Introduction to probability
Lecture Notes

George Z. Miao

2023
Fall

Chapter 1

1.2 Sample Space

Definition 1.1.

*(1) An **Experiment** is a process that with a set of possible outcomes.*

*(2) **Sample Space** S is the set of possible outcomes.*

(3) An event is a set $E \subseteq S$

Ex 1.1. *Roll dice, $S = \{1, 2, 3, 4, 5, 6\}$, $E = \{1, 2, 3\}$, event an event number is rolled.*

Ex 1.2. *Roll dice four times. Note sequence if numbers rolled.*

$$S = \{(a_1, a_2, a_3, a_4) \mid 1 \leq a_i \leq 6\} \text{ where } a_i \text{ is } i^{th} \text{ roll}$$
$$|S| = 6^4$$

Note An event $E \subseteq S$ **occurred** if the outcome is in E.

1.3 Naive Definition of Probability

Definition 1.2. *A sample space S is **Simple** if*

1. *$|S|$ is finite, and*
2. *All outcomes are equally likely*

Definition 1.3 (Probability)**.** *If S is a simple sample space, and $E \subseteq S$, then the probability of E is*

$$P(E) = \frac{|E|}{|S|}$$

1.4 How to count

Theorem 1.1 (Multiplication Rule)**.** *Assume an experiment is performed in 2 steps where*

1. *Step A can be completed a ways*
2. *Step B can be completed b ways*

Then the total number of outcomes the experiment is ab.

Theorem 1.2 (Sampling without replacement)**.** *Assume we have n objects, and we want to choose k of them in order. Then the number of ways to do so is*

$$n(n-1)(n-2)\dots(n-k+1)$$

Definition 1.4 (Permutations). *n ordered sample without replacement is called a permutation (n objects taken k at a time). Number of such permutation is denoted with*

$$P(n,k) = \frac{n!}{(n-k)!}$$

Ex 1.3 (Birthday Problem). *If $k < 365$ people are in a room, what is the probability that at least two of them have the same birthday? Assume birthdays are evenly distributed and ignore leap years.*

Answer. Let A be the event 2 people have a birthday in common.

$$\begin{aligned}
|S| &= 365^k \\
P(A) &= \frac{|A|}{|S|} \\
&= \frac{365^k - \left|A^{\complement}\right|}{|S|} \\
&= \frac{365^k - P(365,k)}{|S|} \\
&= 1 - \frac{P(365,k)}{|S|} \\
&\approx .506 \text{ if } k = 23
\end{aligned}$$

Definition 1.5 (Combination). *A **combination** of n objects taken k at a time $(0 \leq k \leq n)$ is an **unordered** selection of k of the objects.*

Theorem 1.3. *The number of combinations is $C(n,k) = \binom{n}{k} = \frac{n!}{(n-k)!k!}$*

Ex 1.4 (Sampling with repition, but order does not matter). *Assume a bakery has k types of cookies, how many ways to choose n if order does not matter?*

Answer. Number of ways to put the cookies is $\binom{n+k-1}{k-1} = \binom{n+k-1}{n}$

Summary: Sampling Methods

1. *Ordered with repetition:* n^k
2. *Ordered without repetition:* $P(n,k) = \frac{n!}{(n-k)!}$
3. *Unordered without repetition:* $C(n,k) = \binom{n}{k} = \frac{n!}{(n-k)!k!}$
4. *Unordered with repetition:* $\binom{n+k-1}{n}$

Theorem 1.4 (Binomial Theorem).

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

Theorem 1.5 (Pascal's Triangle Identity).

$$\binom{k}{i-1} + \binom{k}{i} = \binom{k+1}{i}$$

1.6 Formal Definition of Probability

Definition 1.6 (Probability Space)**.** *A* ***Probability Space*** *consists of a sample space S, a probability function P that assigns a real number $P(A)$ to each $A \subseteq S$ such that*

- $P(A) \in [0, 1]$
- $P(\varnothing) = 0, P(S) = 1$
- *If $A_1, A_2, A_3, \dots$ is an infinite sequence with $A_i \cap A_j = \varnothing$ if $i \neq j$, then*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Theorem 1.6 (Property of Probability)**.**

- *If $A_1, A_2, A_3, \dots, A_n$ are events with $A_i \cap A_j = \varnothing$ if $i \neq j$, then*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

- $P\left(A^{\complement}\right) = 1 - P(A)$
- *If $A \subseteq B$, then $P(A) \leq P(B)$*
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ***(Inclusion-Exclusion)***
- $P(A) = P(A \cap B) + P(A \cap B^{\complement})$

Remark 1.1. *Inclusion-Exclusion can be written as $P(A \cap B) = P(A) + P(B) - P(A \cup B)$*

Ex 1.5. *Roll 2 dice and compute sum*

1. *What is $P(\{4\}), P(\{5\})$*
2. *Make a table of probability*

Definition 1.7. *When $S = \{S_1, S_2, \dots, S_n\}$, write $P_i = P(\{S_i\}) = P(S_i)$*

Theorem 1.7. *Let $A_1, A_2, \dots, A_n$ be the events in P.S. S. Then*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq 1} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq 1} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Recall

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$
$$e^{-1} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}$$

By Alternating Series Test, it converges

Ex 1.6. *8 firends including Amy and Brad sit at a 8 seat round table. What is the probability that Amy and Brad sit next to each other?*

a) Keep track of Amy's seat and Brad's seat.

   (1) Choose Amy's seat: 8 ways

   (2) Choose Brad's seat: 2 ways

   Total: 56 ways, so $\frac{16}{56} = \frac{2}{7}$.

b) Keep track of seats A and B. $|S| = C(8, 2) = 28$ E be seats that are together, $|E| = 8$, so $\frac{8}{28} = \frac{2}{7}$.

c) Amy has 2 neighbors out of 7 people, so $\frac{2}{7}$.

Chapter 2

Conditional Probability

2.2 Formal Definition

Definition 2.1 (Conditional Probability). *Let A and B be events in a probability space S and $P(B) > 0$. The* ***conditional probability*** *of A given B is*

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

Ex 2.1. *$S = \{1, 2, 3, 4, 5, 6\}$, $A = \{x \mid x \geq 4\} \subseteq S$, $E = \{2, 4, 6\} \subseteq S$. Then $P(A \mid E) = \frac{P(A \cap E)}{P(E)} = \frac{P(\{4,6\})}{P(\{2,4,6\})} = \frac{2}{3}$*

Ex 2.2. *Family have at least 1 girl, what is the probability both children are girls?*

Answer. $S = \{GG, GB, BG, BB\}$, B is the event at least one girl, so $B = \{GG, GB, BG\}$, E is the event of 2 girls, so $E = \{GG\}$, and

$$P(E \mid B) = \frac{P(E \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Ex 2.3. *Show the probability of $P(A^{\complement} \mid B) = 1 - P(A \mid B)$ if $P(B) \neq 0$.*

Proof.

$$\begin{aligned} P(A^{\complement} \mid B) &= \frac{P(A^{\complement} \cap B)}{P(B)} \\ &= \frac{P(B) - P(A \cap B)}{P(B)} \\ &= 1 - \frac{P(A \cap B)}{P(B)} \\ &= 1 - P(A \mid B) \end{aligned}$$

□

Ex 2.4. *Envelope contains 3 cards. 2 are green on both sides and 1 is green on one side and red on the other. Pick a card at random and see one side is green. What is the probability the other side is green?*

Answer.

$$\begin{aligned}
S &= \{(1,G_1),(1,G_2),(2,G_1),(2,G_2),(3,G_1),(3,R_2)\} && \text{(Sample Space)}\\
A &= \{(1,G_1),(1,G_2),(2,G_1),(2,G_2)\} && \text{(Both sides are green)}\\
B &= \{(1,G_1),(1,G_2),(2,G_1),(2,G_2),(3,G_1)\} && \text{(One side is green)}
\end{aligned}$$

$$\begin{aligned}
P(A \mid B) &= \frac{P(A\cap B)}{P(B)}\\
&= \frac{P(A)}{P(B)}\\
&= \frac{4/6}{5/6}
\end{aligned}$$

Ex 2.5. *100 student surveyed, 27 are juniors who sleep late, 19 are juniors who attend class, 31 are seniors who sleep late and 23 are seniors who attend class. Pick one student at random, then 1) what's the probability of a senior, 2) what's the probability that they prefer sleeping and 3) given that they prefer sleeping, what's the probability they're a senior.*

Answer.

	Sleep	Class
Junior	27	19
Senior	31	23

1. $31+23=54$, so 0.54.
2. $27+31=58$, so 0.58.
3. Let A be senior and B be prefer sleeping, then $P(A\cap B)=0.31$, $P(A \mid B)=\frac{P(A\cap B)}{P(B)}=\frac{0.31}{0.58}=\frac{31}{58}$.

Ex 2.6. $A,B \subseteq S$, *P.S.,* $P(A)=.6, P(B)=.45, P(A\cup B)=.9$, *find* $P(A\cap B)$ *and* $P(B \mid A)$.

Answer.

$$\begin{aligned}
P(A\cup B) &= P(A)+P(B)-P(A\cap B) && \text{(Inc-Exc)}\\
.9 &= .6+.45-P(A\cap B)\\
P(A\cap B) &= .15\\
P(A \mid B) &= \frac{P(A\cap B)}{P(B)}\\
&= \frac{.15}{.45}\\
&= \frac{1}{3}\\
P(B \mid A) &= \frac{P(A\cap B)}{P(A)}\\
&= \frac{.15}{.6}\\
&= \frac{1}{4}
\end{aligned}$$

2.3 Baye's Rule and the Law of Total Probability

Theorem 2.1. *Let* $A_1,A_2,\dots,A_n \subseteq S$, *P.S., with* $P\left(\bigcap_{i=1}^{n-1} A_i\right) \neq 0$, *then*

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1\cap A_2)\dots P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right)$$

Ex 2.7. *20 balls with 12 red and 8 green. Draw balls one at a time without replacement.*

1. Probability 1 at least green: $\frac{8}{20}\frac{12}{19}\frac{7}{18}$
2. Probability 1st drawn in green: $\frac{8}{20}$
3. Probability 12^{th} drawn in green: $\frac{8}{20}$

Theorem 2.2 (Baye's Rule, Version 1)**.** *Assume $A, B \subseteq S$, $P(A) > 0, P(B) > 0$, then*

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Definition 2.2 (Partition)**.** *Let S be an sample space, a **partition** of S is a set of events*

$$\{A_1, A_2, \dots, A_n\} \quad \textit{such that}$$

1. $\bigcup_{i=1}^n A_i = S$
2. $A_i \cap A_j = \varnothing$ *if* $i \neq j$
3. $P(A_i) > 0$

Theorem 2.3 (L.O.T.P.)**.** *Let $A_1, \dots, A_n$ be a partiton of p.s. S and $B \subseteq S$ be any event, then*

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

Theorem 2.4 (Baye's Rule, Version 2)**.** *Assume $A_1, \dots, A_n$ in a partition of S p.s,. and $B \subseteq S$, then*

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^n P(B \mid A_j)P(A_j)}$$

Proof.

$$\begin{aligned} P(A_i \mid B) &= \frac{P(B \mid A_i)P(A_i)}{P(B)} && \text{(Baye's I)} \\ &= \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^n P(B \mid A_j)P(A_j)} && \text{(L.O.T.P.)} \end{aligned}$$

□

Ex 2.8. *One has 2 cookie jars, 1^{st} is green with 10 chocolate chip cookies, and 14 ginger cookies. 2^{nd} is red with 5 chocolate chip cookies and 20 ginger cookies. Pick green jar with probability $2/3$ and take random cookie. Pick red jar with probability $1/3$ and take a random cookie. Then:*

1. *What is the probability of a chocolate chip cookie?*
2. *If a chocolate chip cookie is picked, what is the probability it came from the green jar?*

Answer.

$$\begin{array}{lll} A_1 & \text{green jar chosen} & P(A_1) = \dfrac{2}{3} \\ A_2 & \text{red jar chosen} & P(A_2) = \dfrac{1}{3} \\ B & \text{chocolate chip cookie} & P(B \mid A_1) = \dfrac{10}{24},\ P(B \mid A_2) = \dfrac{5}{25} \end{array}$$

1. What is the probability of a chocolate chip cookie?

$$\begin{aligned} P(B) &= P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) \\ &= \frac{10}{24} \cdot \frac{2}{3} + \frac{5}{25} \cdot \frac{1}{3} \end{aligned}$$

2. If I got a c.c.c., what is the probabilities it came from green jar?

$$\begin{aligned} P(A_1 \mid B) &= \frac{P(B \mid A_1)P(A_1)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2)} \\ &= \frac{\frac{10}{24} \cdot \frac{2}{3}}{\frac{10}{24} \cdot \frac{2}{3} + \frac{5}{25} \cdot \frac{1}{3}} > \frac{2}{3} \end{aligned}$$

Definition 2.3. *$P(A_1), \ldots, P(A_n)$ are **prior** probabilities. $P(A_1 \mid B), \ldots, P(A_n \mid B)$ are **posterior** probabilities.*

2.4 Independent Events

Definition 2.4. *Let S be the probability space, $A, B \subseteq S$. Then A and B are **independent** if*

$$P(A \cap B) = P(A)P(B)$$

Caution 2.1.

1. *Independent is **NOT** disjoint.*
2. *$P(A \cap B) \neq P(A)P(B)$ unless they're independent.*

Theorem 2.5. *Assume A and B are independent, and $P(B) \neq 0$, then*

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

Ex 2.9. *If A and B are independent, and $P(A) = 0.6, P(B) = 0.3$, then*

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A)P(B) \\ &= 0.6 + 0.3 - 0.18 \\ &= 0.72 \end{aligned}$$

Theorem 2.6. *If A and B are independent, $A^{\complement}$ and B are independent.*

Proof.

$$\begin{aligned} P(A^{\complement} \cap B) &= P(B) - P(A \cap B) \\ &= P(B) - P(A)P(B) \\ &= [1 - P(A)]P(B) \\ &= P(A^{\complement})P(B) \\ &\Rightarrow A^{\complement} \text{ and } B \text{ are independent} \end{aligned}$$

$\square$

Definition 2.5. *Let $A_1, A_2, \ldots, A_n \subseteq S$, P.S., then $A_1, A_2, \ldots, A_n$ is an **independent collection** if for any $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, then*

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j})$$

Ex 2.10.

$$\begin{aligned}
S &= \{1, 2, \ldots, 12\} \\
A &= \{3, 4, 6, 8, 10, 12\} \quad P(A) = \frac{6}{12} = \frac{1}{2} \\
B &= \{3, 6, 9, 12\} \quad P(B) = \frac{4}{12} = \frac{1}{3} \\
C &= \{1, 5, 6, 7, 8, 12\} \quad P(C) = \frac{6}{12} = \frac{1}{2} \\
P(A \cap B) &= P(\{6, 12\}) = \frac{1}{6} = P(A)P(B) \\
P(A \cap C) &= P(\{6, 8, 12\}) = \frac{1}{4} = P(A)P(C) \\
P(B \cap C) &= P(\{6, 12\}) = \frac{1}{6} = P(B)P(C) \\
P(A \cap B \cap C) &= P(\{6, 12\}) = \frac{1}{6} \neq P(A)P(B)P(C)
\end{aligned}$$

So A, B, C is not an independent collection.

Ex 2.11. *A, B, C independent, $P(A) = .6, P(B) = .3, P(C) = .2$, find $P(A \cup B \cup C)$*

Answer.

$$\begin{aligned}
P(A \cup B \cup C) &= P(A) + P(B) + P(C) - [P(A \cap B) + P(A \cap C) + P(B \cap C)] + P(A \cap B \cap C) \\
P(A \cup B \cup C) &= P(A) + P(B) + P(C) - [P(A)P(B) + P(A)P(C) + P(B)P(C)] + P(A \cap B \cap C) \\
&= .6 + .3 + .2 - [.18 + .12 + .06] + .036
\end{aligned}$$

Definition 2.6 (Conditional Independence)**.** *$A, B, E \subseteq S$, $P(E) > 0$, A and B are **independent given** E if*

$$P(A \cap B \mid E) = P(A \mid E)P(B \mid E)$$

Caution 2.2.

- *A and B are independent does **NOT** imply A and B are independent given E.*
- *A and B are independent given E does **NOT** imply A and B are independent.*

Ex 2.12. *Let $S = \{1, 2, 3, \ldots, 12\}$ be a simple P.S., and*

$$\begin{aligned}
A &= \{2, 4, 6, 7, 9\} \\
B &= \{3, 6, 7, 11\} \\
P(A) &= \frac{|A|}{|S|} = \frac{5}{12} \\
P(B) &= \frac{4}{12} \\
A \cap B &= \{3, 6\} \\
P(A \cap B) &= \frac{2}{12} = \frac{1}{6} \neq P(A)P(B) = \frac{5}{36}
\end{aligned}$$

So A and B are not independent. Now define

$$
\begin{aligned}
E &= \{1,2,3,4,5,6\} \\
P(A \mid E) &= \frac{P(A \cap E)}{P(E)} = \frac{\frac{3}{12}}{\frac{1}{2}} = \frac{1}{2} \\
P(B \mid E) &= \frac{P(B \cap E)}{P(E)} = \frac{\frac{2}{12}}{\frac{1}{2}} = \frac{1}{3} \\
P(A \cap B \mid E) &= \frac{P(A \cap B \cap E)}{P(E)} = \frac{\frac{1}{12}}{\frac{1}{2}} = \frac{1}{6} \\
&= P(A \mid E) P(B \mid E)
\end{aligned}
$$

So A and B are independent given E.

Chapter 3

Random Variables and their distributions

3.1 Random Variables

Definition 3.1. *Let S be a probability space. A **random variable** is a function $X : S \to \mathbb{R}$.*

Ex 3.1. *Role 3 dices, outcome are (g, r, y) where g is green role, r is red role, and y is yellow role. Then $|S| = 6^3$. Now let $X(g, r, y) = g + r + y$.*

$$\begin{aligned} P(X = x) &= P(\{(g, r, y) \mid g + r + y = x\}) \\ P(X = 4) &= P(\{(2, 1, 1), (1, 2, 1), (1, 1, 2)\}) \\ &= \frac{3}{6^3} = \frac{1}{72} \end{aligned}$$

Definition 3.2 (Discrete Random Variable)**.** *A R.V. X that takes values in a sequence of numbers $x_1, x_2, \ldots$ (finite or infinite) is called a **discrete random variable**.*

Definition 3.3 (Continuous Random Variable)**.** *A R.V. X that takes any value in some interval $I = (a, b)$ where $b > a$ is called a **continuous random variable**.*

Ex 3.2. *Toss a coin until a head is tossed. Let X be the number of tosses until a head is tossed. X takes values in $1, 2, 3, \ldots$, so X is a discrete random variable.*

$$\begin{gathered} S = \{H, TH, TTH, \ldots\} \\ P(X = x) = P\left(\underbrace{TT \ldots T}_{x-1} H\right) = \left(\frac{1}{2}\right)^x \end{gathered}$$

3.2 Probability Mass Function

Definition 3.4 (Probability Mass Function)**.** *The* ***probability mass function*** *of a discrete random variable X is the function $P_X(x) = P(X = x)$, and $P_X(x) = 0$ if x not in the sequence of values of X.*

Theorem 3.1. *If X is discrete,*

$$P_X(x) = P(\{s \in S \mid X(s) = x\})$$

Ex 3.3. *In example rolling 3 dices*

$$\begin{aligned} P_X(3) &= \frac{1}{6^3} \\ P_X(4) &= \frac{3}{6^3} \\ P_X(5) &= P(\{(3,1,1),(2,2,1),(2,1,2),(1,3,1),(1,2,2),(1,1,3)\}) \\ &= \frac{1}{36} \end{aligned}$$

Definition 3.5 (Support)**.** *X is a discrete R.V. with p.m.f. $P_X(x)$. The* ***support*** *of X is $\{x \mid P_X(x) \neq 0\}$.*

Definition 3.6 (Valid Probability Mass Function)**.** *$P_X(x)$ is a function with support $x_1, x_2, \dots$ s.t.*

- *$P_X(x) \geq 0$ for all x*
- *$\sum_x P_X(x) = 1$*

Then $P_X(x)$ is a valid P.M.F.

Ex 3.4. *Roll 2 dices X given total of rolls*

Answer.

$$\begin{aligned} \sum_x P_X(x) &= 2\left(\frac{1}{36}\right) + 2\left(\frac{2}{36}\right) + 2\left(\frac{3}{36}\right) + 2\left(\frac{4}{36}\right) + 2\left(\frac{4}{36}\right) + \frac{2}{36} \\ &= \frac{2+4+6+8+10+6}{35} \\ &= \frac{36}{36} = 1 \end{aligned}$$

Ex 3.5. *Roll a dice until a 6 is rolled. X counts number of rolls before a 6 is rolled.*

Answer.

$$\begin{aligned} X\left(\underbrace{?????}_{x}6\right) &= x \\ P_X(x) &= \left(\frac{5}{6}\right)^x \frac{1}{6} \text{ if } x = 0,1,2,3,\dots \\ \sum_{x=0}^{\infty} P_X(x) &= \sum_{x=0}^{\infty} \left(\frac{5}{6}\right)^x \frac{1}{6} \\ &= \frac{\frac{1}{6}}{1-\frac{5}{6}} = 1 \\ P(2 < X \leq 5) &= P(X \in \{3,4,5\}) \\ &= P_X(3) + P_X(4) + P_X(5) \\ &= \left(\frac{5}{6}\right)^3 \frac{1}{6} + \left(\frac{5}{6}\right)^4 \frac{1}{6} + \left(\frac{5}{6}\right)^5 \frac{1}{6} \end{aligned}$$

Definition 3.7 (Geometric Series)**.**

$$\sum_{i=0}^{\infty} ar^i = \frac{a}{1-r}$$

3.3 Bernoulli and Binomial R.V.

Definition 3.8. *A random variable with p.m.f.*

$$0 < p < 1, P_X(x) = \begin{cases} p & \textit{if } x = 1 \\ q = 1 - P & \textit{if } x = 0 \\ 0 & \textit{otherwise} \end{cases}$$

is called a ***Bernoulli random variable****, notation* $X \sim \mathit{Bern}\,(p)$.

Ex 3.6. *S P.S.,* $A \subseteq S$*,* $X : S \to \mathbb{R}$

Definition 3.9. *A R.V.* X *has a* ***Binomial distribution*** *if p.m.f. is*

$$P_X(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \textit{if } x = 0, 1, \ldots, n \\ 0 & \textit{otherwise} \end{cases}$$

$$\begin{aligned} \sum_{x=0}^{n} P_X(x) &= \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} \\ &= (p+q)^n \\ &= 1 \end{aligned}$$

Write $X \sim \mathit{Bin}\,(n, p)$

Ex 3.7. *I have 8 chickens. Every day each lays egg with probability p,* $0 < p < 1$*. Outcome* $NYYNNYNY$ *is an 8 letter word in* $\{Y, N\}$*.* X *is number of eggs laid in a day.*

$$P(NYYNNYNY) = p^4 q^4$$

X *is R.V. counts number of eggs.* X *can take values* $0, 1, 2, \ldots, 8$.

$$\begin{aligned} P_X(2) &= \binom{8}{2} p^2 q^{8-2} \\ P_X(x) &= \binom{8}{x} p^x q^{8-x} \\ X &\sim \mathit{Bin}\,(8, p) \end{aligned}$$

Now define

$$\begin{aligned} X_1 &= \begin{cases} 1 & \textit{if chicken 1 lays egg} \\ 0 & \textit{otherwise} \end{cases} \\ X_2 &= \begin{cases} 1 & \textit{if chicken 2 lays egg} \\ 0 & \textit{otherwise} \end{cases} \\ &\vdots \\ X_8 &= \begin{cases} 1 & \textit{if chicken } 8 \textit{ lays egg} \\ 0 & \textit{otherwise} \end{cases} \\ X_i &\sim \mathit{Bern}\,(P) \\ X &= \sum_{i=1}^{8} X_i \end{aligned}$$

Definition 3.10. *A Bernoulli Trial is an experiment with 2 outcomes, "S" success and "F" failure, which gives R.V.*

$$X = \begin{cases} 1 & \text{if } S \\ 0 & \text{if } F \end{cases}$$

If 1 have n independent Bernoulli trials, X_i is the R.V. for i^{th} trial, then

$$X_1 + X_2 + \cdots + X_n \sim \mathit{Bin}\,(n, p)$$

3.4 Hypergeometric Distribution

Ex 3.8. *Box containing g green and r red balls. Pick n balls at random without repetition.*

$$|S| = \binom{g+r}{n}$$

X counts number of green balls.

$$P(X = k) = \frac{\binom{g}{k}\binom{r}{n-k}}{\binom{g+r}{n}}, \quad 0 \le k \le n$$

Definition 3.11. *If X has p.m.f. f, $0 \le k \le n$, and*

$$f_X(k) = \frac{\binom{g}{k}\binom{r}{n-k}}{\binom{g+r}{n}}$$

*then X has a **hypergeometric distribution**, notation $X \sim$ HGeom (g, r, n).*

Ex 3.9. *X is the number of aces in a poker hand. $X \sim$ HGeom $(4, 48, 5)$.*

$$P_X(2) = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}}$$

3.5 Discrete Uniform

Definition 3.12. *$C = \{x_1, x_2, \ldots, x_n\}$ is the set of n real numbers. X is R.V. with p.m.f.*

$$\begin{aligned} P_X(x) &= \frac{1}{n}, \quad x = x_1, x_2, \ldots, x_n \\ P_X(x) &\ge 0 \\ \sum_x P_X(x) &= \sum_{i=1}^{n} P_X(x_i) \\ &= \sum_{i=1}^{n} \frac{1}{n} \\ &= 1 \end{aligned}$$

Writen as

$$\begin{aligned} X &\sim \mathit{DUnif}\,(C) \\ X &\sim \mathit{DUnif}\,(x_1, x_2, \ldots, x_n) \end{aligned}$$

Theorem 3.2. *If $X \sim$ Bern $\left(\frac{1}{2}\right)$, then $X \sim$ DUnif $(0, 1)$*

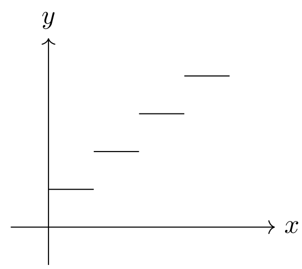
3.6 Cumulative Distribution Functions

Definition 3.13. *If X is a R.V., its **cumulative distribution function (C.D.F.)** is*

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

Ex 3.10.

$$X \sim \textit{Bin}\,(3, .4)$$
$$P_X(x) = \begin{cases} .214 & \textit{if } x = 0 \\ .432 & \textit{if } x = 1 \\ .288 & \textit{if } x = 2 \\ .064 & \textit{if } x = 3 \end{cases}$$
$$\begin{aligned} F(2,3) &= P(X \leq 2.3) \\ &= P_X(0) + P_X(1) + P_X(2) \\ &\simeq .936 \end{aligned}$$



Ex 3.11. *Box contains 50 balls, numberred 1 to 50. Pick 7 at random, one at a time without replacement. X is the number of balls where number is divisible by 7.*

$$X_i = \begin{cases} 1 & \textit{if ball } i \textit{ is divisible by 7} \\ 0 & \textit{otherwise} \end{cases}$$
$$X = X_1 + X_2 + \cdots + X_7$$

(a) *What is distribution of X_i?*

$$P_{X_i}(x) = \begin{cases} \frac{7}{50} & \textit{if } x = 1 \\ \frac{43}{50} & \textit{if } x = 0 \end{cases}$$
$$X_i \sim \textit{Bern}\left(\frac{7}{50}\right)$$

(b) *What is distribution of X?*

$$P(x) = \frac{\binom{7}{x}\binom{43}{7-x}}{\binom{50}{7}} \quad \textit{if } x = 0, 1, 2, \ldots, 7$$

(c)

$$P(X_2 = 1 \mid X_1 = 1) = \frac{6}{49} \neq P(X_2 = 1) = \frac{7}{50}$$
$$P(X_2 = 1 \mid X_1 = 0) = \frac{7}{49} \neq P(X_2 = 1) = \frac{7}{50}$$

Value $X_1, X_2, \ldots, X_7$ take are not "independent" of each other!

Theorem 3.3 (Properties of C.D.F.). *Let X be a R.V. with C.D.F. $F_X(x)$.*

1. *If $X_1 < X_2$ then $F_X(x_1) \leq F_X(x_2)$*

2. $\lim_{x \to a^+} F(x) = F(a)$ *(Right Continuous)*

3. $\lim_{x \to a^-} F(x) = P(x < a)$

4. $\lim_{x \to \infty} F(x) = 1, \lim_{x \to -\infty} F(x) = 0$

5. *If $a < b$ then*

$$F(b) - F(a) = P(a < X \leq b)$$

Proof.

1.

$$\begin{aligned} F_X(x_2) &= P(X \leq x_2) \\ &= P(X \in (-\infty, x_1] \cup (x_1, x_2]) \\ &= P(X \in (-\infty, x_1]) + P(X \in (x_1, x_2]) \\ &\geq F_X(x_1) + 0 \end{aligned}$$

2. clear form the graph of $y = F_X(x)$

3. graph

4. Follows from def of Prob. Space

5. $P(X \leq h) = P(X \leq a) + P(a < x \leq h)$

□

3.7 Functions of a R.V.

Definition 3.14. *If X is a R.V. defined on P.S. S, and $g : \mathbb{R} \to \mathbb{R}$ is defined at every $X(s)$ where $s \in S$. Then a new R.V. $Y = g(X)$ is defined on S by $y(s) = g(X(s))$.*

Ex 3.12.

$$X \sim \textit{DUnif}\,(\{1,2,3,4,5,6\})$$
$$g(x) = (x-3)^3$$

x	$P_X(x)$	$g(x)$
1	1/6	4
2	1/6	1
3	1/6	0
4	1/6	1
5	1/6	4
6	1/6	9

$$\begin{aligned} P_Y(4) &= P(Y=4) \\ &= P(x \in \{1,5\}) \\ &= P_X(1) + P_X(5) \\ &= \frac{1}{3} \end{aligned}$$

y	$P_Y(y)$
0	1/6
1	2/6
4	2/6
9	1/6

$$P_Y(y) = \begin{cases} 1/6 & y = 0, 9 \\ 1/3 & y = 1, 4 \end{cases}$$

Ex 3.13 (Random Walk)**.** *Start at 0 on real line. Every second, jump to the right 1 unit with probability p and jump to the left with probability $q = 1-p$. X is the number of jumps to the right and Y is the final position after n seconds.*

$$\begin{aligned} Y &= X - (n - X) \\ &= 2X - n \\ X &\sim \textit{Bin}\,(n,p) \\ P_Y(y) &= P(Y=y) \\ &= P(2x - n = y) \\ &= P(X = \frac{y+n}{2}) \\ &= \binom{n}{\frac{y+n}{2}} p^{\frac{y+n}{2}} q^{\frac{n-y}{2}} && (\text{If } \tfrac{n-y}{2} = 0, 1, 2, \ldots, n) \end{aligned}$$

Theorem 3.4. *Let $Y = g(X)$ be R.V. defined on P.S. S. If X is discrete with p.m.f. $P_X(x)$, then Y is discrete with p.m.f.*

$$P_Y(y) = \sum_{x \in g^{-1}(y)} P_X(x)$$

Theorem 3.5. *If X is a discrete R.V. with p.m.f. $P_X(x)$ then*

1. $F_X(x) = \sum_{t \le x} P_X(t)$
2. $P_X(x) = P(X = x) = F_X(x) - \lim_{t \to x^-} F(t)$

Ex 3.14. *Roll a red dice and a blue dice.*

$$S = \{(r,b) \mid 1 \le r \le 6, 1 \le b \le 6\}$$

Define:

$$
\begin{aligned}
X : S &\longrightarrow \mathbb{R} \\
(r, b) &\longmapsto r \\
Y : S &\longrightarrow \mathbb{R} \\
(r, b) &\longmapsto b \\
T : S &\longrightarrow \mathbb{R} \\
(r, b) &\longmapsto r + b \\
g(x, y) &= x + y \\
Z : S &\longrightarrow \mathbb{R} \\
s &\longmapsto g(X(s), Y(s))
\end{aligned}
$$

Ex 3.15. *In the last example, let $g(x, y) = \min(x, y)$, find p.m.f. $P_M(m)$ where M can be $1, 2, 3, 4, 5, 6$.*

Answer.

$$
\begin{aligned}
P_M(1) &= P(M = 1) = P(\{1,1\}, \{1,2\}, \ldots, \{1,6\}, \{2,1\}, \{3,1\}, \ldots, \{6,1\}) = \frac{6+5}{36} = \frac{11}{36} \\
P_M(2) &= P(\{2,2\}, \{2,3\}, \ldots, \{2,6\}, \{3,2\}, \{4,2\}, \ldots, \{6,2\}) \\
&= \frac{5+4}{36} = \frac{1}{4} \\
&\vdots \\
P_M(6) &= \frac{1}{36} \\
P_M(m) &= \frac{13 - 2m}{36} \qquad (m = 1, 2, 3, 4, 5, 6)
\end{aligned}
$$

3.8 Independent R.V.

Definition 3.15 (Independent Discrete R.V.)**.** *Let X, Y be R.V. defined on a P.S. S. If*

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

*for all $x, y \in \mathbb{R}$ then say X and Y are **independent**.*

Note 3.1. *If X and Y are discrete then $P(X = x, Y = y) = P_X(x)P_Y(y)$*

Ex 3.16.

$$
\begin{aligned}
X &\sim \mathit{DUnif}\,(\{1, 2, 3, 4, 5, 6\}) \\
Y &\sim \mathit{Bin}\,(4, 0.3) \\
P(X = Y) &= \sum_{i=1}^{4} P(X = Y = i) \\
&= \sum_{i=1}^{4} P(X = i)P(Y = i) \\
&= \frac{1}{6}\sum_{i=1}^{4} \binom{4}{i}(0.3)^i(0.7)^{4-i} \\
P(X + Y = 3) &= P(X = 1)P(Y = 2) + P(X = 2)P(Y = 1) + P(X = 3)P(Y = 0) \\
&= \frac{1}{6}\binom{4}{2}(0.3)^2(0.7)^2 + \frac{1}{6}\binom{4}{1}(0.3)^1(0.7)^3 + \frac{1}{6}\binom{4}{0}(0.3)^0(0.7)^4
\end{aligned}
$$

Ex 3.17. *$X \sim$ DUnif $(\{1,2,3\})$, $Y \sim$ DUnif $(\{1,2,3\})$, X and Y are independent R.V.s. Show that $X+Y$ and $X-Y$ are not independent.*

Proof.

$$\begin{aligned}
P(X+Y=4, X-Y=1) &= 0 \\
P(X+Y=4) &= P(X=1,Y=3)+P(X=2,Y=2)+P(X=3,Y=1) \\
&= P(X=1)P(Y=3)+P(X=2)P(Y=2)+P(X=3)P(Y=1) \\
&= \frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{1}{3} \\
&= \frac{1}{3} \\
P(X-Y=1) &= P(X=2,Y=1)+P(X=3,Y=2) \\
&= P(X=2)P(Y=1)+P(X=3)P(Y=2) \\
&= \frac{2}{9}
\end{aligned}$$

$$P(X+Y=4, X-Y=1) = 0 \neq P(X+Y=4)P(X-Y=1) = \frac{1}{3}\cdot\frac{2}{9}$$

□

Definition 3.16 (General Independent R.V.)**.** *$X_1, X_2, \ldots, X_n$ R.V.'s are **independent** if for all $x_1, x_2, \ldots, x_n \in \mathbb{R}$*

$$P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) = P(X_1 \le x_1)P(X_2 \le x_2)\ldots P(X_n \le x_n)$$

Ex 3.18. *Assume X_1, X_2, X_3 satisfies*

$$P(X_1 \le x_1, X_2 \le x_2, X_3 \le x_3) = P(X_1 \le x_1)P(X_2 \le x_2)P(X_3 \le x_3)$$

Take $\lim_{x_2\to\infty}$

$$P(X_1 \le x_1, X_3 \le x_3) = P(X_1 \le x_1)P(X_3 \le x_3)$$

So X_1 and X_3 are independent.

Caution 3.1. *$X_1, \ldots, X_n$ R.V.'s. Take $i \neq j$, X_i, X_j independent. $X_1, \ldots, X_n$ are not necessarily independent.*

Definition 3.17 (Independent Identically Distributed)**.** *$X_1, \ldots, X_n$ are **independent identically distributed (i.i.d.)** if they are independent and have the same C.D.F. $F(X)$ (If discrete, same p.m.f., $P(X)$)*

Ex 3.19. *Roll a die m times. X_i gives value of i^{th} roll. $X_i \sim$ DUnif $(\{1,2,3,4,5,6\})$, $X_1, \ldots, X_m$ are i.i.d.*

Theorem 3.6. *Assume $X_1, X_2, \ldots, X_n$ i.i.d. where $X_i \sim$ Bern (p) then*

$$X = \sum_{i=1}^{n} X_i \sim \textit{Bin}\,(n,p)$$

Theorem 3.7. *Assume $X \sim$ Bin (n,p) and $Y \sim$ Bin (m,p) are independent. Then*

$$X+Y \sim \textit{Bin}\,(n+m,p)$$

Proof (I). Let $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ be i.i.d. R.V.'s where $X_i \sim \text{Bern}\,(p)$ and $Y_i \sim \text{Bern}\,(p)$. Then

$$\begin{aligned}
X &= \sum_{i=1}^{n} X_i \sim \text{Bin}\,(n,p) \\
Y &= \sum_{i=1}^{m} Y_i \sim \text{Bin}\,(m,p) \\
X+Y &= \sum_{i=1}^{n} X_i + \sum_{i=1}^{m} Y_i \sim \text{Bin}\,(n+m,p)
\end{aligned}$$

□

Proof (II).

$$
\begin{aligned}
P(X+Y=k) &= \sum_{i=0}^{k} P(X=i, Y=k-i) \\
&= \sum_{i=0}^{k} P(X=i)P(Y=k-i) \\
&= \sum_{i=0}^{k} \binom{n}{i} p^i q^{n-i} \binom{m}{k-i} p^{k-i} q^{m-(k-i)} \\
&= \sum_{i=0}^{k} \binom{n}{i} \binom{m}{k-i} p^k q^{(n+m)-k} \\
&= \sum_{i=0}^{k} \binom{n+m}{k} p^k q^{(n+m)-k} \\
&\sim \text{Bin}\,(n+m, p)
\end{aligned}
$$

□

Theorem 3.8. *X, Y are distinct R.V.'s.*

1. *The **conditional p.m.f.** for X given $Y=y$ is*

$$P_{X|Y}(x \mid y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

   where $P(Y=y) \neq 0$

2. *Similarly,*

$$P_{Y|X}(y \mid x) = \frac{P(X=x, Y=y)}{P(X=x)} = P(Y=y \mid X=x)$$

Note 3.2.

1. *$P_{X|Y}(x \mid y)$ is a valid p.m.f. as long as $P(Y=y) \neq 0$*
2. *If X, Y are independent, then*

$$
\begin{aligned}
P_{X|Y}(x \mid y) &= \frac{P(X=x, Y=y)}{P(Y=y)} \\
&= \frac{P(X=x)P(Y=y)}{P(Y=y)} \\
&= P(X=x)
\end{aligned}
$$

Ex 3.20. *X, Y satisfies $P(X=x, Y=y)=k$ if $1 \leq y \leq x \leq 6$ and x, y are integers.*

1. *Find k*
2. *Find $P_{X|Y}(x \mid 3)$*

Answer.

1. There are $\sum_1^6 = 21$ points (x, y) can take, then

$$1 = \sum_{(x,y)} P(X=x, Y=y) = 21k$$

$$k = \frac{1}{21}$$

2.

$$
\begin{aligned}
P_{X|Y}(x \mid 3) &= \frac{P(X = x, Y = 3)}{P(Y = 3)} \\
&= \frac{P(X = x, Y = 3)}{\sum_{x=3}^{6} P(X = x, Y = 3)} \\
&= \frac{\frac{1}{21}}{4 \cdot \frac{1}{21}} \\
&= \frac{1}{4}
\end{aligned}
$$

We have DUnif $(\{3, 4, 5, 6\})$ R.V.

Definition 3.18. *X, Y, Z three R.V. on S P.S., then X and Y are **conditionally independent given** Z if*

$$P(X \leq x, Y \leq y \mid Z = z) = P(X \leq x \mid Z = z)P(Y \leq y \mid Z = z)$$

Ex 3.21. *A school has n sophomores and m seniors. Each decides to attend a basketball game, independently. Sophomore attend with probability p_1 and senior attend with probability p_2. X is number of sophomore attending and Y is number of seniors attending. Then $X \sim$ Bin (n, p_1) and $Y \sim$ Bin (m, p_2) and total number attending is $X + Y$. Assume we know that $X + Y = k$, what is $P(X = x \mid X + Y = k)$?*

Answer.

$$
\begin{aligned}
P(X = x \mid X + Y = k) &= \frac{P(X + Y = k \mid X = x)P(X = x)}{P(X + Y = k)} && \text{(Bayes I)} \\
&= \frac{P(Y = k - x)P(X = x)}{P(X + Y = k)}
\end{aligned}
$$

Assume $p_1 = p_2 = p$, then

$$
\begin{aligned}
X + Y &\sim \text{Bin}\,(n + m, p) \\
P(X = x \mid X + Y = k) &= \frac{P(Y = k - x)P(X = x)}{P(X + Y = k)} \\
&= \frac{\binom{n}{x} p^x q^{n-x} \binom{m}{k-x} p^{k-x} q^{m-(k-x)}}{\binom{n+m}{k} p^k q^{n+m-k}} \\
&= \frac{\binom{n}{x}\binom{m}{k-x} p^k q^{n+m-k}}{\binom{n+m}{k} p^k q^{n+m-k}} \\
&= \frac{\binom{n}{x}\binom{m}{k-x}}{\binom{n+m}{k}} \\
&\sim \text{HGeom}\,(n, m, k) && \text{(if } x = 0, 1, \ldots, k\text{)}
\end{aligned}
$$

Theorem 3.9. *If $X \sim$ Bin (n, p), $Y \sim$ Bin (m, p) are independent R.V., the conditional p.m.f. of X given $X + Y = k$ is HGeom (n, m, k).*

Chapter 4

Expected Value

4.1 Expected Value

Definition 4.1 (Expected Value)**.** *If X is a discrete R.V., taking values $x_1, x_2, \ldots$, then*

$$E(X) = \sum_i x_i P(X = x_i) = \sum_i x_i P_X(x_i)$$

Ex 4.1. *Roll die. X is number rolled, then $X \sim$ DUnif $(\{1,2,3,4,5,6\})$, then*

$$E(X) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{1 + \cdots + 6}{6} = 3.5$$

Note 4.1. *We require $\sum_i x_i P_X(x_i)$ to converge absolutely for $E(X)$ to exist.*

Ex 4.2. *$X \sim$ Bern (p), then*

$$E(X) = 1 \cdot p + 0 \cdot q = p$$

Note 4.2.

1. *$E(X)$ defined only on **distrbition** of X meaning $P_X(x)$ gives $E(X)$*
2. $$\begin{aligned} E(X) &= \sum_x P(X = x) \\ &= \sum_x xP(\{s \in S \mid X(s) = x\}) \end{aligned}$$

Theorem 4.1. *$X \sim$ Bin (n, p) then $E(X) = n \cdot p$*

4.2 Linearity of Expected Values

Theorem 4.2. *If X, Y R.V.s defined on P.S. S, then*

1. *$E(X + Y) = E(X) + E(Y)$*
2. *$E(cX) = cE(X)$, for c constant*

Proof.

$$
\begin{aligned}
E(X) + E(Y) &= \sum_{s \in S} X(s)P(s) + \sum_{s \in S} Y(s)P(s) \\
&= \sum_{s \in S} (X(s) + Y(s))P(s) \\
&= E(X + Y) \\
E(cX) &= \sum_{s \in S} cX(s)P(s) \\
&= c \sum_{s \in S} X(s)P(s) \\
&= cE(X)
\end{aligned}
$$

□

Ex 4.3. $X \sim \textit{Bern}(p) \quad E(X) = p$

Ex 4.4.

$$
\begin{gathered}
X \sim \textit{Bin}(p) \quad E(X) = np \\
x_1, \ldots, x_n \sim \textit{Bern}(p) \\
\sum_{i=1}^{n} x_i \sim \textit{Bin}(n, p) \\
E(X) = E\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} E(x_i) = np
\end{gathered}
$$

Note 4.3. *$E(x)^2$ and $E(x^2)$ are not same!*

Ex 4.5. $X \sim \textit{DUnif}(\{-1, 0, 1\})$*,*

$$
\begin{aligned}
E(X) &= (-1) \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} \\
&= 0 \\
E(X)^2 &= 0 \\
E(X^2) &= (-1)^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3} \\
&= \frac{2}{3}
\end{aligned}
$$

Ex 4.6. *X, Y R.V.s on P.S. S, $E(X) = 3$ and $E(Y) = -2$. Find*

- $E(X + Y) = E(X) + E(Y) = 1$
- $E(3X - 2Y) = 3E(X) - 2E(Y) = 13$

Theorem 4.3. *If $X \sim \textit{HGeom}(g, r, n)$ then*

$$
E(X) = n \cdot \frac{g}{g + r}
$$

Proof. Within g green marbles and r red marbles, pick n with replacement, X is the # of green marble.

$$x_i = \begin{cases} 1 & \text{if } i^{th} \text{ ball is green} \\ 0 & \text{otherwise} \end{cases}$$

$$x_i \sim \text{Bern}\left(\frac{g}{g+r}\right)$$

$$\sum_{i=1}^{n} \sim \text{HGeom}\,(g, r, n)$$

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^{n} x_i\right) \\ &= \sum_{i=1}^{n} E(x_i) \\ &= n \cdot \frac{g}{g+r} \end{aligned}$$

□

Recall 4.1. *If* $|r| < 1$

$$\sum_{n=1}^{\infty} ar^{n+1} = \sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$$

Ex 4.7.

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{3^{n+4}}{7^{n-2}} &= \sum_{n=0}^{\infty} \frac{81 \cdot 3^n}{49^{-1} 7^n} \\ &= 49 \cdot 81 \sum_{n=0}^{\infty} \left(\frac{3}{7}\right)^n \\ &= \frac{49 \cdot 81}{1 - \frac{3}{7}} \\ &= 49 \cdot 81 \cdot \frac{7}{4} \end{aligned}$$

Ex 4.8. *Roll a die until* 6 *is rolled. Let* X *be the number of rolls* ***before*** *a 6.*

$$S = \left\{ \underbrace{NN \ldots NN}_{n} 6 \right\} \mid n \geq 0$$

$$P(X = n) = \frac{1}{6}\left(\frac{5}{6}\right)^n, n = 0, 1, 2, \ldots$$

$$\begin{aligned} \sum_{n=0}^{\infty} P(X = n) &= \sum_{n=0}^{\infty} \frac{1}{6}\left(\frac{5}{6}\right)^n \\ &= \frac{1}{6} \cdot \frac{1}{1 - \frac{5}{6}} = 1 \end{aligned}$$

4.3 Geometric and Negative Binomial Distributions

Definition 4.2 (Geometric Distribution)**.** *We say* $X \sim$ *Geom* (p) *has a* ***geometric distribution*** *if* $0 < p < 1$ *and* X *has p.m.f.*

$$P_X(x) = \begin{cases} (1-p)^x p = q^x p & \textit{if } x = 0, 1, 2, \ldots \\ 0 & \textit{otherwise} \end{cases}$$

.

Note 4.4. *If we repeat Bernoulli trials until first success, then X is the number of failures until first success.*

Definition 4.3 (First Success Distribution)**.** *If we count number of trials until 1^{st} success we get **first success** distribution,*

$$Y \sim FS\,(p)$$

Y has p.m.f.

$$P_Y(y) = q^{y-1}p, \quad y = 1, 2, \dots$$

and $Y = X + 1$. If $X \sim Geom\,(p)$ then $Y \sim FS\,(p)$.

Proposition 4.1. *if $0 < x < 1$ then*

$$\sum_{k=0}^{x} kx^{k-1} = \frac{1}{(1-x)^2}$$

Theorem 4.4. *If $X \sim Geom\,(p)$ then $E(X) = \frac{q}{p}$*

Proof.

$$\begin{aligned} E(X) &= \sum_x xP_X(x) \\ &= \sum_x xq^x p \\ &= pq \sum_x xq^{x-1} \\ &= pq \cdot \frac{1}{(1-q)^2} \\ &= pq \cdot \frac{1}{p^2} \\ &= \frac{q}{p} \end{aligned}$$

□

Corollary 4.1. *If $Y \sim FS\,(p)$ then $E(Y) = \frac{1}{p}$*

Proof. If $X \sim \text{Geom}\,(p)$, $X + 1 \sim \text{FS}\,(p)$,

$$E(Y) = E(X+1) = E(X) + E(1) = \frac{q}{p} + 1 = \frac{p+q}{p} = \frac{1}{p}$$

□

Ex 4.9. *Roll 2 dice until sum is 11. What is expected number of rolls?*

Answer. Probability of rolling 11 is $\frac{2}{3}6 = p$. Let Y is number of rolls, then $Y \sim \text{FS}\left(\frac{2}{36}\right)$,

$$E(Y) = \frac{1}{p} = \frac{1}{\frac{2}{36}} = 18$$

Definition 4.4 (Negative Binomial Distribution)**.** *A R.V. with p.m.f.*

$$P_X(x) = P(X = x) = \binom{x+r-1}{r-1} p^r q^x, \quad x = 0, 1, 2, \dots$$

*with $0 < p < 1$, $q = 1 - p$, $r \geq 1$ are interger, is called a **Negative Binomial R.V.**. Denoted $X \sim NBin\,(r, p)$.*

Ex 4.10. *Roll a die until r 6's appears. Each roll are i.i.d. Bernoulli Trials, $p = \frac{1}{6}$. X is number of failures before r^{th} 6 is rolled. Number of rolls with exactly x N's is*

$$\binom{x+r-1}{r-1}$$

each has probability $q^x p^r$, then

$$P(X = x) = \binom{x+r-1}{r-1} \left(\frac{1}{6}\right)^r \left(\frac{5}{6}\right)^x$$

Ex 4.11. *If $X \sim$ NBin (r, p), find $E(X)$.*

Answer. Let X_i be the number of failures after $(i-1)^{\text{th}}$, before i^{th} success.

$$\begin{aligned}
X_i &\sim \text{Geom}\,(p) \\
X &= \sum_{i=1}^{r} X_i \\
E(X) &= \sum_{i=1}^{r} E(X_i) \\
&= \sum_{i=1}^{r} \frac{q}{p} \\
&= \frac{rq}{p}
\end{aligned}$$

Note 4.5.

$$E(X) = \sum_{n=0}^{\infty} x \binom{x+r-1}{r-1} p^r q^x = \frac{rq}{p}$$

Ex 4.12 (Collector Problem). *One of n toys given randomly each visit. What is expected number of visits to get a completed set?*

Answer. Let X_i be number of visits to get i^{th} toy after getting $(i-1)^{\text{th}}$ toy and $X_i \sim \text{FS}\left(\frac{n-i+1}{n}\right)$. $X = \sum_{i=1}^{n} X_i$ is the total number of visits.

$$\begin{aligned}
E(X) &= E(\left(\sum_{i=1}^{n} x_i\right)) \\
&= \sum_{i=1}^{n} E(x_i) \\
&= \sum_{i=1}^{n} \frac{n}{n+1-i} \\
r = 6 \quad & E(X) \approx 14.7 \\
r = 8 \quad & E(X) \approx 21.7
\end{aligned}$$

4.4 Indicator R.V. and The Fundamental Bridge

Definition 4.5. *$A \subseteq S$, S is the P.S., **indicator R.V.** of A is*

$$\begin{aligned}
I_A &: S \longrightarrow A \\
I_A(S) &= \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{otherwise} \end{cases} \\
I_A &\sim \textit{Bern}\,(P(A)) \\
E(A) &= P(A) = p
\end{aligned}$$

Theorem 4.5. *$A_1, A_2, \ldots, A_n \subseteq S$, $I_{\bigcup_{i=1}^n A_i} \leq \sum_{i=1}^n I_{A_n}$*

$$\begin{aligned}
P\left(\bigcup_{i=1}^n A_i\right) &= E(I_{\bigcup_{i=1}^n A_i}) \\
&\leq E\left(\sum_{i=1}^n A_i\right) \\
&= \sum_{i=1}^n E(A_i) \\
&= \sum_{i=1}^n P(A_i)
\end{aligned}$$

Theorem 4.6. *S P.S.,*

- *If $A, B \subseteq S$, $I_{A\cup B} = I_A I_B$*
- *If $A_1, A_2, \ldots, A_n \subseteq S$, then*

$$I_{\bigcap_{i=1}^n A_i} = \prod_{u=1}^n I_{A_i}$$

- $I_{A^\complement} = 1 - I_A$

Ex 4.13 (Inclusion and Exclusion for three sets)**.** *$A_1, A_2, A_3 \subseteq S$,*

$$\begin{aligned}
I_{A_1\cup A_2\cup A_3} &= 1 - I_{(A_1\cup A_2\cup A_3)^\complement} \\
&= 1 - I_{A_1^\complement \cap A_2^\complement \cap A_3^\complement} \\
&= 1 - I_{A_1^\complement} I_{A_2^\complement} I_{A_3^\complement} \\
&= 1 - (1 - I_{A_1})(1 - I_{A_2})(1 - I_{A_2}) \\
&= 1 - [1 - I_{A_1} - I_{A_2} - I_{A_2} + I_{A_1}I_{A_2} + I_{A_2}I_{A_3} + I_{A_1}I_{A_3} - I_{A_1}I_{A_2}I_{A_2} \\
&= I_{A_1} + I_{A_2} + I_{A_2} - I_{A_1}I_{A_2} - I_{A_2}I_{A_3} - I_{A_1}I_{A_3} + I_{A_1}I_{A_2}I_{A_2}
\end{aligned}$$

4.5 L.O.T.U.S.

Theorem 4.7 (Law of the Unconscious Statistician)**.** *If X is a discrete R.V. with p.m.f. $P_X(x)$ then*

$$E(g(X)) = \sum_x g(x) P_X(x)$$

Proof. If $y \in \mathbb{R}$, let $A_y = \{x \in \mathbb{R} \mid g(x) = y\}$.

$$\begin{aligned}
P_Y(y) = P(x \in A_y) &= \sum_{A \in A_y} P_X(x) \\
E(Y) &= \sum_y y \cdot P_Y(y) \\
&= \sum_y y \sum_{x \in A_y} P_X(x) \\
&= \sum_x g(x) P_X(x)
\end{aligned}$$

□

Ex 4.14. *If $X \sim$ Bin $(2, .3)$, find $E(Y)$ where $Y = 3^x$.*

Answer.

$$\begin{aligned} E(Y) &= \sum_x g(x)P_X(x) \\ &= \sum_{x=0}^{2} 3^x \binom{2}{x}(.3)^x(.7)^{2-x} \\ &= \sum_{x=0}^{2} \binom{2}{x}(.9)^x(.7)^{2-x} \end{aligned}$$

Ex 4.15. *$X \sim$ Geom $\left(\frac{3}{4}\right)$, find $E(Y)$ where $Y = 2^x$.*

Answer.

$$\begin{aligned} E(Y) &= \sum_{x=0}^{\infty} 2^x \frac{1}{4}^x \frac{3}{4} \\ &= \frac{3}{4}\sum_{x=0}^{\infty} \frac{1}{2}^x \\ &= \frac{3/4}{1 - 1/2} \\ &= \frac{3}{2} \end{aligned}$$

4.6 Variance

$E(X)$ is the "average" value X takes. How spread out are values of X? Could try $E(|X - E(X)|)$, not desirable.

Definition 4.6 (Variance)**.** *If X is a R.V., its **variance** is*

$$\sigma^2 = \mathrm{Var}(X) = E((X - E(X))^2), (\sigma > 0)$$

Definition 4.7 (Standard Deviation)**.** *The **standard deviation** is*

$$\sigma_x = \sigma = \sqrt{\mathrm{Var}(X)}$$

Ex 4.16. *$X \sim$ Bern (P),*

$$\begin{aligned} \mathrm{Var}(X) &= E((X - E(X)^2)) \\ &= E((X - p)^2) \\ &= (0 - p)^2 q + (1 - p)^2 p \\ &= p^2 q + q^2 p = pq(q + p) = pq \\ \sigma_x &= \sqrt{pq} \end{aligned}$$

Theorem 4.8. *X is R.V. then*

$$\mathrm{Var}(X) = E(X^2) - E(X)^2$$

Proof.

$$\begin{aligned}\mathrm{Var}(X) &= E((X - E(X))^2) \\ &= E((X^2) - 2E(X)X + E(X)^2) \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2\end{aligned}$$

□

Lemma 4.1. *Let $0 < x < 1$, then*

(i)

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

(ii)

$$\sum_{k=1}^{\infty} k^2 x^{k-1} = \frac{1+x}{(1-x)^3}$$

Proof.

(i) Computed to find $E(X)$, $X \sim \mathrm{Geom}\,(p)$.

(ii) Multiply (ii) by x.

$$\sum_{k=1}^{\infty} kx^{k} = \frac{x}{(1-x)^2}$$

Differentiate w.r.t. x

$$\begin{aligned}\sum_{k=1}^{\infty} k^2 x^{k-1} &= \frac{(1-x)^2 \cdot 1 - x \cdot 2(1-x)^1(-1)}{(1-x)^4} \\ &= \frac{1 - x + 2x}{(1-x)^3} \\ &= \frac{1+x}{(1-x)^3}\end{aligned}$$

□

Theorem 4.9. *If $X \sim$ Geom (p), then*

$$\mathrm{Var}(X) = \frac{q}{p^2}$$

Proof.

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{\infty} k^2 q^k p \\
&= pq \sum_{k=0}^{\infty} k^2 q^{k-1} \\
&= pq \sum_{k=1}^{\infty} k^2 q^{k-1} \\
&= pq \frac{1+q}{(1-q)^3} \\
&= pq \frac{1+q}{p^3} \\
&= \frac{q(1+q)}{p^2} \\
\mathrm{Var}(X) &= E(X^2) - E(X)^2 \\
&= \frac{q(1+q)}{p^2} - \left(\frac{q}{p}\right)^2 \\
&= \frac{q+q^2-q^2}{p^2} = \frac{q}{p^2}
\end{aligned}$$

□

Theorem 4.10. *If a and b are constants, then*

$$\mathrm{Var}(aX+b) = a^2 \,\mathrm{Var}(X)$$

Proof.

$$\begin{aligned}
\mathrm{Var}(aX+b) &= E((aX+b)^2) - E(aX+b)^2 \\
&= E(a^2X^2 + 2abX + b^2) - (aE(X)+b)^2 \\
&= E(a^2X^2) + E(2abX) + E(b^2) - (a^2E(X)^2 + 2abE(X) + b^2) \\
&= a^2(E(X^2) - E(X)^2) \\
&= a^2 \,\mathrm{Var}(X)
\end{aligned}$$

□

Theorem 4.11. *If $X_1, \ldots, X_n$ are independent then*

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$$

Ex 4.17. *X is a R.V., $E(X) = 3$, $\mathrm{Var}(X) = 5$, find $E(2X^2 + 7X + 5)$*

Answer.

$$\begin{aligned}
\mathrm{Var}(X) &= E(X^2) - 9 \\
E(X^2) &= 5 + 9 = 14 \\
E(2X^2 + 7X + 5) &= 2E(X^2) + 7E(X) + 5 \\
&= 2 \cdot 14 + 7 \cdot 3 + 5
\end{aligned}$$

Theorem 4.12. *If* $X \sim$ *Bin* (n, p) *then* $\operatorname{Var}(X) = mpq$.

Proof. $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim$ Bern (p), $X = X_1 + \cdots + X_n$,

$$\begin{aligned} \operatorname{Var}(X) &= \operatorname{Var}(X_1 + \cdots + X_n) \\ &= \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n) \\ &= npq \end{aligned}$$

□

Theorem 4.13. *If* $X \sim$ *NBin* (r, p) *then* $\operatorname{Var}(X) = r\frac{q}{p^2}$

Proof. We have $X = \sum_{i=1}^{r} X_i$ i.i.d $X_i \sim$ Geom (p),

$$\begin{aligned} \operatorname{Var}(X) &= \operatorname{Var}\left(\sum_{i=1}^{r} X_i\right) \\ &= \sum_{i=1}^{r} \operatorname{Var}(X_i) \\ &= \sum_{i=1}^{r} \frac{q}{p^2} \\ &= r\frac{q}{p^2} \end{aligned}$$

□

4.7 Poisson Distribution

Ex 4.18. *Customer arriving in 1 hour has expected value* λ. *Equally likely to get customer at any time. Assume*

1. *customers show up independently*
2. *in any time interval at most 1 customer will show up*

$$X_i = \begin{cases} 1 & \textit{if customer in } i^{th} \textit{ interval} \\ 0 & \textit{otherwise} \end{cases}$$

$$X_i \sim \textit{Bin}\left(\frac{\lambda}{n}\right)$$

Let X *be the total number of contomers is* $\sum_{i=1}^{n} X_i \sim$ *Bin* $\left(n, \frac{\lambda}{n}\right)$

$$\begin{aligned} P(X = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\ldots(n-k+1)}{k!n^k} \cdot \lambda^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{1}{k!}\binom{n}{n}\binom{n-1}{n}\binom{n-k+1}{n}\lambda^k \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \end{aligned}$$

Fix k, *take* $\lim_{n \to \infty}$

$$\begin{aligned} P(X = k) &= \frac{1}{k!} \cdot 1 \cdot 1 \ldots 1 \cdot \lambda^k \frac{e^{-\lambda}}{(1-0)^k} \\ &= \frac{e^{-\lambda}\lambda^k}{k!} \end{aligned}$$

Definition 4.8. *A random variable with p.m.f.*

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

*is a **Poisson R.V.** with parameter $\lambda > 0$. Denoted $X \sim$ Pois (λ).*

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \qquad \text{(Mac. Series)}$$

$$\begin{aligned}
\sum_k P_X(k) &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{-k}}{k!} \\
&= e^{-\lambda} e^{\lambda} = 1
\end{aligned}$$

Ex 4.19. *If $X \sim$ Pois (λ) find*

1. $E(X)$

$$\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k P_X(k) \\
&= \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\
&= e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^(k-1)}{(k-1)!} \\
&= e^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \lambda e^{\lambda} = \kappa
\end{aligned}$$

2. $E(X(X-1))$

$$\begin{aligned}
E(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} \\
&= e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \\
&= e^{-\lambda} \lambda^2 \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \lambda^2 e^{\lambda} = \lambda^2
\end{aligned}$$

Theorem 4.14. *If $X \sim$ Pois (λ) then* $\mathrm{Var}(X) = \lambda$

Proof.

$$\begin{aligned} E(X(X-1)) &= \lambda^2 \\ E(X^2 - X) &= \lambda^2 \\ E(X^2) - E(X) &= \lambda^2 \\ E(X^2) &= \lambda^2 + E(X) = \lambda^2 + \lambda \\ \mathrm{Var}(X) &= E(X^2) - E^2(X) \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

□

Ex 4.20. *Roll 3 dices, (red, blue, green),*

$$S = \{(r, b, g) \mid 1 \leq r, b, g \leq 6\}$$

Event total rolls 6:

$$E = \{(1,1,4),(1,2,3),(1,3,2),(1,4,1),(2,1,3),(2,2,2),(2,3,1),(3,1,2),(3,2,1),(4,1,1)\}$$

with $|S| = 6^3$*, simple S.S.*

$$\begin{aligned} I_E : S &\longrightarrow \mathbb{R} \\ s &\longmapsto \begin{cases} 1 & \textit{if } s \in E \\ 0 & \textit{otherwise} \end{cases} \\ I_E &\sim \textit{Bern}\left(\frac{10}{6^3}\right) \\ I_{E^\complement} &\sim \textit{Bern}\left(\frac{26}{6^3}\right) \end{aligned}$$

4.8 Poisson and Binomial Connections

Theorem 4.15. *Assume $X \sim$ Pois (λ_1), $Y \sim$ Pois (λ_2) are independent, then*

$$X + Y \sim \textit{Pois}\,(\lambda_1 + \lambda_2)$$

Proof. If $k \geq 0$, then

$$\begin{aligned} P(X+Y=k) &= \sum_{i=0}^{k} P(X=i, Y=k-i) \\ &= \sum_{i=0}^{k} P(X=i)P(Y=k-i) \\ &= \sum_{i=0}^{k} \frac{e^{-\lambda_1}\lambda_1^i}{i!} \frac{e^{-\lambda_2}\lambda_2^{k-i}}{(k-i)!} \\ &= e^{-\lambda_1}e^{-\lambda_2} \sum_{i=0}^{k} \frac{\lambda_1^i \lambda_2^{k-i}}{i!(k-i)!} \end{aligned}$$

$$
\begin{aligned}
&= \frac{e^{-\lambda_1}e^{-\lambda_2}}{k!}\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\lambda_1^i\lambda_2^{k-i} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda_1^i\lambda_2^{k-i} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}(\lambda_1+\lambda_2)^k \sim \text{Pois}\left(\lambda_1+\lambda_2\right)
\end{aligned}
$$

□

Theorem 4.16 (Poisson given a sum of Poisson)**.** *Assume $X \sim$ Pois $\left(\lambda_1\right), Y \sim$ Pois $\left(\lambda_2\right)$ are independent. The conditional distribution of X given $X+Y=n$ is*

$$
\textit{Bin}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)
$$

Proof.

$$
\begin{aligned}
P(X=k \mid X+Y=n) &= \frac{P(X=k \mid X+Y=n)P(X=k)}{P(X+Y=n)} && \text{(Baye's Rule)} \\
&= \frac{P(Y=n-k)P(X=k)}{P(X+Y=n)} \\
&= \frac{\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k!)}\frac{e^{-\lambda_1}\lambda_1^k}{k!}}{\frac{e^{\lambda_1+\lambda_2}(\lambda_1+\lambda_2)^n}{n!}} \\
&= \frac{\frac{\lambda_2^{n-k}}{(n-k!)}\frac{\lambda_1^k}{k!}}{\frac{(\lambda_1+\lambda_2)^n}{n!}} \\
&= \frac{\lambda_2^{n-k}}{(n-k!)}\frac{\lambda_1^k}{k!}\frac{n!}{(\lambda_1+\lambda_2)^n} \\
&= \frac{n!}{(n-k!)}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n k!} \\
&= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-k} \\
&\sim \text{Bin}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)
\end{aligned}
$$

□

$X \sim$	$P_X(x)$	$E(X)$	$\text{Var}(X)$
Bern (p)	$(x = 1) \to 1, (x \neq 1) \to 0$	p	pq
Bin (n, p)	$\binom{n}{x} p^x q^{n-x}$	np	npq
DUnif (S)	$\frac{1}{\lvert S \rvert}$	$\sum_{s \in S} \frac{s}{\lvert S \rvert}$	
HGeom (g, r, n)	$\frac{\binom{g}{x}\binom{r}{n-x}}{\binom{g+r}{n}}$	$\frac{ng}{g+r}$	
Geom (p)	$q^x p$	$\frac{q}{p}$	$\frac{q}{p^2}$
FS (p)	$q^{x-1} p$	$\frac{1}{p}$	
NBin (r, p)	$\binom{x+r-1}{r-1} p^r q^x$	$r \cdot \frac{q}{p}$	$r \cdot \frac{q}{p^2}$
Pois (λ)	$\frac{e^{-\lambda}\lambda^k}{k!}$	λ	λ

Chapter 5

Continuous R.V.

Definition 5.1 (Continuous R.V.). *X is a **continuous random variable** if its C.D.F. $F(X)$ is continuous everywhere and $F'(X)$ exists, except at finitely many points.*

Definition 5.2. *If X is a cont R.V. with C.D.F. $F(X)$, its **probability density function** is*

$$f_X(x) = \begin{cases} F'(x) & \text{when } F'(X) \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

Remark 5.1. *If X is continuous, $P(X = a) = P(a) - \lim_{x \to a} F(x) = F(a) - F(a) = 0$.*

Remark 5.2. *If $a < b$ then*

$$\int_a^b f_X(x)dx = \int_a^b F_X'(x)dx = F_X(b) - F_X(a) = P(a < X \leq b)$$

Theorem 5.1. *If X is a R.V. with p.d.f. $f_X(x)$ then*

1. $f_X(x) \geq 0$
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$

Ex 5.1. *X has p.d.f.*

$$f_X(x) = \begin{cases} ax^2 & \text{if } -1 < x < 0 \\ 0 & \text{otherwise} \end{cases}$$

1. *Find a*
2. *Compute $P\left(-\frac{1}{2} < x < 0\right)$*

Answer.

1.

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x)dx &= \int_{-1}^{0} ax^2 dx \\ &= \frac{a}{3}x^3 \Big|_{-1}^{0} \\ &= \frac{a}{3} = 1 \\ \frac{a}{3} &= 1 \Rightarrow a = 3 \end{aligned}$$

2.

$$
\begin{aligned}
P\left(-\frac{1}{2} < x < 0\right) &= P\left(-\frac{1}{2} < x <= 0\right) \\
&= \int_{-\frac{1}{2}}^{0} f_X(x)dx \\
&= \int_{-\frac{1}{2}}^{0} 3x^2 dx \\
&= x^3 \Big|_{-\frac{1}{2}}^{0} = \frac{1}{8}
\end{aligned}
$$

Definition 5.3 (Expected Value)**.** *If X has p.d.f. $f_X(x)$, the **expected value** of X is*

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx \qquad \text{(If converges absolutely)}$$

Theorem 5.2 (L.O.T.U.S.)**.** *If X has p.d.f. $f_X(x)$ and $Y = g(x)$, then*

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x)dx$$

Definition 5.4.

$$
\begin{aligned}
\mathrm{Var}(X) &= E((X - E(X))^2) \\
&= E(X^2) - E(X)^2
\end{aligned}
$$

Ex 5.2. *For the last example, compute* $\mathrm{Var}(X)$

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x)dx \\
&= \int_{-1}^{0} 3x^3 dx = \frac{3x^4}{4}\Big|_{-1}^{0} \\
&= -\frac{3}{4} \\
E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \\
&= \int_{-1}^{0} x^2 \cdot 3x^2 dx \\
&= \frac{3}{5} x^5 \Big|_{-1}^{0} = \frac{3}{5} \\
\mathrm{Var}(x) &= E(X^2) - E(X)^2 \\
&= \frac{3}{5} - \left(-\frac{3}{4}\right)^2 \\
&= \frac{48 - 45}{80} = \frac{3}{80}
\end{aligned}
$$

Ex 5.3. *The **logistic R.V.** has C.D.F. $F_X(x) = \frac{e^x}{1+e^x}$. The p.d.f. is*

$$f_X(x) = F_X'(x) = \frac{(1+e^x)e^x - e^x(0+e^x)}{(1+e^2)^2} = \frac{e^x}{(1+e^x)^2}$$

5.2 Uniform Distribution

Definition 5.5 (Uniform Distribution). *A R.V. with p.d.f.*

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

is called a ***uniform R.V.*** *on* (a, b). *Denoted* $X \sim \mathit{Unif}(a, b)$.

Ex 5.4. *If* $X \sim \mathit{Unif}(a, b)$, $a < b$, *find* $E(X)$ *and* $\mathrm{Var}(X)$.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \frac{1}{b-a} \int_a^b x dx \\ &= \frac{1}{b-a} \cdot \frac{1}{2}(b^2 - a^2) \\ &= \frac{1}{2}(a+b) \\ E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\ &= \frac{1}{b-a} \int_a^b x^2 dx \\ &= \frac{1}{3} \cdot \frac{1}{b-a}(b^3 - a^3) \\ &= \frac{b^2 + ab + a^2}{3} \\ \mathrm{Var}(X) &= \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

Theorem 5.3. *Assume* $X \sim \mathit{Unif}(a, b)$, *conditional distribution of* X, *given* $c < X < d$ *where* $a < c < d < b$, *is* $\mathit{Unif}(c, d)$.

Theorem 5.4 (Remaining C.D.F. from p.d.f.). *If* X *has p.d.f.* $f_X(x)$, *its C.D.F. is*

$$F_X(x) = \int_{-\infty}^{x} f_X(t) dt$$

Proof.

$$\begin{aligned} p(X \le x) &= \lim_{x \to -\infty} p(a \le X \le x) \\ &= \lim_{x \to -\infty} \int_a^x f_X(t)\, dt \\ &= \int_{-\infty}^{x} f_X(t)\, dt \end{aligned}$$

□

Ex 5.5. $X \sim \mathit{Unif}(a, b)$, *find C.D.F. for* X.

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

Theorem 5.5. *If X has p.d.f. $f_X(x)$, C.D.F. $F_X(x)$, $Y = F^{-1}(a)$, $a \sim$ Unif$(0,1)$, then Y has same distribution as X.*

5.4 Normal Distribution

Definition 5.6. *A R.V. z with p.d.f.*

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

is called a ***standard normal R.V.***.

Theorem 5.6. *We must have $I = \int_{-\infty}^{\infty} \phi(z)\, dz = 1$.*

Proof.

$$\begin{aligned}
I^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx \int_{-\infty}^{\infty} e^{-y^2} dy \\
&= \frac{1}{2\pi} \iint_{\mathbb{R}^2} e^{-\left(\frac{x^2+y^2}{2}\right)}\, dA \\
&= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} r\, dr\, d\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi} \left(-e^{-\frac{1}{2}r^2} \Big|_0^{\infty} \right) d\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi} 1\, d\theta \\
&= \frac{1}{2\pi} \cdot 2\pi = 1
\end{aligned}$$

□

Remark 5.3. $\phi(-z) = \phi(z)$

Definition 5.7. *If $Z \sim$ S.N. R.V., its* ***C.D.F.*** *is*

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \phi(t)\, dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt$$

Theorem 5.7. *If $z \in \mathbb{R}$, then $\Phi(z) + \Phi(-z) = 1$.*

Definition 5.8. *If Z is a S.N. R.V., σ, μ be two constants and $X = \sigma Z + \mu$ be a normal R.V., write*

$$X \sim N(\mu, \sigma)$$

Theorem 5.8. *If $X \sim N(\mu, \sigma)$, then*

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

Theorem 5.9 (68, 95, 99.7 Rule)**.** *If $X \sim N(\mu, \sigma)$, then*

$$\begin{aligned}
P(|x-u| < \sigma) &\approx 0.68 \\
P(|x-u| < 2\sigma) &\approx 0.95 \\
P(|x-u| < 3\sigma) &\approx 0.997
\end{aligned}$$

5.5 Exponential Distribution

Definition 5.9. *A R.V. with p.d.f. $f_X(x) = \lambda e^{-\lambda x}$ is called an **exponential R.V.** with parameter $\lambda > 0$. Denoted $X \sim$ Expo (λ).*

Theorem 5.10. *The C.D.F. of $X \sim$ Expo (λ) is*

$$\begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Theorem 5.11. *If $X \sim$ Expo (λ), then*

1. $E(X) = \frac{1}{\lambda}$
2. $\operatorname{Var} X = \frac{1}{\lambda^2}$

Theorem 5.12 (Memoryless Propertiy of Exponential)**.** *If $X \sim$ Expo (λ), then*

$$P(X \geq s + t \mid X \geq s) = P(X \geq t)$$

Proof.

$$\begin{aligned} P(X \geq x) &= 1 - P(X < x) \\ &= 1 - [1 - e^{-\lambda x}] \\ &= e^{-\lambda x} \\ P(X \geq s + y \mid x \geq s) &= \frac{P(X \geq s+t)}{P(X \geq s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} \\ &= P(X \geq t) \end{aligned}$$

□

Ex 5.6. *In Sunnyville, trains run every 30 min reliably. In Cloudyville, trains run as an Expo $\left(\frac{1}{15}\right)$ R.V. In both, expected waiting time is 15 min. In Sunnyville, if you wait 15 min, your new expected waiting time is $\frac{15}{2}$. In Cloudyville, if you wait 15 min, your new expected waiting time is still 15 min.*

Chapter 6

Moments

Definition 6.1. *X R.V. The n^{th} **moment** of X is $E(X^n)$.*

(a) If X has p.m.f. $P_X(x)$, then

$$E(X^n) = \sum_x x^n P_X(x)$$

(b) If X has p.d.f. $f_X(x)$, then

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

Requires absolute convergence.

6.1 Summaries of a R.V.

Definition 6.2. *$c \in \mathbb{R}$ is a **median** for X if $P(X \leq c) = \frac{1}{2}$ and $P(X \geq c) = \frac{1}{2}$.*

Ex 6.1. *Roll a die, XDUnif$(\{1, 2, 3, 4, 5, 6\})$, then any $c \in [3, 4]$ is a median.*

Ex 6.2. *If $X \sim$ Expo(3), then*

$$\begin{aligned} P(X \leq c) = \frac{1}{2} &= 1 - e^{-3c} \\ e^{3c} &= 2 \\ c &= \frac{\ln 2}{3} \end{aligned}$$

Definition 6.3. *$c \in \mathbb{R}$ is a **mode** for X if*

(a) $P_X(c) \geq P_X(x)$, for all x.

(b) $f_X(c) \geq f_X(x)$, for all x.

Theorem 6.1. *Let X have a R.V. with $E(X) = \mu$ and median m, then*

1. *the value of c that minimizes the expected value of $E((X - c)^2)$ is $c = \mu$.*

   Proof.

$$\begin{aligned} E((X - c)^2) &= E(X^2 - 2cX + c^2) \\ &= E(X^2) - 2c\mu + c^2 \\ &= E(X^2) - E^2(X) + \mu^2 - 2c\mu + c^2 \\ &= \mathrm{Var}(X) + (\mu - c)^2 \\ c &= \mu \end{aligned}$$

□

2. *the value of c that minimizes the expected value of $E(|X-c|)$ is $c=m$.*

Definition 6.4. *Let X be a R.V. with $E(X)=\mu$. Then*

1. *$E(X^n)$ is the n-th **moment**of X.*
2. *$E(X-\mu)^n$ is the n-th **central moment** of X.*
3. *$E\left(\left(\frac{X-\mu}{\sigma}\right)^n\right)$ is the n-th **standardized moment** of X.*

Note 6.1. $E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right)$ *is called the **skewness** of X. If X has p.d.f. $f_X(x)$ where*

$$f_X(\mu-c)=f_X(\mu+c) \text{ for all } c$$

then

$$E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right)=0$$

6.4 Moment Generating Functions

Definition 6.5 (Moment Generating Functions)**.** *Given R.V. X the **moment generating function** (m.g.f) of X is*

$$\Psi(t)=\Psi_X(t)=E\left(e^{tX}\right)$$

Theorem 6.2.

$$\Psi(0)=E\left(e^{0X}\right)=E(1)=1$$

Theorem 6.3. *If $X_1,\ldots,X_n$ are independent R.V., and*

$$X=\sum_{i=1}^{n}X_i$$

then

$$\Psi_X(t)=\prod_{i=1}^{n}\Psi_{X_i}(t)$$

Theorem 6.4. *If X,Y be R.V. such that $\Psi_X(t)=\Psi_Y(t)$, when $-a<t<a$, for some $a>0$, then X and Y are the same distribution (same p.m.f. and p.d.f.).*

Theorem 6.5 (The m.g.f. of a location-scale transformaton)**.** *Let X be a R.V. with a,b constants, then the m.g.f. of $Y=aX+b$ is*

$$\Psi_Y(t)=e^{bt}\Psi_X(at)$$

Proof.

$$\begin{aligned}
\Psi_Y(t)&=E\left(e^{tY}\right)\\
&=E\left(e^{t(aX+b)}\right)\\
&=\int_{-\infty}^{\infty}e^{t(ax+b)}f_X(x)dx\\
&=e^{bt}\int_{-\infty}^{\infty}e^{axt}f_X(x)dx\\
&=e^{bt}E\left(e^{(at)X}\right)\\
&=e^{bt}\Psi_X(at)
\end{aligned}$$

□

Theorem 6.6. *Let X be a R.V. with m.g.f. $\Psi_X(t)$ defined for $-a < t < a$, for some $a > 0$. Then all moments of X exist and*

$$E(X^n) = \Psi_X^{(n)}(0)$$

Note 6.2.

$$\Psi_X(t) = \sum_{n=0}^{\infty} \frac{\Psi_X^{(n)(0)}}{n!} t^n$$

Ex 6.3. *If $X \sim$ Bern (p), then*

$$\begin{aligned}\Psi_X(t) &= e^{1t} P_X(1) + e^{0t} P_X(0) \qquad \text{(LOTUS)}\\ &= pe^t + q\end{aligned}$$

Ex 6.4. *$X \sim$ Geom (p), $P_X(x) = p^x p$ if $x = 0, 1, 2, \ldots$, then*

$$\begin{aligned}\Psi_X(t) &= \sum_{x=0}^{\infty} e^{tx} q^x p\\ &= \sum_{x=0}^{\infty} p(qe^t)^x\\ &= \frac{p}{1 - qe^t}\end{aligned}$$

Ex 6.5. *$X \sim$ Unif (a, b), then*

$$\begin{aligned}\Psi_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x) dx\\ &= \frac{1}{b-a} \int_a^b e^{tx} dx\\ &= \frac{1}{b-a} \left(\frac{e^{tx}}{t} \bigg|_a^b \right)\\ &= \begin{cases} \frac{e^{tb} - e^{ta}}{t(b-a)} & \text{if } t \neq 0\\ 1 & \text{if } t = 0 \end{cases}\end{aligned}$$

Ex 6.6. *If $X \sim$ Pois (λ) then*

$$\begin{aligned}\Psi_X(t) &= \int_{-0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!}\\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}\\ &= e^{-\lambda} e^{\lambda e^t}\\ &= e^{\lambda(e^t - 1)}\end{aligned}$$

Ex 6.7. *Assume $X \sim$ Pois (λ_1) and $Y \sim$ Pois (λ_2) are independent. Then what is the distribution of $X + Y$?*

$$\begin{aligned}\Psi_{X+Y}(t) &= \Psi_X(t) \Psi_Y(t)\\ &= e^{\lambda_1(e^t - 1)} e^{\lambda_2(e^t - 1)}\\ &= e^{(\lambda_1 + \lambda_2)(e^t - 1)}\\ \implies X + Y &\sim \text{Pois}\,(\lambda_1 + \lambda_2)\end{aligned}$$

Ex 6.8. *Let $X \sim$ Pois (λ), find $E(X)$ and $E(X^2)$.*

$$\begin{aligned} \Psi(t) &= e^{\lambda(e^t-1)} \\ \Psi'(t) &= \lambda e^t e^{\lambda(e^t-1)} \\ \Psi''(t) &= \lambda e^t e^{\lambda(e^t-1)} + \left(\lambda e^t\right)^2 e^{\lambda(e^t-1)} \\ E(x) = \Psi'(0) &= \lambda \\ E(X^2) = \Psi''(0) &= \lambda + \lambda^2 \\ \operatorname{Var} X &= E(X^2) - E(X)^2 = \lambda \end{aligned}$$

Ex 6.9. *If $X \sim$ Expo (λ), then*

$$\Psi_X(t) = \frac{\lambda}{\lambda - t}$$

$$\begin{aligned} f_X(x) &= \lambda e^{-\lambda x} \\ \Psi_X(t) &= \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx \\ &= \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx \\ &= \frac{\lambda}{\lambda - t} && \text{(for } \lambda - t > 0\text{)} \end{aligned}$$

Ex 6.10. *If $X \sim$ NBin (r, p) R.V., find $\Psi_X(t)$.*

$$\begin{aligned} X_i &\sim \textit{Geom}\,(p) \ \textit{i.i.d} \\ X &= \sum_{i=1}^r X_i \\ \Psi_X(t) &= \Psi_{\sum_{i=1}^r X_i}(t) \\ &= \prod_{i=1}^r \Psi_{X_i}(t) \\ &= \prod_{i=1}^r \frac{p}{1 - qe^t} \\ &= \left(\frac{p}{1 - qe^t}\right)^r \end{aligned}$$

Ex 6.11. *If $Z \sim N(0, 1)$, then*

$$\begin{aligned} \Psi_Z(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{tz} e^{-\frac{1}{2}z^2}\, dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{1}{2}(z^2 - 2tz)}\, dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{1}{2}(z-t)^2 + \frac{1}{2}t^2}\, dz \\ &= e^{\frac{1}{2}t^2} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{1}{2}(z-t)^2}\, dz\right] \\ &= e^{\frac{1}{2}t^2} \end{aligned}$$

Ex 6.12. *If $X \sim N(\mu, \sigma)$, then*

$$\begin{aligned} X &= \sigma Z + \mu \\ \Psi_X(t) &= e^{\mu t}\Psi_Z(\sigma t) \\ &= e^{\mu t} e^{\frac{1}{2}(\sigma t)^2} \\ &= e^{\frac{1}{2}\sigma^2 t^2 + \mu t} \end{aligned}$$

Ex 6.13. *If $X_i \sim N(\mu_i, \sigma_i)$ are n independent R.V.'s, $X = \sum_{i=1}^n X_i$ then*

$$\begin{aligned} \Phi_X &= \prod_{i=1}^n \Phi_{X_i} \\ &= \prod_{i=1}^n \exp\left(\frac{1}{2}\sigma_i^2 t^2 + \mu_i t\right) \\ &= \exp\left(\frac{1}{2}\sum_{i=1}^n \sigma_i^2 t^2 + \sum_{i=1}^n \mu_i t\right) \sim N\left(\sum_{i=1}^n \mu_i, \sqrt{\sum_{i=1}^n \sigma_i^2}\right) \end{aligned}$$

Chapter 7

Multivariate Distribution

7.1 Bivariate Distribution

Definition 7.1. *Assume X, Y are two R.V. defined on P.S. S, then (x, y) given a* ***random point*** *in $\mathbb{R}^2$.*

7.2 Joint Distribution

Definition 7.2. *If X, Y are* ***jointly distributed*** *and (X, Y) takes values in a sequence $\{(x_1, y_1), (x_2, y_2), \dots\}$, which can be finite or infinite, then X, Y have a* ***joint discrete distribution****, and the* ***joint p.m.f.*** *of X, Y is*

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

Note 7.1.

$$P_{X,Y}(x, y), \geq 0 \quad \sum_{(x,y)} P_{X,Y}(x, y) = 1$$

Definition 7.3. *If we have joint p.m.f. $P_{X,Y}(x, y)$, then*

1. *The* ***marginal p.m.f.*** *for X is*
$$P_X(x) = \sum_y P_{X,Y}(x, y)$$
2. *The* ***marginal p.m.f.*** *for Y is*
$$P_Y(y) = \sum_x P_{X,Y}(x, y)$$

Definition 7.4. *If X, Y are 2 R.V. on P.S. S, the*

1. *the* ***joint distribution function*** *of X, Y is*
$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$
2. *the* ***marginal distribution*** *for X is*
$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y)$$
3. *the* ***marginal distribution*** *for Y is*
$$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y)$$

Definition 7.5. *X, Y have joint p.m.f $P_{X,Y}(x, y)$.*

1. *If $P_X(x) \neq 0$, then the* ***conditional*** *p.m.f for Y given $X = x$ is*

$$P_{Y|X}(y \mid x) = \frac{P_{X,Y}(x, y)}{P_X(x)}$$

2. *If $P_Y(y) \neq 0$, then the* ***conditional*** *p.m.f for X given $Y = y$ is*

$$P_{X|Y}(x \mid y) = \frac{P_{X,Y}(x, y)}{P_Y(y)}$$

Note 7.2.

$$\begin{aligned}\sum_y P_{Y|X}(y \mid x) &= \sum_y \frac{P_{X,Y}(x, y)}{P_X(x)} \\ &= \frac{1}{P_X(x)} \sum_y P_{X,Y}(x, y) \\ &= \frac{1}{P_X(x)} \cdot P_X(x) \\ &= 1\end{aligned}$$

Theorem 7.1. *If X, Y are discrete then X and Y are independent if and only if*

$$P_{X,Y}(x, y) = P_X(x) P_Y(y)$$

Theorem 7.2. *X, Y have joint cdf $F_{X,Y}(x, y)$ with marginal cdf $F_X(x), F_Y(y)$. Then X, Y are independent if and only if*

$$F_{X,Y}(x, y) = F_X(x) F_Y(y)$$

Ex 7.1. *X, Y have p.m.f. given by table:*

Y \ X	1	2	3	$P_Y(y)$
1	2/24	3/24	4/24	9/24
2	0	4/24	5/24	9/24
3	0	0	6/24	6/24

$$\begin{aligned}P_X(1) &= \sum_y P_{X,Y}(1, y) \\ &= P_{X,Y}(1, 1) + P_{X,Y}(1, 2) + P_{X,Y}(1, 3) \\ P_Y(y) &= \sum_x P_{X,Y}(x, y)\end{aligned}$$

Ex 7.2. *Compute $P_{Y|X}(y \mid 3)$.*

$$\begin{aligned}P_{Y|X}(y \mid 3) &= \frac{P_{X,Y}(3, y)}{P_X(3)} \\ &= \begin{cases} \frac{4/24}{15/24} & \text{if } y = 1 \\ \frac{5/24}{15/24} & \text{if } y = 2 \\ \frac{6/24}{15/24} & \text{if } y = 3 \end{cases}\end{aligned}$$

Ex 7.3. *N is the number of egg layed, $N \sim$ Pois (λ). Each egg hatches with probability p. Let X be the number of eggs that hatch and Y are the number of eggs that do not hatch. $X+Y=N$. Show that X, Y are independent.*

Proof.

$$\begin{aligned}
P_{X,Y}(i,j) &= P(X=i, Y=j) \\
&= \sum_{n=0}^{\infty} P(X=i, Y=j \mid N=n) P(N=n) && \text{(L.O.T.P.)} \\
&= \binom{i+j}{i} p^i q^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} \\
&= e^{-\lambda} \frac{(i+j)!}{i!j!} p^i q^j \frac{\lambda^{i+j}}{(i+j)!} \\
&= e^{-\lambda} \frac{(p\lambda)^i (q\lambda)^j}{i!j!} \\
&= e^{-p\lambda} \frac{(p\lambda)^i}{i!} e^{-q\lambda} \frac{(q\lambda)^j}{j!} && (e^{-p\lambda} \cdot e^{-q\lambda} = e^{-\lambda}) \\
P_X(x) &= \sum_j \frac{e^{-p\lambda}(p\lambda)^i}{i!} \frac{e^{-q\lambda}(q\lambda)^j}{j!} \\
&= \frac{e^{-p\lambda}(p\lambda)^i}{i!} \\
X &\sim \text{Pois}\,(p\lambda) \\
Y &\sim \text{Pois}\,(q\lambda) \\
P_{X,Y}(i,j) &= P_X(i) P_Y(j) \Rightarrow X, Y \text{ are independent}
\end{aligned}$$

□

Theorem 7.3 (LOTUS)**.** *X, Y defined P.S. S, $Z = g(X,Y)$*

1. *If X, Y have joint p.m.f. $P_{X,Y}(x,y)$, then*

$$E(Z) = \sum_{x,y} g(x,y) P_{X,Y}(x,y)$$

2. *If X, Y have joint p.d.f. $f_{X,Y}(x,y)$, then*

$$E(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy$$

7.3 Covariance and Correlation

Question 7.1. *If X, Y not independent, what is $\mathrm{Var}(X+Y)$? Let $E(X) = \mu_X, E(Y) = \mu_Y$,*

$$\begin{aligned}
\mathrm{Var}(X+Y) &= E((X+Y-(\mu_X+\mu_Y))^2) \\
&= E(((X-\mu_X)+(Y-\mu_Y))^2) \\
&= E((X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2) \\
&= E((X-\mu_X)^2) + 2E((X-\mu_X)(Y-\mu_Y)) + E((Y-\mu_Y)^2) \\
&= \mathrm{Var}(X) + 2E((X-\mu_X)(Y-\mu_Y)) + \mathrm{Var}(Y)
\end{aligned}$$

Definition 7.6 (Covariance)**.** *The* ***covariance*** *of* X, Y *is*

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E((X-\mu_X)(Y-\mu_Y)) \\ &= E(XY - -\mu_X Y - \mu_Y X + \mu_X\mu_Y) \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

Definition 7.7. *The* ***correlation*** *of* X, Y *is*

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

where σ_X, σ_Y *are the standard deviation of* X, Y*, and*

$$-1 \leq \mathrm{Corr}(X,Y) \leq 1$$

Caution 7.1. *If* $\mathrm{Var}(X,Y) = 0$*, then* $\mathrm{Corr}(X,Y) = 0$*, but* X, Y *need* ***not*** *to be independent.*

Theorem 7.4 (Properties of Covariant)**.** *Given* X, Y, Z, W *R.V. defined on P.S.* S

1. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
2. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
3. $\mathrm{Cov}(X,c) = 0$
4. $\mathrm{Cov}(aX,Y) = a\,\mathrm{Cov}(X,Y)$
5. $\mathrm{Cov}(X,Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$
6. $\mathrm{Cov}(X+Y,Z+W) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(X,W) + \mathrm{Cov}(Y,Z) + \mathrm{Cov}(Y,W)$
7. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$
8. $\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{1\leq i\leq j\leq n} \mathrm{Cov}(X_i, X_j)$

Theorem 7.5. *If* X, Y *are independent, then* $\mathrm{Cov}(X,Y) = 0$.

Ex 7.4. X, Y *has p.m.f. by the table*

Y \ X	2	3	$f_Y(y)$
1	.1	.2	.3
4	.3	.1	.4
6	.2	.1	.3
$f_X(x)$	.6	.4	

$$
\begin{aligned}
E(X) &= \sum_x x P_X(x) \\
&= 2 \cdot (.6) + 3 \cdot (.4) \\
&= 2.4 \\
E(Y) &= \sum_y y P_Y(y) \\
&= 1 \cdot (.3) + 4 \cdot (.4) + 6 \cdot (.3) \\
&= 3.7 \\
E(XY) &= \sum_{x,y} xy P_{X,Y}(x,y) \\
&= 2 \cdot 1 \cdot (.1) + 2 \cdot 4 \cdot (.3) + 2 \cdot 6 \cdot (.2) \\
&\quad + 3 \cdot 1 \cdot (.2) + 3 \cdot 4 \cdot (.1) + 3 \cdot 6 \cdot (.1) \\
&= 8 \\
\mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) \\
&= 8 - 2.4 \cdot 3.7
\end{aligned}
$$

Ex 7.5. *If* $X \sim$ *HGeom* (g, r, n)*, find* $\mathrm{Var}(X)$.

$$
\begin{aligned}
X_i &= \begin{cases} 1 & \text{if } i\text{th ball is green} \\ 0 & \text{otherwise} \end{cases} \\
X_i &\sim \textit{Bern}\left(p = \frac{g}{g+r}\right) \\
X &= \sum_{i=1}^{n} X_i \\
\mathrm{Var}(X_i) &= pq = \frac{gr}{(g+r)^2} \\
\mathrm{Cov}(X_i, X_j) &= \mathrm{Cov}(X_1, X_2) \\
P(X_1, X_2 = 1) &= P(X_1 = 1)P(X_2 = 1 \mid X_1 = 1) \\
&= p \cdot \frac{g-1}{g+r-1} \\
\mathrm{Cov}(X_1, X_2) &= E(X_1X_2) - E(X_1)E(X_2) \\
&= p \cdot \frac{g-1}{g+r-1} - \left(\frac{g}{g+r}\right)^2 \\
&= p\left(\frac{g-1}{g+r-1}\right) - p^2 \\
\mathrm{Var}(X) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \le i \le j \le n} \mathrm{Cov}(X_i, X_j) \\
&= n \cdot \frac{gr}{(g+r)^2} + \frac{n(n-1)}{2} \cdot \left[\frac{g}{g+r}\left(\frac{g-1}{g+r-1}\right) - \left(\frac{g}{g+r}\right)^2\right]
\end{aligned}
$$

Chapter 8

Transformatons

8.1 Change of Variables

Ex 8.1. *Let $X \sim N(0,1)$ and $Y = |X|$, find p.d.f. for Y.*

Y has C.D.F.

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= \begin{cases} P(-Y < X < Y) & \text{if } y > 0 \\ 0 & \text{if } y \leq 0 \end{cases} = \begin{cases} \Psi(y) - \Psi(-y) & \text{if } y > 0 \\ 0 & \text{if } y \leq 0 \end{cases} \\
&= \begin{cases} 2\Psi(y) - 1 & \text{if } y > 0 \\ 0 & \text{if } y \leq 0 \end{cases} \\
f_Y(y) &= F'_Y(y) \\
&= \begin{cases} 2\Psi'(y) & \text{if } y > 0 \\ 0 & \text{if } y \leq 0 \end{cases} \\
&= \begin{cases} \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} & \text{if } y > 0 \\ 0 & \text{if } y \leq 0 \end{cases}
\end{aligned}$$

Theorem 8.1. *Assume X is a R.V. with $P(a < x < b) = 1$. Let $f_X(x)$ be p.d.f. of X. Assume $g(X)$ is Differentiate function with $g'(x) \neq 0$ for $a < x < b$. Then $Y = g(X)$ has p.d.f.*

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

for $g(a) < y < g(b)$.

Proof. Y has C.D.F.

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(g(X) \leq y) \\
&= P(X \leq g^{-1}(y)) \\
&= F_X(g^{-1}(y)) \\
f_Y(y) &= F'_Y(y) \\
&= F'_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y) \\
&= f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)
\end{aligned}$$

$\square$

Ex 8.2. $X \sim$ *Expo* (λ), $Y = X^2$, $(a, b) = (\alpha, \beta) = (0, \infty)$

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= \lambda e^{-\lambda\sqrt{y}} \cdot \frac{1}{2\sqrt{y}} \\ &= \frac{\lambda}{2\sqrt{y}} e^{-\lambda\sqrt{y}} \end{aligned}$$

Chapter 9

Conditional Expectation

9.1 Conditional Expectation Given an Event

Definition 9.1. *X in R.V. R on S, $A \subseteq S$ is an event with $P(A) \neq 0$. Then*

1. $E(X \mid A) = \sum_x x \underbrace{P(X = x \mid A)}_{\text{condition pmf}}$
2. $E(X \mid A) = \int_{-\infty}^{\infty} \underbrace{f_{X|A}(X = x \mid A)}_{\text{conditional pdf}}\ dx$

Ex 9.1. *X, Y jointly distributed with p.d.f. given by the same table in **Ex 7.4**. Find $E(Y \mid X = 3)$, $E(Y)$ and $E(X|Y = 6)$.*

$$\begin{aligned}
P_{Y|X}(y \mid 3) &= \frac{P_{X,Y}(3, y)}{P_X(3)} = \begin{cases} \frac{.2}{.4} = .5 & \text{if } y = 1 \\ \frac{.1}{.4} = .25 & \text{if } y = 4 \text{ or } 6 \end{cases} \\
E(Y \mid X = 3) &= \sum_y y P_{Y|X}(y = 3) \\
&= 1 \cdot (.5) + 4 \cdot (.25) + 6 \cdot (.25) = 3 \\
E(Y) &= \sum_y y P_Y(y) \\
&= 1 \cdot (.3) + 4 \cdot (.4) + 6 \cdot (.3) = 3.7
\end{aligned}$$

Ex 9.2. *X, Y has joint p.d.f. $f_{X,Y}(x, y) = \frac{3}{64}x$ inside the triangle with vertices $(0, 0), (4, 0), (4, 4)$. Find $E(Y)$ and $E(X \mid Y = 1)$.*

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx \\
&= \int_y^4 \frac{3}{64} x\ dx \\
&= \frac{3}{8} - \frac{3}{128} y^2 \qquad (0 < y < 4) \\
E(Y) &= \int_{-\infty}^{\infty} y f_Y(y)\ dy \\
&= \int_0^4 y\ \left(\frac{3}{8} - \frac{3}{128} y^2\right)\ dy
\end{aligned}$$

$$
\begin{aligned}
&= \frac{21}{8} \\
f_{X|Y}(x \mid 1) &= \frac{f_{X,Y}(x,1)}{f_Y(1)} \\
&= \frac{\frac{3}{64}x}{\frac{3}{8} - \frac{3}{128}} \\
&= \frac{8}{7}x && \text{(if } 0 < x < 4\text{)} \\
E(X \mid Y = 1) &= \int_{-\infty}^{\infty} x f_{X|Y}(x \mid 1)\, dx \\
&= \int_1^4 \frac{3x}{24 - \frac{3}{2}}\, dx \\
&= \left(\frac{3}{24 - \frac{3}{2}}\right) \frac{1}{2}x^2 \Big|_1^4 \\
&= \left(\frac{3}{24 - \frac{3}{2}}\right) \left(\frac{15}{2}\right)
\end{aligned}
$$

Chapter 10

Inequalities and Limit Theorem

10.1 Inequalities

Theorem 10.1 (Markov's Inequality)**.** *X R.V. with $P(x > 0) = 1$. If $a > 0$, then*

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Proof. X has p.d.f $f_X(x)$.

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x)\, dx \\
&= \int_{0}^{\infty} x f_X(x)\, dx \\
&= \int_{0}^{a} x f_X(x)\, dx + \int_{a}^{\infty} x f_X(x)\, dx \\
&\geq \int_{a}^{\infty} x f_X(x)\, dx \\
&\geq \int_{a}^{\infty} a f_X(x)\, dx \\
&= a \int_{a}^{\infty} f_X(x)\, dx \\
&= a P(X \geq a) \\
P(X \geq a) &\leq \frac{E(X)}{a}
\end{aligned}$$

□

Theorem 10.2. *Given any R.V. X, let $Y = |X|$. Then $P(Y \geq 0) = 1$. Apply Markov's Inequality to Y, we have*

$$P(|X| \geq a) \leq \frac{E(|X|)}{a}$$

Theorem 10.3 (Chebyshev's Inequality)**.** *X R.V. with $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$, then for $c > 0$*

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Proof. Let $Y = (X-\mu)^2$ and $P(Y \geq 0) = 1$, then

$$\begin{aligned} E(Y) &= E((X-\mu)^2) = \mathrm{Var}(X) = \sigma^2 \\ P(|X-\mu| \geq c) &= P((X-\mu)^2 \geq c^2) \\ &\leq \frac{E((X-\mu)^2)}{c^2} = \frac{\sigma^2}{c^2} \end{aligned}$$

□

Note 10.1. *If we replace c by $c\sigma$, then*

$$P(|X-\mu| \geq c\sigma) \leq \frac{1}{c^2}$$

using Complement Rule, we have

$$P(|X-\mu| < c\sigma) \geq 1 - \frac{1}{c^2}$$

Ex 10.1. $X \sim$ *Expo* (4), $E(X) = \frac{1}{4}$, $a = 3$.

$$P(X \geq a) \leq \frac{\frac{1}{3}}{4} = \frac{1}{12}$$

exact value is

$$P(X \geq 3) = 1 - \int_0^3 4e^{-4x}\, dx \approx 0.000006$$

Ex 10.2. *If $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$, then*

$$P(|X-\mu| < 3\sigma) \geq 1 - \frac{1}{9} = \frac{8}{9} \quad \left(< 0.9\right)$$

Ex 10.3. *If $X \sim N(\sigma, \mu)$ then*

$$\begin{aligned} P(|X-\mu| < 3\sigma) &= P(-3\sigma < X-\mu < 3\sigma) \\ &< P(-3 < \frac{X-\mu}{\sigma} < e) \\ &= \Phi(3) - \Phi(-3) \approx 0.9973 \end{aligned}$$

10.2 Law of Large Numbers

Let X_i i.i.d., assume $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ both exists.

$$\begin{aligned} \text{Let } \bar{X_n} &= \frac{X_1 + \cdots + X_n}{n} \\ E(\bar{X_n}) &= E\left(\frac{X_1 + \cdots + X_n}{n}\right) \\ &= \frac{E(X_1) + \cdots + E(X_n)}{n} \\ &= \frac{n\mu}{n} = \mu \\ \mathrm{Var}(\bar{X_n}) &= \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) \\ &= \frac{\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)}{n^2} \\ &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \\ \bar{X_n} \text{ has standard deviation } &\frac{\sigma}{\sqrt{n}} \to 0 \text{ as } n \to \infty \end{aligned}$$

Theorem 10.4 (Weak Law of Large Numbers, WLLN). *X_i i.i.d., $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$, then*

$$P(\left|\bar{X}_n - \mu\right| > \epsilon) \to 0 \text{ as } n \to \infty$$

for any $\epsilon > 0$.

Proof. We have

$$0 \leq P(\left|\bar{X}_n - \mu\right| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty$$

so

$$P(\left|\bar{X}_n - \mu\right| > \epsilon) \to 0 \text{ as } n \to \infty$$

□

Theorem 10.5 (Strong Law of Large Numbers, SLLN). *X_i i.i.d., $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$, then*

$$P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$$

Ex 10.4. *X_i i.i.d., $E(X) = 4.2$, $\mathrm{Var}(X_i) = \sigma^2 = 7$, find n such that*

$$P(\left|\bar{X}_n - 4.2\right| < 0.1) \geq 0.95$$

By Chebyshev's Inequality,

$$\begin{aligned} P(\left|\bar{X}_n - 4.2\right| < 0.1) &\geq 1 - \frac{\sigma^2}{n(0.1)^2} \\ &= 1 - \frac{7}{n(0.01)} \\ &= 0.95 \\ \Rightarrow n &\geq 14,000 \end{aligned}$$

Ex 10.5. *X_i i.i.d. $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$. Find c such that*

$$P\left(\lim_{n\to\infty} \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^\nu X_i^2} = c\right) = 1$$

Solution.

$$\begin{aligned} \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} &= \frac{\frac{1}{n}\sum_{i=1}^n X_i}{\frac{1}{n}\sum_{i=1}^n Y_i} && \text{(where } Y_i = X_i^2\text{)} \\ E(Y_i) &= E(X_i^2) = \sigma^2 + \mu^2 \\ P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) &= 1 \\ P\left(\lim_{n\to\infty} \bar{Y}_n = \sigma^2 + \mu^2\right) &= 1 \\ P\left(\lim_{n\to\infty} \frac{\bar{X}_n}{\bar{Y}_n} = c\right) &= 1 \textit{ if } c = \frac{\mu}{\sigma^2 + \mu^2} \end{aligned}$$

10.3 Central Limit Theorem

Theorem 10.6 (Central Limit Theorem).

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \to N(0,1) \text{ as } n \to \infty$$

Theorem 10.7 (Central Limit Theorem, Approximation form). *For large n, the distribution of $\bar{X}_n$ is approx-mated by*

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$