ST2334

AY22/23 Sem 1

github.com/jasonqiu212

01. Basic Concepts of Probability

Event Operations

- **Mutually Exclusive** - $A \cap B = \emptyset$
- **Contained** - $A \subset B$
- **Equivalence** - $A \subset B$ and $A \supset B \to A = B$
- **Distributive** - $A \cap (B \cup C) = (A \cup B) \cup (A \cup C)$
- **DeMorgan's** - $(A \cup B)' = A' \cap B'$
- $A = (A \cap B) \cup (A \cap B')$

Counting Methods

- **Multiplcation Principle** - Given r experiments performed **sequentially** and each has $n_1, n_2, ..., n_r$ outcomes. After r experiments, there are $n_1 n_2 ... n_r$ outcomes.
- **Addition Principle** - Given experiment can be done in k different ways and each has $n_1, n_2, ..., n_r$ ways. There are $n_1 + n_2 + ... + n_k$ total ways.
- **Permutation** - ${}_nP_r = \frac{n!}{(n-r)!}$
- **Combination** - $\binom{n}{r} = \frac{n!}{(n-r)!r!}$

Probability

Axioms of Probability

1. For any event A, $0 \leq P(A) \leq 1$
2. P(S) = 1
3. If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$

- $P(A') = 1 - P(A)$
- $P(A) = P(A \cap B) + P(A \cap B')$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If $A \subset B$, then $P(A) < P(B)$

Finite Sample Space with Equally Likely Outcomes

Given sample space $S = \{a_1, ..., a_k\}$ and all outcomes are **equally likely**, i.e. $P(a_1) = ... = P(a_k)$:

$$\text{For any event } A \subset S, P(A) = \frac{\text{No. of sample points in A}}{\text{No. of sample points in S}}$$

Conditional Probability

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

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 $A_1, ..., A_n$ are mutually exclusive events and $\bigcup_{i=1}^n A_i = S$, then $A_1, ..., A_n$ are partitions
- If $A_1, ..., A_n$ 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 $A_1, ..., A_n$ be partitions of S. For any event B:

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

02. Random Variables

- Motivation: Assign value to outcome of experiment
- **Random Variable** - Let S be sample space. Function X which maps $\mathbb{R}$ to every $s \in S$

Probability Distribution

- 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 Distribution

- **Probability Function** - Given $R_x = \{x_1, ...\}$. For each x_i, there's some probability that $X = x_i$:

$$f(x) = P(X = x)$$

- $p.f.$ must satisfy:
  1. $f(x_i) = P(X = x_i)$ for $x_i \in R_x$
  2. $f(x_i) = 0$ for $x_i \notin R_x$
  3. $\sum_{i=1}^{\infty} f(x_i) = 1$
  4. $\forall B \subseteq \mathbb{R}, P(X \in B) = \sum_{x_i \in B \cap R_x} f(x_i)$
- **Probability Distribution** - Collection of pairs $(x_i, f(x_i))$

Continuous Probability Distribution

- **Probability Function** - Given R_x is interval. Quantifies probability that X is in some range.
- $p.f.$ must satisfy:
  1. $f(x) \geq 0$
  2. $f(x) = 0$ for $x \notin R_x$
  3. $\int_{R_x} f(x)dx = 1$
  4. $\forall a, b$ s.t. $a \leq b, P(a \leq X \leq b) = \int_a^b f(x)dx$
- Note: $P(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$

Cumulative Distributive Function

Given RV X, which can be discrete or continuous:

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

- $F(x)$ is non-decreasing and $0 \leq F(x) \leq 1$
- **For discrete RV**: Step function

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

  - $P(a \leq X \leq b) = F(b) - \lim_{x \to a^-} F(x)$
  - $0 \leq f(x) \leq 1$
- **For continuous RV**:

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

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

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

Expectation of Random Variable

- **Mean of discrete RV**:

$$\mu = E(X) = \sum_{x \in R_x} x_i f(x_i) = \sum_{i=1}^{\infty} P(X \geq i)$$

  - Let g be some function. $E(g(x)) = \sum_{x \in R_x} g(x)f(x)$
- **Mean of continuous RV**:

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

  - Let g be some function. $E(g(x)) = \int_{x \in R_x} g(x)f(x)dx$
- $E(aX + b) = aE(X) + b$
- Linearity of expectation: $E(X + Y) = E(X) + E(Y)$

Variance of Random Variable

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

- **Variance of discrete RV**:

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

- **Variance of continuous RV**:

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

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

03. Joint Distributions

- Motivation: What if interested in more than 1 RV simultaneously?
- Given sample space S. Let X and Y be functions mapping $s \in S \to \mathbb{R}$:

$$(X, Y) \text{is 2D random vector}$$

$$\text{Range space: } R_{X,Y} = \{(x, y) | x = X(s), y = Y(s), s \in S\}$$

- **Discrete 2D RV** - If no. of possible values of $(X(s), Y(s))$ are finite or countable
- **Continuous 2D RV** - If no. of possible values of $(X(s), Y(s))$ can be any value in Euclidean space $\mathbb{R}^2$
- If both X and Y are discrete/continuous, then (X, Y) is discrete/countinuous.

Joint Probability Function

- **For discrete**:

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

  - $f_{X,Y}(x,y) \geq 0$ 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 $A \subseteq R_{X,Y}$. $P((X,Y) \in A) = \sum\sum_{(x,y)\in A} f_{X,Y}(x,y)$

- **For continuous**:

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

  - $f_{X,Y}(x,y) \geq 0$ for any $(x,y) \in R_{X,Y}$
  - $f_{X,Y}(x,y) = 0$ for any $(x,y) \notin R_{X,Y}$
  - $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$

Marginal Probability Function

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

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

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

- $f_Y(y)$ defined similarly
- Intuition: Marginal distribution for X ignores presence of Y
- $f_X(x)$ is a $p.f.$

Conditional Distribution

Let (X,Y) be a 2D RV with joint probability function $f_{X,Y}(x,y)$. Then $\forall x$ s.t. $f_X(x) > 0$:

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

- 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
- But, $f_{Y|X}(y|x)$ is not a $p.f.$ for x
- $P(Y \leq y | X = x) = \int_{-\infty}^{y} f_{Y|X}(y|x)dy$
- $E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy$

Independent Random Variables

$$X \perp Y \leftrightarrow \forall x, y, f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

- Necessary condition: $R_{X,Y}$ must be a product space. Else, dependent.

Properties

Suppose X, Y are independent RV:

- If $A, B \subseteq \mathbb{R}$, then events $X \in A$ and $Y \in B$ are independent:

$$P(X \in A; Y \in B) = P(X \in A)P(Y \in B)$$

- $g_1(X)$ and $g_2(Y)$ are independent
- Independence is related with conditional distribution:

$$f_X(x) > 0 \rightarrow f_{Y|X}(y|x) = f_Y(y)$$

$$f_Y(y) > 0 \rightarrow f_{X|Y}(x|y) = f_X(x)$$

Quick way to check independence

1. $R_{X,Y}$ is a product space. i.e. R_X does not depend on Y and vice versa.
2. $f_{X,Y}(x,y)$ can be written as $cg_1(x)g_2(y)$ where g_1 depends on x only and g_2 depends on y only.
3. For discrete: $f_X(x) = \frac{g_1(x)}{\sum_{t \in R_X} g_1(t)}$
4. For continuous: $f_X(x) = \frac{g_1(x)}{\int_{t \in R_X} g_1(t)dt}$

Expectation

Given 2 variable function $g(x,y)$:

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

$$\text{If } (X,Y) \text{ is continuous, } E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dydx$$

- $E(XY) = E(X)E(Y)$ if $X \perp Y$

Covariance

$$cov(X,Y) = E((X - E(X))(Y - E(Y)))$$

$$\text{If } (X,Y) \text{ is discrete, } cov(X,Y) = \sum_x\sum_y (x - \mu_X)(y - \mu_Y)f_{X,Y}(x,y)$$

$$\text{If } (X,Y) \text{ is cont., } cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f_{X,Y}(x,y)dxdy$$

- $cov(X,Y) = E(XY) - E(X)E(Y)$
- $X \perp Y \rightarrow cov(X,Y) = 0$. But converse is not always true.
- $cov(aX + b, cY + d) = (ac)cov(X,Y)$
- $V(aX + bY) = a^2V(X) + b^2V(Y) + 2abcov(X,Y)$
- $X \perp Y \rightarrow V(X + Y) = V(X) + V(Y)$

04. Special Probability Distributions

Discrete Uniform Distribution

- If X has values $x_1, x_2, ..., x_k$ with **equal probability**
- $p.f.$: $f_X(x) = \frac{1}{k}$ where $x = x_1, ..., x_k$ and 0 otherwise
- Expectation: $\mu_X = E(X) = \sum_{i=1}^{k} x_i f_X(x_i) = \frac{1}{k}\sum_{i=1}^{k} x_i$
- Variance: $\sigma_X^2 = V(X) = E(X^2) - (E(X))^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \mu_X^2$

Bernoulli

- **Bernoulli Trial** - Random experiment with 2 possible outcomes (success and failure)

Bernoulli Random Variable

- Number of successes in Bernoulli trial (Either 1 or 0)
- Let $0 \leq p \leq 1$ be the probability of success in Bernoulli trial

$$f_X(x) = P(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & otherwise \end{cases}$$

- $f_X(x) = p^x(1-p)^{1-x}$ for $x = 0$ or 1
- Notation: $X \sim Ber(p)$ and $q = 1 - p$
- $\mu_X = E(X) = p$ and $\sigma_X^2 = V(X) = p(1-p)$

Bernoulli Process

- Sequence of repeatedly performed **independent and identical** Ber. trials
- Generates sequence of **independet and identically distributed (i.i.d.)** Ber. RVs: $X_1, X_2, ...$

Binomial Distribution

- **Binomial RV** - Counts the number of successes in n trials in a Ber. process
- Given n trials with each trial having probability p of success:

$$P(X = x) = \binom{n}{x} p^x(1-p)^{n-x}$$

- Notation: $X \sim B(n,p)$
- $E(X) = np$ and $V(X) = np(1-p)$

Negative Binomial Distribution

- Let X = Number of i.i.d. Bernoulli(p) trials until kth success occurs

$$P(X = x) = \binom{x-1}{k-1} p^k(1-p)^{x-k}$$

- Notation: $X \sim NB(k,p)$
- $E(X) = \frac{1}{p}$ and $V(X) = \frac{(1-p)}{p^2}$

Geometric Distribution

- Let X = Number of i.i.d. Bernoulli(p) trials until 1st success occurs

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

- Notation: $X \sim G(p)$
- $E(X) = \frac{1}{p}$ and $V(X) = \frac{1-p}{p^2}$

Poisson Distribution

- **Poisson RV** - Denotes number of events happening in **fixed period of time or fixed region**

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

- Notation: $X \sim Poisson(\lambda)$ where $\lambda > 0$ is expected number of occurrences during given period/region
- $E(X) = \lambda$ and $V(X) = \lambda$

Poisson Process

- Continuous time process, where we count number of occurrences within some interval of time
- Given Poisson process with rate parameter α:
  - Expected number of occurrences in interval of length T is αT
  - No simultaneous occurrences
  - Number of occurrences in disjoint intervals are independent
- Number of occurrences in any interval T of Poisson process follows $Poisson(\alpha T)$ distribution

Poisson Approximation of Binomial Distribution

Let $X \sim B(n,p)$. Suppose $n \to \infty$ and $p \to 0$ s.t. $\lambda = np$ remains constant. Then $X \sim Poisson(\lambda)$ approximately.

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

- Approximation is good when $n \geq 20$ and $p \leq 0.05$, or $n \geq 100$ and $np \leq 10$

Continuous Uniform Distribution

X follows uniform distribution over interval (a, b) if $p.f.$ is given by:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & otherwise \end{cases}$$

- Notation: $X \sim U(a, b)$
- $E(X) = \frac{a+b}{2}$ and $V(X) = \frac{(b-a)^2}{12}$
- $c.d.f.$ is given by:

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

Exponential Distribution

X follows exponential distribution with parameter $\lambda > 0$ if $p.f.$ is given by:

$$f_x(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

- Notation: $X \sim Exp(\lambda)$
- $E(X) = \frac{1}{\lambda}$ and $V(X) = \frac{1}{\lambda^2}$
- $c.d.f.$ is given by:

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

- Suppose X has exponential distribution with parameter $\lambda > 0$. Then for any positive numbers s and t, we have:

$$P(X > s + t | X > s) = P(X > t)$$

Normal Distribution

X follows normal distribution with mean μ and variance σ^2 if $p.f.$ is given by:

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

- Notation: $X \sim N(\mu, \sigma^2)$
- $E(X) = \mu$ and $V(X) = \sigma^2$
- $p.f.$ is bell-shaped curve and symmetric about $x = \mu$
- Total area under curve is 1
- 2 normal curves are identical in shape if they have same σ^2. They differ in location by $\mu_1 - \mu_2$.
- As σ increases, curve becomes more spread out
- If $X \sigma N(\mu, \sigma^2)$ and let $Z = \frac{X-\mu}{\sigma}$

Standardized Normal Distribution

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

$$Z = \frac{X - \mu}{\sigma}$$

then $Z \sim N(0, 1)$

- $E(Z) = 0$ and $V(Z) = 1$
- $p.f.$ is given by:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

- Importance of standardizing normal distribution is that it allows us to use tables to find probabilities
- Let $X \sim N(\mu, \sigma^2)$. We can compute $P(x_1 < X < x_2)$ by standadization:

$$x_1 < X < x_2 \leftrightarrow \frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma}$$

- $c.d.f.$ is given by:

$$\phi(z) = F_Z(z) = \int_{-\infty}^{z} f_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

- $P(Z \ge 0) = P(Z \le 0) = \phi(0) = 0.5$
- For any z, $\phi(z) = P(Z \le z) = P(Z \ge -z) = 1 - \phi(-z)$
- $-Z \sim N(0, 1)$
- If $Z \sim N(0, 1)$, then $\sigma Z + \mu \sim N(\mu, \sigma^2)$
- **Quantile** - The αth quantile is x_α that satisfies:

$$P(X \ge x_\alpha) = \alpha$$

  e.g. The 0.05th quantile of $Z \sim N(0, 1)$ is 1.645, i.e. $z_{0.05} = 1.645$.
- $P(Z \ge z_\alpha) = P(Z \le -z_\alpha) = \alpha$

Normal Approximation to Binomial Distribution

Let $X \sim B(n, p)$, then as $n \to \infty$:

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

- Approximation is good when $np > 5$ and $n(1-p) > 5$