual distribution of X, ignoring the value of Y. "Projection" of 2D function $f_{X,Y}(x, y)$ to 1D function.

Let (X, Y) be 2D RV with joint probability function $f_{X,Y}(x, y)$:

If Y is **discrete**, $f_X(x) = \sum_y f_{X,Y}(x, y)$

If Y is **continuous**, $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$

- $f_Y(y)$ defined similarly
- $f_X(x)$ is a $p.f.$, satisfies all properties of prob. fn.

Conditional Distribution

Let (X, Y) be 2D RV with joint probability function $f_{X,Y}(x, y)$. Then $\forall x$ s.t. $f_X(x) > 0$: (X marg prob fn.)
Conditional probability function of Y given $X = x$:

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

- Intuition: Distribution of Y given $X = x$
- Only defined for x s.t. $f_X(x) > 0$
- $f_{Y|X}(y|x)$ is a $p.f.$ if we fix x, satisfies prop. of prob.fn.
- But, $f_{Y|X}(y|x)$ is not a $p.f.$ for x: No need for sum / integral over x = 1. Hence,
  If $f_X(x) > 0$: $f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y|x)$
  If $f_Y(y) > 0$: $f_{X,Y}(x, y) = f_Y(y) f_{X|Y}(x|y)$
- **Probability Y $\le y$, Average Y given $X = x$**
- $P(Y \le y | X = x) = \int_{-\infty}^{y} f_{Y|X}(y|x) dy$
- $E(Y | X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$