Mathematical Probability Theorems and Formulas

Jason Lim

12/10/2020

Contents

1 Set Theory

1.1 De Morgan's Laws

Suppose $\{A_j\}_{j=1}^k$ are a collection of sets in B, i.e. $A_j \subseteq B \ \forall j = 1, \dots, k$

$$(A \cup B)' = A' \cap B'$$
$$(A \cap B)' = A' \cup B'$$

General form:

$$\left(\bigcup_{j=1}^{k} A_j\right)' = \bigcap_{j=1}^{k} A_j'$$
$$\left(\bigcap_{j=1}^{k} A_j\right)' = \bigcup_{j=1}^{k} A_j'$$

2 Probability and Events

2.1 Probability Space

Definition 2.1 (Probability space)**.** A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where:

- Ω is a nonempty set called the static space.
- $\mathcal{F}$ is a collection of subsets of Ω (note: $\emptyset, \Omega, \in \mathcal{F}$)
  - An element $A \in \mathcal{F}$ is called an event and $A \subseteq \Omega$
  - (in this course, $\mathcal{F} = \{\text{all subsets of } \Omega\}$
- A function $\mathbb{P} : \mathcal{F} \longrightarrow [0, 1]$ satisfying:
  1. $\mathbb{P}(A) \geq 0$ for any $A \in \mathcal{F}$
  2. $\mathbb{P}(\Omega) = 1$
  3. If $\{A_j\}_{j=1}^k$ are events such that $A_i \cap A_j = \emptyset$ for $i \neq j$ (mutually exclusive), then

$$\mathbb{P}\left(\bigcup_{j=1}^{k} A_j\right) = \sum_{j=1}^{k} \mathbb{P}(A_j)$$

and when $k = +\infty$,

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$$

Proposition 2.1. $\mathbb{P}(\emptyset) = 0$. *This lets us say if* $A \cap B = \emptyset$, $\mathbb{P}(A \cap B = \mathbb{P}(\emptyset) = 0$

Proposition 2.2. *If A is an event, then $A' = \Omega \backslash A$ is an event and*
$\mathbb{P}(A') = 1 - \mathbb{P}(A)$

Proposition 2.3. *If $A \subseteq B$, then* $\mathbb{P}\left(B \backslash A\right) = \mathbb{P}\left(B\right) - \mathbb{P}\left(A\right)$

Proposition 2.4. *If $A \subseteq B$ then* $\mathbb{P}\left(A\right) \leq \mathbb{P}\left(B\right)$

2.2 Inclusion-exclusion principle

Theorem 2.1 (Inclusion-exclusion principle)**.** *For any events $A, B \subseteq \Omega$,*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

2.3 Independent events

Definition 2.2. $A, B \subset \Omega$ are <u>independent</u> if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Otherwise, we say that they are <u>dependent</u>.

Note: Don't confuse independent with disjoint ($A \cap B = \emptyset$). If A, B are both independent and disjoint, then $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$.

Proposition 2.5. *If A, B are independent events, then so are:*

1. *A and B'*
2. *A' and B'*

Definition 2.3. We say events $A_1, \ldots, A_n \subseteq \Omega$ are <u>mutually independent</u> if, given $1 \leq k \leq n$ and $1 \leq j_1 < j_2 < \cdots < j_k \leq n$, we have

$$\mathbb{P}\left(\bigcap_{l=1}^{k} A_{j_l}\right) = \prod_{l=1}^{k} \mathbb{P}(A_{j_l})$$

3 Methods of Enumeration

3.1 The multiplication principle

Let $r \in \mathbb{N}_{>0}$. Suppose that we run r independent experiments and:

- The first experiment has n_1 possible outcomes.
- The second experiment has n_2 possible outcomes.
  ...
- The r^{th} experiment has n_r possible outcomes.

Unordered without replacement: When ignoring order, each chosen set of size r is considered equivalent to all its $r!$ permutations, so the number of distinct possibilities is simply divided by $r!$.

Unordered with replacement: If we take unordered samples of size r from a set of n objects with replacement, then the number of samples is

$${}_{n+r-1}C_r = \binom{n+r-1}{r}$$

3.2 The Binomial Theorem

Theorem 3.1 (The Binomial Theorem)**.** *If $n \geq 0$, then*

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}$$

$\binom{n}{r}$ *shows up in this formula because:*

- *When $(x+y)^n$ is expanded, each part of the resulting sum will have exactly n factors.*
- *Since multiplication is commutative, two addends are considered equivalent if they contain the same number of x's and y's. For example, $xxxyx = xxyxx$.*
- *Each possible combination of r x's and $r-n$ y's appears once.*

Thus the coefficient of $x^r y^{n-r}$ is the number of ways of choosing r x's out of n x's and y's, or $\binom{n}{r}$

Theorem 3.2 (The Multinomial Coefficient)**.** *For $1 \leq r \leq n$ and $n_1 = \cdots + n_r = n$. Then, there are*

$$\binom{n}{n_1, n_2, \ldots, n_r} := \frac{n!}{n_1! n_2! \cdots n_r!}$$

distinguishable permutations.

4 More Probability

Definition 4.1 (Conditional Probability)**.** Let $B \subseteq \Omega$ be an event so that $\mathbb{P}(B) \neq 0$. The probability of an event $A \subseteq \Omega$ <u>conditioned on the event B</u> is given by

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

"probability that A occurs given B has occurred"

Theorem 4.1 (Properties of the conditional probability)**.** *If $B \subseteq \Omega$ is an event such that $\mathbb{P}(B) = 0$, then $\mathbb{P}(\cdot|B)$ is a probability measure, i.e. $\mathbb{P}(\cdot|B) : \mathcal{F} \to [0,1]$ satisfies:*

1. $\mathbb{P}(\Omega|B) = 1$
2. *(Countable additivity) If $\{A_j\}_{j=1}^k$ are mutually exclusive events, then*

$$\mathbb{P}\left(\bigcup_{j=1}^{k} A_j \middle| B\right) = \sum_{j=1}^{k} \mathbb{P}(A_j|B)$$

And when $k = +\infty$,

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j \middle| B\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j|B)$$

Proposition 4.1. *If $A, B \subseteq \Omega$ are independent and $\mathbb{P}(B) \neq 0$, then*

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

Theorem 4.2 (The Law of Total Probability)**.** *Let $A \subseteq \Omega$ be an event and $\{B_j\}_{j=1}^k \subseteq \Omega$ be mutually exclusive, satisfying $\mathbb{P}(B_j) \neq 0$ for every $j = 1, \dots, k$, and*

$$A \subseteq \bigcap_{j=1}^{k} B_j$$

Then

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|B_1)\,\mathbb{P}(B_1) + \cdots + \mathbb{P}(A|B_k)\,\mathbb{P}(B_k) \\ &= \sum_{j=1}^{k} \mathbb{P}(A|B_j)\,\mathbb{P}(B_j)\end{aligned}$$

4.1 Bayes' Theorem

Theorem 4.3 (Bayes' Theorem)**.** *If $A, B \subseteq \Omega$ are events such that $\mathbb{P}(A), \mathbb{P}(B) \neq 0$, then*

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

Theorem 4.4 (Bayes' Theorem v2)**.** *If $A \subseteq \Omega$ is an event, $\{B_j\}_{j=1}^k \subseteq \Omega$ are mutually exclusive events so that $\mathbb{P}(A), \mathbb{P}(B_j) \neq 0$, and*

$$A \subseteq \bigcap_{j=1}^{k} B_j$$

Then for any $1 \leq \ell \leq k$, we have

$$\mathbb{P}(B_\ell|A) = \frac{\mathbb{P}(A|B_\ell)\,\mathbb{P}(B_\ell)}{\sum_{j=1}^{k} \mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}$$

This follows trivially from Bayes' Theorem and the law of total probability.

5 Random Variables

Definition 5.1 (Random variable)**.** Given a set S and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable is a function

$$X : \Omega \to S$$

Notation: If $x \in s$, and $A \subseteq S$, we write

$$\begin{aligned}\mathbb{P}(X = x) &:= \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) \\ \mathbb{P}(X \in A) &:= \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})\end{aligned}$$

Definition 5.2 (Discrete random variable)**.** A random variable $X : \Omega \to S$ is discrete if $S \subseteq \mathbb{R}$ is finite or countable (i.e. in one-to-one correspondence with $\mathbb{N}$)

We can think of discrete random variables as random numbers.

5.1 PMF and CDF

Given a discrete random variable X taking values in $S \subseteq \mathbb{R}$, we define:

- the probability mass function (PMF) of X as the function $p_X : S \to [0,1]$ defined by

$$p_X(x) = \mathbb{P}(X = x)$$

- the cumulative distribution function (CDF) of X as the function $F_X : S \to [0,1]$ defined by

$$F_X(x) = \mathbb{P}(X \leq x)$$

- We say that two random variables X and Y are identically distributed if they have the same CDF and we write $X \sim Y$

Proposition 5.1. *If X is a discrete random variable and $A \subseteq \mathbb{R}$ is any set, then*

$$\mathbb{P}(X \in A) = \sum_{x \in A \cap S} p_X(x)$$

Proposition 5.2. *if X is a discrete random variable and $A \subseteq \mathbb{R}$ is any set, then*

$$F_X(x) = \sum_{\substack{y \in S \\ y \leq x}} p_X(y)$$

Proposition 5.3. *If X is a discrete random variable and $a < b$, then*

$$\mathbb{P}(a < x \leq b) = F_X(b) - F_X(a)$$

5.2 Expected Value

Definition 5.3. If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$, its expected value is defined to be

$$\mathbb{E}[X] = \sum_{x \in s} x p_X(x)$$

provided the sum converges.

Notation: we also write $\mu_X = \mathbb{E}[X]$

Proposition 5.4. *If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$, and $g : S \to \mathbb{R}$ is a function, then the expected value of $g(X)$ is*

$$\mathbb{E}[g(X)] = \sum_{x \in S} g(x) p_X(x)$$

provided the sum converges.

Proposition 5.5 (Linearity of Expectation)**.** *If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$. If $a, b \in \mathbb{R}$ and $g, h : S \to \mathbb{R}$, then*

$$\mathbb{E}[ag(X) + bh(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)]$$

Proposition 5.6. *If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$ and $g, h : S \to \mathbb{R}$ such that $g(x) \leq h(x)$ for all $x \in S$, then*

$$\mathbb{E}[g(X)] \leq \mathbb{E}[h(X)]$$

5.3 Moment, Variance, and Standard Deviation

Definition 5.4. If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$ and $b \in \mathbb{R}$, we define the r^{th} moment of X about b to be

$$\mathbb{E}[(X-b)^r]$$

Definition 5.5. Let X be a discrete random variable. We define the variance of X to be

$$\operatorname{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Whenever is converges, we use the notation $\sigma^2 = \operatorname{var}(X)$

The standard deviation of X is $\sigma_X = \sqrt{\operatorname{var}(X)}$

Proposition 5.7. *If X is a discrete random variable and $a, b \in \mathbb{R}$, then:*

$$\mathbb{E}[aX+b] = a\mathbb{E}[X] + b$$
$$\operatorname{var}(aX+b) = a^2 \operatorname{var}(X)$$

Proposition 5.8 (Alternate Formula for Variance)**.** *If X is a discrete random variable, then*

$$\operatorname{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Definition 5.6 (Moment Generating Function)**.** If X is a discrete random variable, we define the moment generating function (MGF) of X to be the function

$$M_X(t) = \mathbb{E}[e^{tX}], \quad t \in \mathbb{R}$$

whenever it exists.

Proposition 5.9. *Let X be a discrete random variable with MGF $M_X(t)$ which is well-defined and smooth for $t \in (-\delta, \delta)$ for some $\delta > 0$. Then*

$$\frac{d^r}{dt^r} M_X\Big|_{t=0} = \mathbb{E}[X^r] \quad r \in \{1, 2, 3, \dots\}$$

Proposition 5.10 (Expectation and Variance from MGF)**.** *Let X be a discrete random variable with MGF $M_X(t)$ which is well-defined and smooth for $t \in (-\delta, \delta)$ for some $\delta > 0$. Then*

$$\frac{d}{dt} \log M_X\Big|_{t=0} = \mathbb{E}[X]$$
$$\frac{d^2}{dt^2} \log M_X\Big|_{t=0} = \operatorname{var}(X)$$

6 Types of Discrete Random Variables

6.1 Bernoulli

Definition 6.1. Let $p \in (0,1)$. We say that a discrete random variable X is a Bernoulli random variable and write $X \sim \text{Bernoulli}(p)$ if it has PMF

$$p_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$$

6.2 Uniform (Discrete)

Definition 6.2. Let $m \geq 1$. A discrete random variable X is uniformly distributed on $\{1, 2, \ldots, m\}$ and we write

$$X \sim \text{Uniform}(\{1, 2, \ldots, m\})$$

if it has PMF

$$p_X(x) = \frac{1}{m} \text{ for } x \in \{1, 2, \ldots, m\}$$

If $X \sim \text{Uniform}(\{1, 2, \ldots, m\})$, then it has CDF

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{k}{m} & \text{if } k \leq x < k+1, \ k \in \{1, \ldots, m-1\} \\ 1 & \text{if } x \geq m \end{cases}$$

6.3 Binomial

Definition 6.3 (Bernoulli trial)**.** An experiment that has probability $p \in (1, 0)$ of success and probability $(1-p)$ of failure

Definition 6.4. If X Binomial(n, p) then it has PMF

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

for $x \in \{0, 1, \ldots, n\}$

Proposition 6.1. *If X Binomial(n, p), then its MGF is*

$$M_X(t) = (1 - p + pe^t)^n$$

Proposition 6.2. *If X Binomial(n, p), then*

$$\mathbb{E}[X] = np$$

Proposition 6.3. *If X Binomial(n, p), then*

$$\text{var}(X) = np(1-p)$$

6.4 Geometric

Definition 6.5. Suppose we run independent, identical Bernoulli trials with probability $p \in (0, 1)$ of success.

- Let X be the trial on which we first achieve success.
- Then X is a discrete rv taking values in $S = \{1, 2, 3, \ldots\}$
- We say that X is a Geometric random variable with parameter p and write $X \sim \text{Geometric}(p)$

Proposition 6.4. *If $X \sim$ Geometric(p), then its PMF is*

$$p_X(x) = (1-p)^{x-1} p \textit{ for } x \in \{1, 2, 3, \ldots\}$$

Proposition 6.5. *If $X \sim$ Geometric(p) and k an integer, then*

$$\mathbb{P}(X \leq k) = 1 - (1-p)^k$$

Proposition 6.6. *If $X \sim$ Geometric(p), then its MGF is*

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t}$$

for $t < -\log(1-p)$

Proposition 6.7. *If $X \sim$ Geometric(p), then its mean is*

$$\mathbb{E}[X] = \frac{1}{p}$$

Proposition 6.8. *If $X \sim$ Geometric(p), then its variance is*

$$\mathrm{var}(X) = \frac{1-p}{p^2}$$

6.5 Negative Binomial

Definition 6.6.

- Suppose we run independent, identical Bernoulli trials with probability $p \in (0,1)$ of success.
- Let $r \geq 1$ and X be the trial on which we first achieve the rth success.
- Then X takes values in $S = \{r, r+1, r+2, \dots\}$
- We say that X is a negative binomial random variable with parameter r, p and write $X \sim$ Negative Binomial(r, p)
- If $r = 1$, then $X \sim$ Negative Binomial$(1, p) \sim$ Geometric(p)

Proposition 6.9. *If $X \sim$ Negative Binomial(r,p), then its PMF is*

$$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \qquad \text{if } x \in \{r, r+1, \dots\}$$

Lemma 6.1. *If $r \geq 1$ is an integer $0 < s < 1$, then*

$$\left(\frac{1}{1-s}\right)^r = \sum_{x=r}^{\infty} \binom{x-1}{r-1} s^{x-r}$$

Proposition 6.10. *If $X \sim$ Negative Binomial(r,p), then its MGF is*

$$M_X(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^r \qquad \text{if } t < \log(1-p)$$

Proposition 6.11. *If $X \sim$ Negative Binomial(r,p), then*

$$\mathbb{E}[X] = \frac{r}{p}$$

$$\mathrm{var}(X) = \frac{r(1-p)}{p^2}$$

6.6 Poisson

Assumptions:

- We make the following assumptions about the arrivals:
    1. If the time intervals $(t_1, t_2], (t_3, t_3], \ldots, (t_n, t_{n+1}]$ are *disjoint*, then the number of arrivals in each time intervals are *independent*
    2. if $h = t_2 - t_1 > 0$ is sufficiently small, then the probability of exactly one arrival in the time interval $(t_1, t_2]$ is λh
    3. if $h = t_2 - t_1 > 0$ is sufficiently small, then the probability of two or more arrivals in the time interval $(t_1, t_2]$ converges rapidly to zero as $h \to 0$
- An arrival process satisfying these assumptions is called an approximate Poisson process.
- The random variable X is called a Poisson random variable and we write $X \sim \text{ Poisson}(\lambda)$

Proposition 6.12. *If $X \sim$ Poisson(λ), then its PMF is*

$$p_X(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad \text{if } x \in \{0, 1, 2, \ldots\}$$

Proposition 6.13. *Consider an approximate Poisson process with rate $\lambda > 0$ per unit time. Let X be the number of arrivals in a time of length $T > 0$ units. Then $X \sim$ Poisson(λT)*

Proposition 6.14. *If $\lambda > 0$ and $X \sim$ Poisson(λT), then its MGF is*

$$M_X(t) = e^{\lambda(e^t - 1)}$$

Proposition 6.15. *If $\lambda > 0$ and $X \sim$ Poisson(λT), then*

$$\begin{aligned}\mathbb{E}[X] &= \lambda \\ \text{var}(X) &= \lambda\end{aligned}$$

7 Continuous Random Variables

Definition 7.1. Let $S \subseteq \mathbb{R}$, and let $X : \Omega \to S$ be a random variable.

- We define the cumulative distribution function of X,
  $F_X : \mathbb{R} \to [0, 1]$ by
  $$F_X(x) = \mathbb{P}(X \leq x)$$
  We have
  $$\lim_{x \to -\infty} F_X(x) = 0 \text{ and } \lim_{x \to \infty} F_X(x) = 1$$
- We say that X is a continuous random variable if there exists a non-negative integrable function $f_X : \mathbb{R} \to [0, \infty)$ such that
  $$\mathbb{P}(X \leq x) = F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$
- We call f_X a probability density function for X.

Proposition 7.1. *If X is a continuous random variable with PDF*
$f_X : \mathbb{R} \to [0, \infty)$, then

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Proposition 7.2. ***Proposition:*** *If X is a continuous random variable with PDF $f_X : \mathbb{R} \to [0, \infty)$ and $a < b$, then*

$$\mathbb{P}(a < X \leq B) = \int_a^b f_X(x)dx$$

as a consequence of the statement $\mathbb{P}(a < X \leq B) = F_X(b) - F_X(a)$

Proposition 7.3. *If X is a continuous random variable with PDF $f_X : \mathbb{R} \to [0, \infty)$, then for any $x \in \mathbb{R}$*

$$\mathbb{P}(X = x) = 0$$

In particular, if $a < b$, then

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b)$$

7.1 Expectation

Definition 7.2. If X is a continuous random variable with PDF $f_X(x)$, we define its expected value to be

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

More generally, if $g : \mathbb{R} \to \mathbb{R}$ is any function, then

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$$

Proposition 7.4. *Let X be a continuous random variable.*

- *if $a \in \mathbb{R}$ is a constant, then*
$$\mathbb{E}[a] = a$$
- *if $a, b \in \mathbb{R}$ are constants and $g, h : \mathbb{R} \to \mathbb{R}$ are functions, then*
$$\mathbb{E}[ag(X) = bh(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)]$$
- *If $g(x) \leq h(x)$ for all $x \in \mathbb{R}$, then*
$$\mathbb{E}[g(X)] \leq \mathbb{E}[h(X)]$$

7.2 Moment, Variance

Definition 7.3. If X is a continuous random variable, then we define its variance to be

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

We will use the notation $\sigma_X^2 = \text{var}(X)$ and define the standard deviation to be $\sigma_X = \sqrt{\text{var}(X)}$. We also have

$$\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Definition 7.4. If X is a continuous random variable we define its moment generating function to be

$$M_X(t) = \mathbb{E}[e^{tX}]$$

for all $t \in \mathbb{R}$ for which this makes sense.

Proposition 7.5. *If $M_X(t)$ is smooth on some interval $(-\delta, \delta)$ where $\delta > 0$, then for all $n \geq 0$,*

$$\frac{d^n}{dt^n} M_X\big|_{t=0} = \mathbb{E}[X^n]$$
$$\frac{d}{dt} \log M_X\big|_{t=0} = \mathbb{E}[X]$$
$$\frac{d^2}{dt^2} \log M_X\big|_{t=0} = \text{var}(X)$$

8 Types of Continuous Random Variables

8.1 Uniform

Definition 8.1 (Uniform random variable).

- Let $a < b$
- Pick a point X at random from the interval $[a, b]$
- If we have an equal probability of picking every point in $[a, b]$, we say X is uniformly distributed on the interval $[a, b]$
- We say $X \sim \text{Uniform}([a, b])$

Proposition 8.1. *If $a < b$ and $X \sim$ Uniform$([a, b])$, then it has PDF*

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

Proposition 8.2. *The CDF of $X \sim$ Uniform$([a, b])$ is $F_X(x) = \mathbb{P}(X \leq x)$, where*

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a < x < b \\ 1 & \text{if } x \geq 0 \end{cases}$$

Proposition 8.3 (Variance for Uniform). *Let $a < b$ and $X \sim$ Uniform$([a, b])$. Then,*

$$\text{var}(X) = \frac{b-a}{12}$$

8.2 Exponential

Definition 8.2. The exponential distribution

- Consider an approximate Poisson process with rate $\lambda > 0$ per unit time
- Let X be the time of the first arrival
- We say that X is exponentially distributed with mean waiting time $\theta = \frac{1}{\lambda}$ and write $X \sim \text{Exponential}(\theta)$

Proposition 8.4. *If $\theta > 0$ and $X \sim$ Exponential(θ), then it has PDF*

$$f_X(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$$

Proposition 8.5. *If $\theta > 0$ and $X \sim$ Exponential(θ), then its MGF is*

$$M_X(t) = \frac{1}{1 - \theta t} \quad \text{if } t < \frac{1}{\theta}$$

Proposition 8.6. *If $\theta > 0$ and $X \sim$ Exponential(θ), then it has mean and variance*

$$\mathbb{E}[X] = \theta$$
$$\text{var}(X) = \theta^2$$

8.3 Gamma

- Consider an approximate Poisson process with rate $\lambda > 0$ per unit time
- Let $\alpha \geq 1$ be an integer, and let X be the time of the αth arrival
- We say that X is gamma distributed with mean parameters α and $\theta = \frac{1}{\lambda}$ and write $X \sim \text{ Gamma}(\alpha, \theta)$
- If $X \sim \text{ Gamma}(1, \theta)$, then $X \sim \text{ Exponential}(\theta)$

If $\alpha \geq 1$ is an integer, $\theta > 0$ and $X \sim \text{ Gamma}(\alpha, \theta)$, X has PDF

$$f_X(x) = \frac{1}{\theta^\alpha (\alpha - 1)!} x^{\alpha - 1} e^{-\frac{x}{\theta}}$$

Its MGF, mean, and variance are

$$\begin{aligned} M_X(t) &= \frac{1}{(1 - \theta t)^\alpha} && \text{if } t < \frac{1}{\theta} \\ \mathbb{E}[X] &= \alpha\theta \\ \text{var}(X) &= \alpha\theta^2 \end{aligned}$$

8.4 Normal

Definition 8.3. We say a continuous random variable X is normally distributed with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ if it has PDF:

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

for $x \in \mathbb{R}$

Proposition 8.7 (MGF for Normal). *If $X \sim \mathcal{N}(\mu, \sigma^2)$, then it has MGF:*

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \text{ for any } t \in \mathbb{R}$$

Proposition 8.8. *If $X \sim \mathcal{N}(\mu, \sigma^2)$, then*

$$\begin{aligned} \mathbb{E}[X] &= \mu \\ \text{var}(X) &= \sigma^2 \end{aligned}$$

Definition 8.4. We define the function

$$F_X(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \, dx$$

Proposition 8.9. *If $x \geq 0$, then*

$$\Phi(-x) = 1 - \Phi(x)$$

9 Bivariate Probability

Let X, Y be a pair of discrete random variables taking values in sets $S_X, S_Y \subset \mathbb{R}$. We define their joint PMF by

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$$

We define the joint PMFs of X and Y to be

$$p_X(x) = \mathbb{P}(X = x) = \sum_{y \in S_Y} p_{X,Y}(x,y)$$
$$p_Y(y) = \mathbb{P}(Y = y) = \sum_{x \in S_X} p_{X,Y}(x,y)$$

We say that X, Y are independent if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y)$$

Let $S = S_X \times S_Y$. If $g : S \to \mathbb{R}$, the expected value of $g(X,Y)$ is

$$\mathbb{E}[g(X,Y)] = \sum_{(x,y) \in S} g(x,y)p_{X,Y}(x,y)$$

This is linear as well:

$$\mathbb{E}[ag(X,Y) + bh(X,Y)] = a\mathbb{E}[g(X,Y)] + b\mathbb{E}[h(X,Y)]$$

We also have that

$$\mathbb{E}[g(X)] = \sum_{x \in S_X} g(x)p_X(x)$$
$$\mathbb{E}[h(Y)] = \sum_{y \in S_Y} h(y)p_Y(y)$$

If X and Y are *independent* discrete random variables, we have

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$$

9.1 Cauchy-Schwarz Inequality

Let X, Y be discrete random variables. Then,

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$$

9.2 Covariance

Let X, Y be a pair of discrete random variables taking values. We define the covariance of X, Y to be

$$\begin{aligned}\operatorname{cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \sum_{(x,y) \in S} (x - \mathbb{E}[X])(y - \mathbb{E}[Y])p_{X,Y}(x,y)\end{aligned}$$

This reduces to:

$$\operatorname{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

If X is a random variable, then

$$\operatorname{cov}(X, X) = \operatorname{var}(X)$$

If X, Y are independent, then $\operatorname{cov}(X, Y) = 0$

If $a, b \in \mathbb{R}$,

$$\operatorname{cov}(aX, bY) = ab \operatorname{cov}(X, Y)$$

9.3 Correlation coefficient

For discrete random variables X, Y, we define the correlation coefficient of X, Y to be

$$\begin{aligned}\rho(X, Y) &= \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}} \\ &= \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}\end{aligned}$$

If X, Y are discrete random variables, then

$$-1 \leq \rho(X, Y) \leq 1$$

9.4 Conditional Distributions

Definition 9.1. Let X, Y be a pair of discrete random variables taking values in $S_X, S_Y \subseteq \mathbb{R}$, respectively.

- For each fixed $y \in S_Y$, we define the random variable $X|y$ with PMF $p_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y)$ for $x \in S_X$
- For each fixed $x \in S_X$, we define the random variable $Y|x$ with PMF $p_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x)$ for $y \in S_Y$

Proposition 9.1. *Let X, Y be a pair of discrete random variables taking values in $S_X, S_Y \subseteq \mathbb{R}$, respectively.*

- *For each fixed $y \in S_Y$, we have*
$$\sum_{x \in S_X} p_{X|Y}(x|y) = 1$$
- *For each fixed $x \in S_X$, we have,*
$$\sum_{y \in S_Y} p_{Y|X}(y|x) = 1$$

Definition 9.2. Let X, Y be a pair of discrete random variables taking values in $S_X, S_Y \subseteq \mathbb{R}$, respectively.

Define the function $g : S_X \to \mathbb{R}$ by

$$g(x) = \mathbb{E}[Y|x]$$

We define the conditional expectation of Y conditioned on X to be the random variable

$$\mathbb{E}[Y|X] = g(X)$$

Theorem 9.1 (The Law of Iterated Expectation)**.** *Let X, Y be discrete random variables. Then*

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$$

Definition 9.3. Let X, Y be a pair of discrete random variables taking values in $S_X, S_Y \subseteq \mathbb{R}$, respectively.

- Define the function $h : S_X \to \mathbb{R}$ by
$$h(x) = \text{var}\,(Y|x)$$
- We define the conditional variance of Y conditioned on X to be the random variable
$$\text{var}\,(Y|X) = h(X)$$

Theorem 9.2 (The Law of Total Variance)**.** *Let X, Y be discrete random variables. Then*
$$\mathbb{E}[\text{var}\,(Y|X)] + \text{var}\,(\mathbb{E}[Y|X]) = \text{var}\,(Y)$$

9.5 Bivariate Distributions of the Continuous Type

Proposition 9.2. *If X, Y are continuous random variables with joint PDF $f_{X,Y}(x, y)$, then*
$$\iint_{\mathbb{R}^2} f_{X,Y}(x, y)\, dx\, dy = 1$$

Definition 9.4. If X, Y are continuous random variables with joint PDF $f_{X,Y}(x, y)$, then

- the marginal PDF of X to be
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$
- the marginal PDF of Y to be
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$$

Definition 9.5 (Independence)**.** Let X, Y be continuous random variables with joint PDF $f_{X,Y}(x, y)$ and marginal PDFs $f_X(x)$, $f_Y(y)$.
We say that X, Y are independent if
$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$
for all $(x, y) \in \mathbb{R}^2$.

10 Several Random Variables

Definition 10.1. Let $X_1, X_2, \cdots, X_n$ be discrete random variables taking values in sets $S_1, S_2, \cdots, S_n \in \mathbb{R}$ and let $S = S_1 \times S_2 \times \cdots \times S_n \subseteq \mathbb{R}^n$. Then $X_1, X_2, \cdots, X_n$ have the joint PMF:
$$p_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = \mathbb{P}\,(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$$
These variables are independent if $p_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = p_{X_1}(x_1) p_{X_2}(x_2) \cdots p_{X_n}(x_n)$ for all $(x_1, x_2, \cdots, x_n) \in S$.

Proposition 10.1 (Linearity of Expectation)**.** *Let $X_1, X_2, \cdots, X_n$ be discrete or continuous random variables. Let $a_1, a_2, \cdots, a_n \in \mathbb{R}$ and let $Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$. Then*
$$\mathbb{E}[Y] = a_1 \mathbb{E}[X_1] + a_2 \mathbb{E}[X_2] + \cdots + a_n \mathbb{E}[X_n]$$

Definition 10.2 (Variance)**.** Let $X_1, X_2, \cdots, X_n$ be discrete or continuous random variables. Let $a_1, a_2, \cdots, a_n \in \mathbb{R}$ and let $Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$. Then
$$\text{var}\,(Y) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \,\text{cov}\,(X_j, X_k)$$

10.1 i.i.d Variables

Definition 10.3. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed. Then we define the sample sum

$$S_n = \sum_{j=1}^{n} X_j = X_1 + X_2 + \cdots + X_n$$

And sample average

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{n}S_n$$

Example 10.1. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed with mean μ and variance σ^2.

$$\mathbb{E}[S_n] = n\mu$$
$$\mathbb{E}[\bar{x}] = \mu$$
$$\operatorname{var}(S_n) = n\sigma^2$$
$$\operatorname{var}(\bar{x}) = \frac{\sigma^2}{n}$$

11 Transformation of Random Variables

11.1 Single Variable

Proposition 11.1. *Let X be a continuous random variable with PDF $f_X(x)$*

- *Let $X \subseteq \mathbb{R}$ so that $f_X(x) = 0$ for all $x \in \mathbb{R}\backslash S$*
- *Let $u : \mathbb{R} \to \mathbb{R}$ be smooth and satisfy $u\prime(x) > 0$ or $u\prime(x) < 0$ for all $x \in S$.*
- *Then $Y = u(Y)$ has PDF*

$$f_Y(y) = \left|\frac{d}{dx}u^{-1}(y)\right| \cdot f_X(u^{-1}(y))$$

Theorem 11.1 (Change of Variable)**.**

- *Let $S \subseteq \mathbb{R}^2$ and $f : S \to \mathbb{R}$ be continuous.*
- *Let $u : \mathbb{R}^2 \to \mathbb{R}^2$ be smooth and invertible and take $(z, w) = u(x, y)$*
- *If $v = u^{-1}$, so that $(x, y) = v(z, w)$, then*

$$\iint_S f(x, y)\, dx\, dy = \iint_{u(S)} f(v(z, w)) \left|\frac{\partial(x, y)}{\partial(z, w)}\right| dz\, dw$$

11.2 Double Variable

Proposition 11.2.

- *Let X, Y be continuous random variables with joint PDF $f_{X,Y}(x, y)$*
- *Let $u : \mathbb{R}^2 \to \mathbb{R}^2$ be smooth and invertible, with inverse $v(z, w)$*
- *Then, the random variables $(Z, W) = u(X, Y)$ are continuous and have joint PDF*

$$f_{Z,W}(z, w) = f_{X,Y}(z(z, w)) \left|\frac{\partial(x, y)}{\partial(z, w)}\right|$$

12 The Law of Large Numbers and Convergence

Proposition 12.1 (Markov's inequality)**.** *Let X be a non-negative random variable. Then, given $\lambda > 0$, we have*

$$\mathbb{P}\left(X \geq \lambda\right) \leq \frac{\mathbb{E}[X]}{\lambda}$$

Proposition 12.2 (Generalized Markov's inequality)**.** *Let X be a non-negative random variable. Then, given $\lambda > 0$ and integer $k \geq 1$, we have*

$$\mathbb{P}\left(X \geq \lambda\right) \leq \frac{\mathbb{E}[X^k]}{\lambda^k}$$

Proposition 12.3 (Chebyshev's inequality)**.** *Let X be a random variable with mean μ and variance σ^2. Then, given $\lambda > 0$, we have*

$$\mathbb{P}\left(|X - \mu| \geq \lambda\right) \leq \frac{\sigma^2}{\lambda^2}$$

Proposition 12.4 (Chernoff bound)**.** *Let X be a random variable. Then, given $\lambda > 0$, we have*

$$\mathbb{P}\left(X \geq \lambda\right) \leq \inf_{t>0}(e^{-t\lambda} M_X(t))$$

Theorem 12.1 (The Weak Law of Large Numbers)**.** *Let $X_1, X_2, \cdots$ be an i.i.d. sequence of random variables with finite mean μ. Then,*

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \text{ as } n \to \infty$$

12.1 The MGF Technique

Proposition 12.5. *Let $X_1, X_2, \cdots, X_n$ be a sequence of independent random variables and let $a_1, a_2, \cdots, a_n \in \mathbb{R}$. Then the random*

$$Y = \sum_{j=1}^{n} a_j X_j$$

has MGF

$$M_Y(t) = \prod_{j=1}^{n} M_{X_j}(a_j t)$$

whenever it is well-defined.

Proposition 12.6. *Let X, Y be continuous random variables with MGFs $M_X(t)$ and $M_Y(t)$. Suppose that for some $h > 0$ and all $t \in (-h, h)$, we have*

$$M_X(t) = M_Y(t)$$

Then, X and Y are identically distributed.

Proposition 12.7. *Let $X_1, X_2, \cdots, X_n$ be an i.i.d. sequence of random variables with common MGF $M(t)$. Then*

$$M_{S_n}(t) = [M(t)]^n$$
$$M_{\bar{X}}(t) = \left[M(\frac{t}{n})\right]^n$$

Proposition 12.8 (Limiting MGF determines the distribution)**.** *Let* $X_1, X_2, \cdots, X_n$ *and* X *be random variables. Suppose that for some* $h > 0$ *and all* $t \in (-h, h)$*, we have*

$$M_{X_n}(t) \to M_X(t) \text{ as } n \to \infty$$

Then, $X_n \to X$ *in distribution as* $n \to \infty$*.*

Theorem 12.2 (The Central Limit Theorem)**.** *Let* $X_1, X_2, \cdots, X_n$ *be an i.i.d. sequence of random variables with finite mean* μ *and variance* σ^2*. Then,*

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \to \sim \mathcal{N}(0, 1) \text{ in distribution as } n \to \infty$$

Proposition 12.9 (DeMoivre-Laplace Correction)**.** *Let* $X \sim$ *Binomial*(n, p)*. Then, we have*

$$\mathbb{P}(X \leq \ell) \approx \Phi\left(\frac{\ell + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

$$\mathbb{P}(X \geq k) \approx 1 - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

$$\mathbb{P}(k \leq X \leq \ell) \approx \Phi\left(\frac{\ell + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$