Chapter 1: Introduction

We mainly cover:

1. Probability model
2. Commonly used random variables

1.1 Basic concepts of probability model

- Probability model
  3 components of a probability model:
  - Sample space
  - Event
  - Probability function
- Sample space:
  All possible outcomes of a random experiment.
  > e.g. Toss a coin twice: $S=\{(H,H),(H,T),(T,H),(T,T)\}$
- Event
  Roughly speaking: Event = subset of the sample space.
  > e.g. E=$\{(H,H)\}$, subset of sample space S.
- Probability function
  Notation P. Probability function is a function of event and satisfies 3 conditions:
  i. $0 \leq P(E) \leq 1$ for any event E
  ii. $P(\text{Sample Space})=1$
  iii. Suppose $E_1, E_2, \ldots$ are a sequece of disjoint events, i.e. $E_i \cap E_j=\emptyset$(empty set) for $i \neq j$.
  Then $P(\bigcup_{i=1}^{\infty} E_i)=\sum_{i=1}^{\infty} P(E_i)$ [Additivity Property]
  > e.g. Toss coin twice
  > Define $P(E)=\frac{\#\text{ of outcomes in } E}{4\text{ (Total } \#\text{ of outcomes in sample space)}}$
  >
  > > Exercise: Verify P is a probablity function
- Property of probability function
  i. If $E_1 \subset E_2$, then $P(E_1) \leq P(E_2)$
  - If $E1$ occurs, then E_2 must occur $\implies E_1 \subset E_2$.
  - If E_1 implies E_2, then $E_1 \subset E_2$.

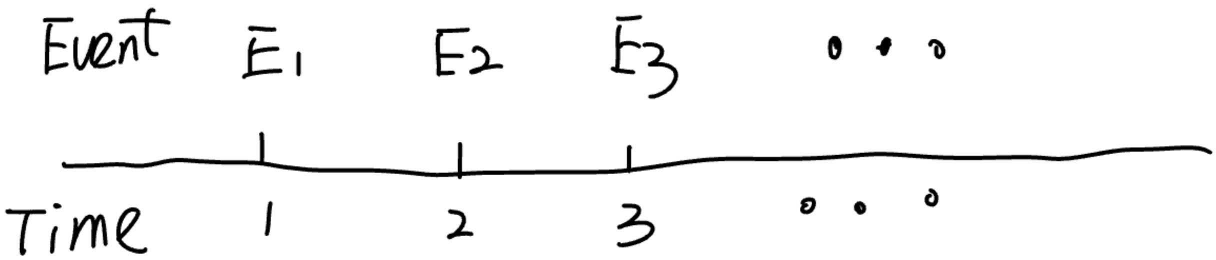
  ii. If $E_1 \cap E_2=\emptyset$, then $P(E_1 \cup E_2)=P(E_1)+P(E_2)$
  iii. $P(\emptyset)=0$
  iv. $P(E)+P(E^c)=1$
  - E^c is the complementary of E.

v. $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

- Independence
  Two events E and F are independent if $P(E \cap F) = P(E) \cdot P(F)$ where $P(E \cap F)$ is the joint probability and $P(E) \cdot P(F)$ is marginal probabilities or unconditional probabilities.
  so, independence means:

$$\text{joint = product of marginals}$$

- A useful result
  Suppose we have a sequence of independent trials and a sequence of events $E_1, E_2, \ldots$.
  Now: E_i only depends on ith trial $\implies$:
  i. all events $E_1, E_2, \ldots$ are independent
  ii. $P\left(\bigcap_{i=1}^{\infty} E_i\right) = \prod_{i=1}^{\infty} P(E_i)$
  iii. $P\left(\bigcap_{i=1}^{m} E_i\right) = \prod_{i=1}^{m} P(E_i)$

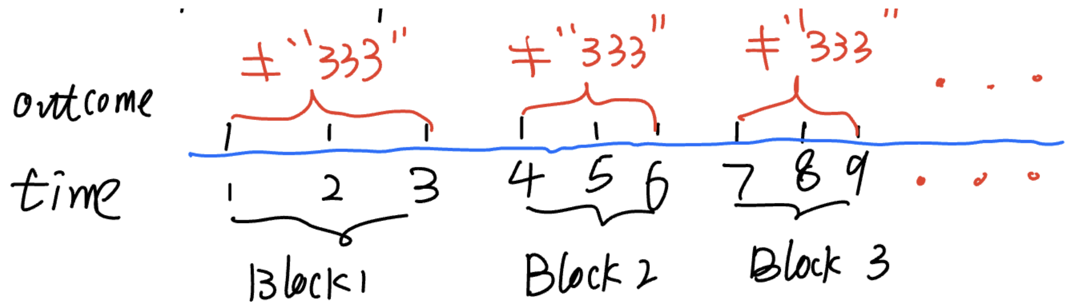


> Example 1.1 (Independent Example):
> Suppose the die is a fair die. If you repeatedly and independently toss a die, then you will get a sequence of numbers.
> Find:
>
> 1. $P($getting a "3" in the sequence of numbers$)$
>
> > Solutions:
> > Let E = "3" in the sequence
> > E^C = No "3" in the sequence
> > Then,
> >
> > $$\begin{aligned} P(E^C) &= P(\text{1st} \neq 3 \cap \text{2nd} \neq 3 \cap \text{3rd} \neq 3 \cap \ldots) \\ &= \prod_{i=1}^{\infty} P(i\text{th} \neq 3) = \frac{5}{6}^{\infty} = 0 \end{aligned}$$
>
> 2. $P($getting a "333" in the sequence of numbers$)$
>   Here we say "333” occurs on the nth toss if the $(n-2)$th outcome is "3”, $(n-1)$th outcome is "3”, and nth outcome is "3”.

Solution: Let F = "333" in the sequence

F^C = No "333" in the sequence

Idea: Create independent blocks of 3 tosses.



Note: F^C implies that block $1 \neq$ "333", block $2 \neq$ "333", block $3 \neq$ "333", $\ldots$

$$\implies F^C \subset \{\text{block } 1 \neq \text{"333"}, \text{block } 2 \neq \text{"333"}, \text{block } 3 \neq \text{"333"}, \ldots\}$$

$$\implies P(F^C) \leq P(\text{block } 1 \neq \text{"333"} \,\&\, \text{block } 2 \neq \text{"333"} \,\&\, \text{block } 3 \neq \text{"333"}, \ldots)$$

$$= \prod_{i=1}^{\infty} P(i\text{th block } \neq \text{"333"})$$

$$= \prod_{i=1}^{\infty} 1 - P(i\text{th block } = \text{"333"})$$

$$= \prod_{i=1}^{\infty} 1 - \left(\frac{1}{6}\right)^3 = \frac{215}{216}^{\infty} = 0$$

Example 1.2:

A coin is continually and independently tossed, where the probability of head (H) on a toss is $1/2$.

Find:

1. $P($1st two tosses give "HH"$)$

> Solution: $P($1st two tosses give "HH"$) = P(H) \cdot P(H) = \frac{1}{4}$

2. $P($1st two tosses give "TH"$)$

> Solution: $P($1st two tosses give "TH”$) = P(T) \cdot P(H) = \frac{1}{4}$

3. $P($"TH" occurs before "HH"$)$

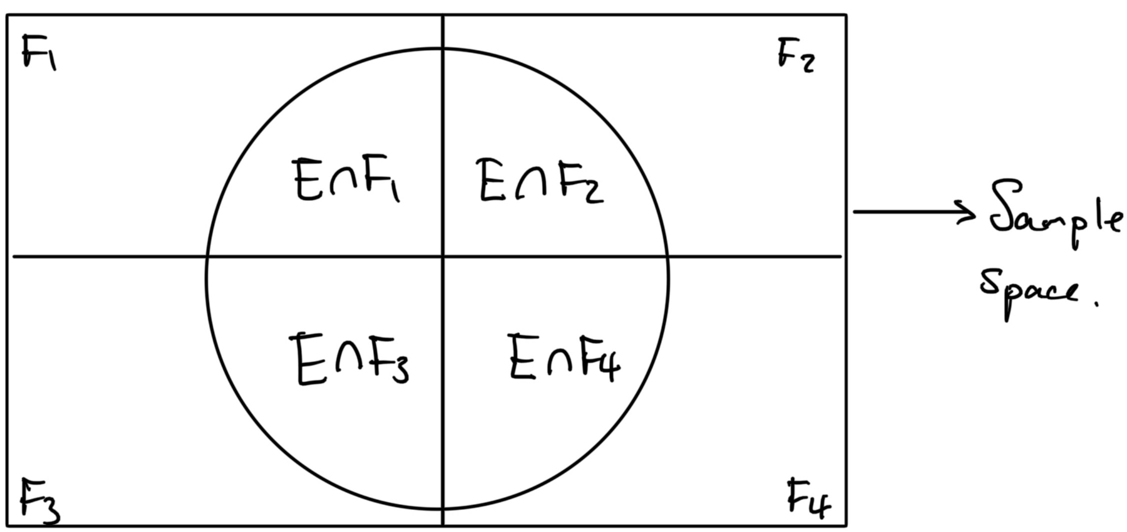
> Solution:
>
> Case 1: Consider 1st outcome is "T", then "TH" occurs before "HH".
>
> Since to observe "TH", we need 1 "H", but to observe "HH", we need 2 "H"s.
>
> $P(\text{Case } 1) = \frac{1}{2}$
>
> Case 2: Consider 1st outcome is "H", 2nd outcome is "T", then "TH" occurs before "HH".
>
> $P(\text{Case } 2) = \frac{1}{4}$
>
> Case 3: Consider 1st outcome is "H", 2nd outcome is "H", then "HH" occurs before "TH".

> $P(\text{Case } 3) = \frac{1}{4}$
> $\implies P(\text{"TH" occurs before "HH"}) = P(\text{Case } 1) + P(\text{Case } 2) = \frac{3}{4}$

- Conditional Probability
  Suppose E and F are two events with $P(F) > 0$. Then

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

  where $P(E|F)$ is the conditional probability, $P(E \cap F)$ is the joint probability, and $P(F)$ is the marginal probability.
  - Result 1: $P(E \cap F) = P(E|F)P(F)$ (Multiplication Rule)
  - Result 2: If E and F are independent, then $P(E|F) = P(E)$, i.e. conditional probability is the same as marginal probability.
    Proof idea: (not required) Use definition of independent and conditional.
  - Bayer's Formula:
    Suppose we have a sequence of events $F_1, F_2, \ldots$ such that:
    a. $P(F_i) > 0$ for all i
    b. $F_i \cap F_k = \emptyset$ for all $i \neq j$
    c. $\bigcup_{i=1}^{\infty} F_i = S$
    Then, for any event E,
    a. $P(E) = \sum_{i=1}^{n} P(E \cap F_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$ (Law of Total Probability)



    b. $P(F_k|E) = \frac{P(E \cap F_k)}{P(E)} = \frac{P(E|F_k)P(F_k)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$ (Bayes' Formula)

> Example 1.3 (Monte Hall Problem):
>
> - There are three doors (say A, B, and C), behind which there are 2 goats and 1 car.

> - Monty knows the location of the car, but you do not.
> - You select a door at random (say A) and at this point your chance of winning the prize is $\frac{1}{3}$.
> - Then Monty opens one of the remaining two doors, either door B or door C, to reveal a goat.
>
> Find $P(\text{winning the car if you switch the door})$.
>
> > Method 1:
> >
> > | | **A** | **B** | **C** | |
> > |---|---|---|---|---|
> > | Case 1 | G | C | G | Monty will open door C |
> > | Case 2 | G | G | C | Monty will open dooor B |
> > | Case 3 | C | G | G | Monty will open either door B or door C |
> >
> > $P(\text{winning if switch}) = P(\text{Case } 1) + P(\text{Case } 2) = \frac{2}{3}$
> >
> > Method 2: Conditional Probability Idea
> > Suppose you choose door A and Monty opens door B (Event E).
> > Let $F_k = \text{car is behind door } k, k = A, B, C$
> > Then, $P(F_A) = P(F_B) = P(F_C) = \frac{1}{3}$
> > $P(\text{win if switch}) = P(F_C|E)$
> > $= \frac{P(E|F_C)P(F_C)}{P(E|F_A)P(F_A)+P(E|F_B)P(F_B)+P(E|F_C)P(F_C)}$
> > $= \frac{1(\frac{1}{3})}{\frac{1}{2}(\frac{1}{3})+1(\frac{1}{3})} = \frac{2}{3}$

1.2 Random Variables (r.v.s)

- Definition of rv
  rv is a function defined on sample space to real line.

$$X : S \rightarrow \mathbb{R}$$

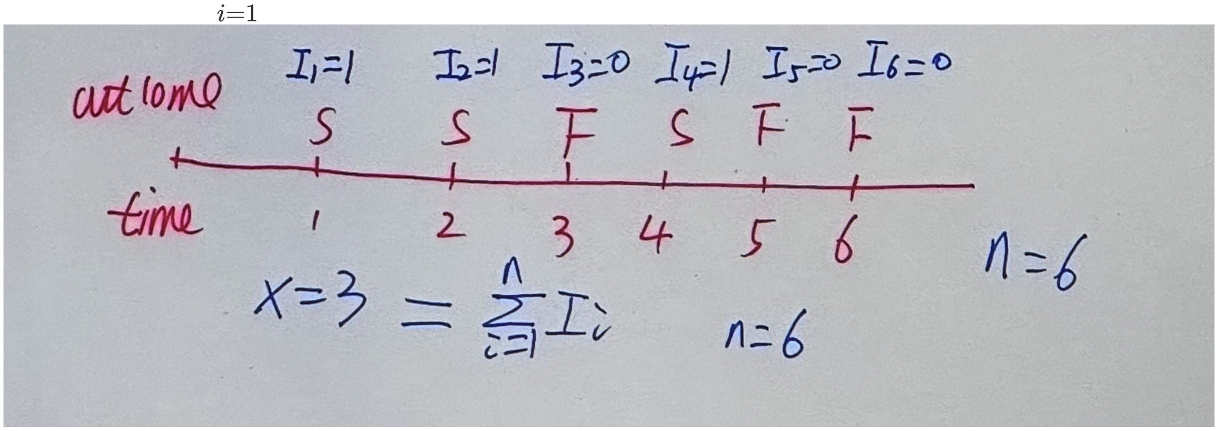
- Two types of rv
  i. Discrete rv: all possible values are at most countable. (e.g. Binomial, Poisson)
  ii. Continuous rv: all possible values contain an interval (e.g. Uniform)
- Review of important rvs:
  - Bernoulli Trials:
    a. Each trial has 2 outcomes: success (S) or failure (F)
    b. All trials are independent
    c. Probability of "S" on each trial is the same:

$$P(\text{S}) = p, P(\text{F}) = 1 - p$$

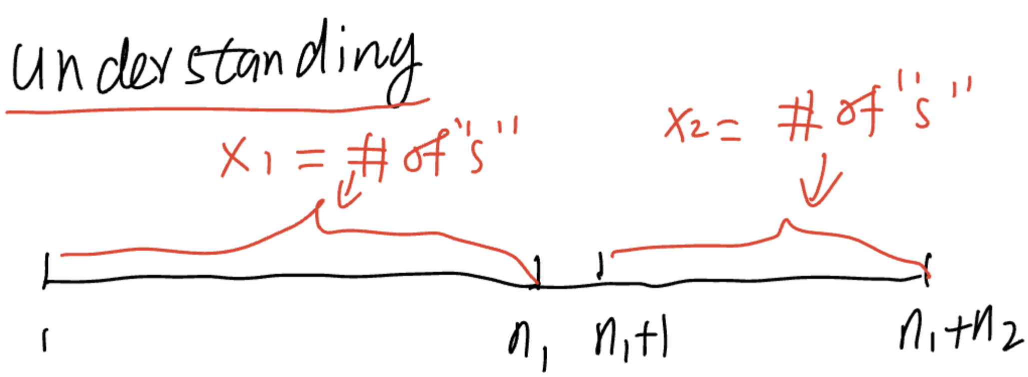
- Bernoulli distribution: $\mathrm{Bernoulli}(p)$
  Let

$$I_i = \begin{cases} 1 & \text{if "s" appears on the } i\text{th trial} \\ 0 & \text{otherwise} \end{cases}$$

  $P(I_i = 1) = p$, $P(I_i = 0) = q = 1 - p$, then $I_1, I_2, \ldots$ are a sequence of iid (independent identically distributed) Bernoulli rvs.
- Binomial rvs: $\mathrm{Bin}(n, p)$
  $X = \#\text{ of "S"s in } n \text{ bernoulli trials} \sim \mathrm{Bin}(n, p)$ where n is the number of trials and p is the probability of success.
  a. Range: $\{0, 1, 2, \ldots, n\}$
  b. Probability mass function (pmf): $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, 2, \ldots, n$
  c. Result 1: $X = \sum_{i=1}^{n} I_i$, where $I_1, .., I_n$ are iid Bernoulli rvs.



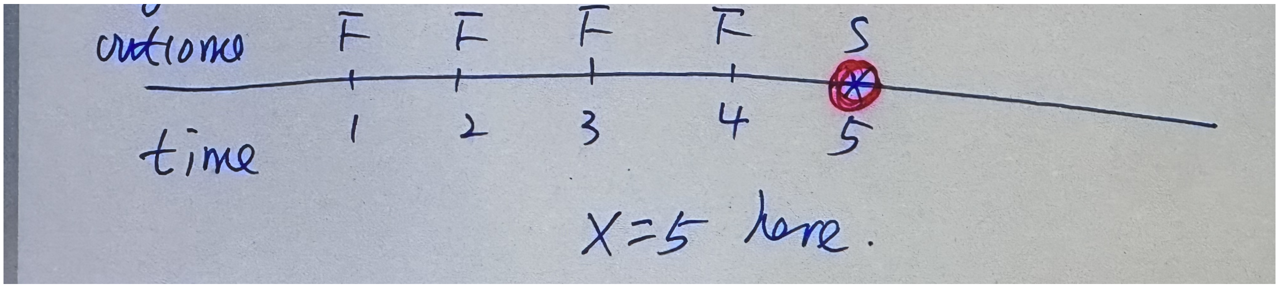
  d. Result 2: If $X_1 \sim \mathrm{Bin}(n_1, p)$ and $X_2 \sim \mathrm{Bin}(n_2, p)$, and both of them are independent, then $X_1 + X_2 \sim \mathrm{Bin}(n_1 + n_2, p)$.



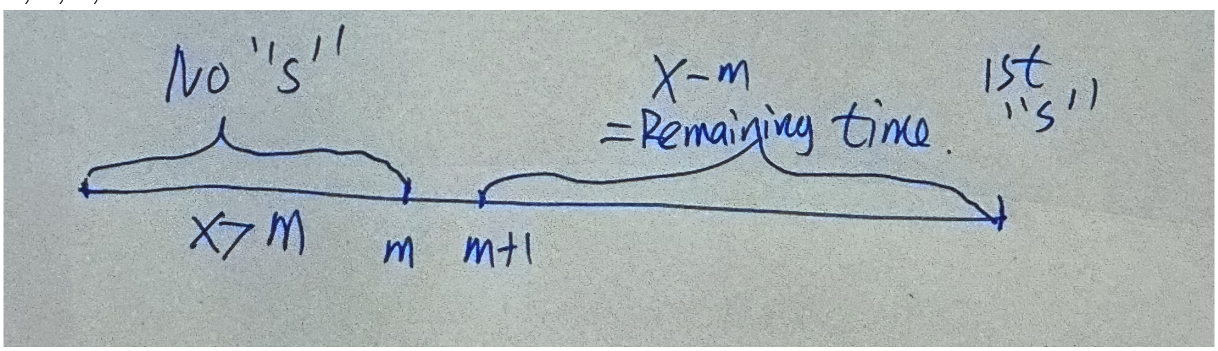
  Then, $x_1 + x_2 =$ # of "s" in $n_1 + n_2$ trials $\sim \mathrm{Bin}(n_1 + n_2, p)$
  Independent: No overlap between first n_1 trials and next n_2 trials.
- Geometric rvs: $\mathrm{Geo}(p)$
  (1st discrete waiting time rv)
  $X = \#\text{ of trials to get 1st "s" in the sequence of Bernoulli trials}$ (including the trial to

observe 1st "s").

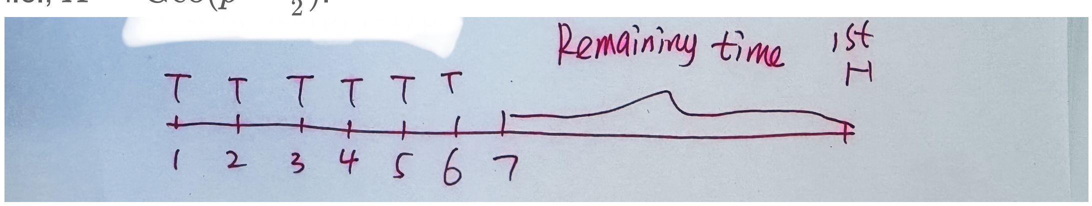
e.g.



a. Range: $\{1,2,3,\ldots\}$
b. pmf: $P(X=k)=(1-p)^{k-1}p$ for $k=1,2,3,\ldots$
   $E(X)=\frac{1}{p}$
c. No-memory property:
   $P(X>n+m|X>m)=P(X-m>n|X>m)=P(X>n)$ for $k,j=1,2,3,\ldots$



Formula tells us: Given that we do not observe Event "s", the remaining time $\sim \operatorname{Geo}(p)$., same as the original time.

> Example 1.4 (Geometric rv example):
>
> A fair coin is tossed repeatedly and independently. The objective is to observe the 1 st head. Let X be the corresponding waiting time. Suppose we get 6 tails in the first 6 tosses.
>
> Note $E(\operatorname{Geo}(p))=\frac{1}{p}$.
>
> > i.e., $X\sim\operatorname{Geo}(p=\frac{1}{2})$.
> >
> > 
> >
>
> Find:
>
> a. $P(X=10|\text{ the first 6 tosses give 6 tails})$.
>
> > Solution:
> >
> > $P(X=10|\text{ the first 6 tosses give 6 tails})$
> > $=P(\text{remainning time})$

> > $= (1-p)^{4-1}p$
> > $= \left(1-\frac{1}{2}\right)^{4-1}\frac{1}{2}$
> > $= \frac{1}{16}$
>
> b. $E(X|\text{ the first 6 tosses give 6 tails})$.
>
> > Solution:
> > $E(X|\text{ the first 6 tosses give 6 tails})$
> > $= E(6 + \text{remainning time})$
> > $= 6 + \frac{1}{p}$
> > $= 8$

- Negtive Binomial $\text{NegBin}(r, p)$
  a. Range: $\{r, r+1, r+2, \ldots\}$
  b. pmf: $P(X=k) = \binom{k-1}{r-1}p^r(1-p)^{k-r}$ for $k = r, r+1, r+2, \ldots$ (not required)
  c. Property: Let $X_1 = \text{Waiting time to observe 1st "s"}$,
  $X_2 = \text{Waiting time to observe 2nd "s" after 1st "s"}$,
  $\cdots$,
  $X_r = \text{Waiting time to observe } r\text{th "s" after } (r-1)\text{th "s"}$
  $X = \sum_{i=1}^{r} X_i \sim \text{NegBin}(r,p)$: $X_1, \cdots, X_r$ are iid $\text{Geo}(p)$ rvs.

> Example 1.5 (Negative Binomial rv example)
> A fair coin is tossed repeatedly and independently. The objective is to observe the two heads in total. Let X be the corresponding waiting time.
> Note $E(\text{Geo}(p)) = \frac{1}{p}$.
> Find:
>
> a. $E(X|\text{the first 3 tosses give "HTT"})$.
>
> > Solution:
> > $E(X|\text{the first 3 tosses give "HTT"})$
> > $= E(3 + \text{remainning time})$
> > $= 3 + \frac{1}{p}$
> > $= 3 + 2$
> > $= 5$
>
> b. $E(X|\text{the first 3 tosses give "TTT"})$.
>
> > Solution:
> > $E(X|\text{the first 3 tosses give "TTT"})$
> > $= E(3 + \text{remainning time of } X_1 + X_2)$
> > $= 3 + \frac{1}{p} + \frac{1}{p}$
> > $= 7$

- Poisson r.v. $\mathrm{Pois}(\lambda)$
  Suppose $X \sim \mathrm{Pois}(\lambda)$
  a. Range: $\{0, 1, 2, \ldots\}$
  b. pmf: $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ for $k = 0, 1, 2, \ldots$
     $\lambda =$ rate parameter / $E(X) = \lambda$
  c. Property:
     If $X_1 \sim \mathrm{Pois}(\lambda_1)$ and $X_2 \sim \mathrm{Pois}(\lambda_2)$, and both of them are independent, then $X_1 + X_2 \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$.
- Exponential r.v.: $\mathrm{Exp}(\lambda)$
  [continuous waiting time rv]
  a. probability density function (pdf): $f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{o.w.} \end{cases}$
  b. $\lambda =$ rate parameter / $E(X) = \frac{1}{\lambda}$
  c. Tail prob: $P(X > t) = e^{-\lambda t} = \int_t^\infty \lambda f(x) dx$ for $t > 0$
  d. No memory property: $P(X > t + s | X > s) = P(\text{remaining time} > t | X > s) = P(X > t)$ for $s, t > 0$
     Meaning: Given that we do not observe Event "s", the remaining time $\sim \mathrm{Exp}(\lambda)$, same as the original time.

> Example 1.6 (Exponential rv example)
> Suppose waiting time (X) for customers coming to Tim Hortons (T.H.) follows $\mathrm{Exp}(2)$. Here 1 unit time = 1 minute. In the first 3 minutes, there is no customer.
>
> > $\lambda = 2, X > 3.$
>
> (a) Find the probability of no customer in the first 5 minutes.
>
> > Solution: $P(X > 5 | X > 3) = P(X > 2) = e^{-4}$
>
> (b) $E(X | \text{no customer in the first 3 minutes})$.
>
> > Solution: $E(X | \text{no customer in the first 3 minutes})$
> > $= E(3 + \text{remainning time})$
> > $= 3 + \frac{1}{\lambda}$
> > $= 3 + \frac{1}{2}$
> > $= 3.5$

1.3 Expectation and Variance

- Expectation of Discrete r.v.
  Range: $\{x_1, x_2, \ldots\}$
  $E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$ = value * probability

> e.g. $X \sim \text{Bernoulli}(p)$, Find $E(X)$.
> Solution: $P(X=1)=p$, $P(X=0)=1-p$, $E(X)=1\cdot p+0\cdot(1-p)=p$

- Expectation of Continuous r.v.
  x has pdf $f(x)$
  $E(X)=\int_{-\infty}^{\infty} xf(x)dx$
- General case: $g(x)$ is a function of x.
  $$E(g(X))=\begin{cases}\sum_{i=1}^{\infty} g(x_i)P(X=x_i) & \text{if } X \text{ is discrete}\\ \int_{-\infty}^{\infty} g(x)f(x)dx & \text{if } X \text{ is continuous}\end{cases}$$
  where $g(X)$ is a new r.v.
  > e.g. $X \sim \text{Bernoulli}(p)$, Find $E(X^2)$.
  >
  > > Solution: Consider $g(x)=x^2$, then
  > >
  > > $$E(X^2)=0^2P(X-0)+1^2P(X=1)=0^2\cdot(1-p)+1^2\cdot p=p=E(X)$$
  > >
  > > Therefore, $X^2=X$ for $X \sim \text{Bernoulli}(p)$.

- Variance
  $Var(X)=E(X^2)-[E(X)]^2$
  > e.g. $X \sim \text{Bernoulli}(p)$, Find $Var(X)$.
  >
  > > Solution: $Var(X)=E(X^2)-[E(X)]^2=p-p^2=p(1-p)=P(X=1)*P(X=0)$

- Properties:
  i. $E(\sum_{i=1}^{n} a_iX_i)=\sum_{i=1}^{n} a_iE(X_i)$ [linearity]
     $a_1,\cdots,a_n$ are constants, $X_1,\cdots,X_n$ are r.v.s
  ii. If $X_1,\cdots,X_n$ are independent, then $Var(\sum_{i=1}^{n} a_iX_i)=\sum_{i=1}^{n} Var(a_iX_i)=\sum_{i=1}^{n} a_i^2Var(X_i)$
     [Aside: $Var(aX+b)=Var(aX)=a^2Var(X)$]
     > e.g. If X_1 and X_2 are independent, then
     > $Var(X_1+X_2)=Var(X_1)+Var(X_2)$, and
     > $Var(X_1-X_2)=Var(X_1)+Var(X_2)$

  iii. In general:
     $Var(\sum_{i=1}^{n} a_iX_i)=\sum_{i=1}^{n} a_i^2Var(X_i)+\sum_{i\neq j} a_ia_jCov(X_i,X_j)$
     > Aside: $Cov(X,Y)=E(XY)-E(X)E(Y)$, and if X and Y are independent, then $Cov(X,Y)=0$.
     > Question: If $Cov(X,Y)=0$, are X and Y independent?

$Var(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 Var(X_i) + 2\sum_{i<j} a_i a_j Cov(X_i, X_j)$

> e.g.
> $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$, and
> $Var(X_1 + X_2 + X_3) = Var(X_1) + Var(X_2) + Var(X_3) + 2Cov(X_1, X_2) + 2Cov(X_1, X_3) + 2Cov(X_2 + X_3)$

1.4 Indicator r.v.

- Indicator r.v.: Only two possible values 0 and 1
  For a given event A, we define $I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$
  Suppose $P(A) = 0$, then $P(I_A = 1) = p$, $P(I_A = 0) = 1 - p = q$.
  By Bernoulli distribution, $E(I_A) = P(I_A = 1) = p$ and $Var(I_A) = P(I_A = 1)P(I_A = 0) = p(1-p)$

  > e.g. Suppose $X \sim \text{Bin}(n, p)$, find $E(X)$ and $Var(X)$
  >
  > > Solution: (1) Def, (2)Mgf
  > > $X = \sum_{i=1}^n I_i : I_1, \cdots, I_n \text{ are iid Bernoulli}(p) \text{ r.v.}$
  > > Therefore $E(X) = \sum_{i=1}^n E(I_i) = np$ [linearlity], and
  > > $Var(X) = Var\left(\sum_{i=1}^n I_i\right) = \sum_{i=1}^n Var(I_i) = np(1-p)$ [$I_1, \cdots, I_n$ are independent]

> Example 1.7
>
> - Suppose we have two boxes: red box and black box.
> - In red box, there are 4 red balls and 6 black balls.
> - In the black box, there are 6 red balls and 4 black balls.
> - An experiment is conducted as follows.
>   - At the first step, a fair coin is tossed.
>   - If it shows head, then you randomly choose a ball in the red box, record the color and put it back; else, you randomly choose a ball in the black box, record the color and put it back.
>   - At the second step, you randomly choose a ball from the box, which has the same color as the ball you chosen at the first step.
>
> Let X be the number of red balls you draw in the first two steps. Find E(X) and Var(X).
>
> > Solution: $X \in \{0, 1, 2\}$. Find $P(X = 0), P(X = 1), P(X = 2)$.
> > Let $I_i = \begin{cases} 1 & \text{if we get } R \text{ ball in } i\text{th step} \\ 0 & \text{otherwise} \end{cases}$, then $X = I_1 + I_2$
> > $P(X_1 = 1) = P(\text{1st ball} = R) = P(\text{1st ball} = R|H)P(H) + P(\text{1st ball} = R|T)P(T) = 0.4 \cdot \frac{1}{2} + 0.6 \cdot \frac{1}{2} \cdot \frac{1}{2} = 0.5$
> > $P(X_2 = 1) = P(\text{2nd ball} = R) = P(\text{2nd ball} = R|H)P(H) + P(\text{2nd ball} = R|T)P(T) = 0.4 \cdot \frac{1}{2} + 0.6 \cdot \frac{1}{2} \cdot \frac{1}{2} = 0.5$

> Therefore, $E(X)=E(I_1)+E(I_2)=P(I_1=1)+P(I_2=2)=0.5+0.5=1$
> Here, $I_1I_2=\begin{cases}1 & I_1=I-2=1 \text{ or } 1\text{st}=R \text{ and } 2\text{nd}=R\\ 0 & \text{otherwise}\end{cases}$, therefore, it is also an indicator r.v.
> $E(I_1I_2)=P(I_1I_2=1)=P(1\text{st}=R \text{ and } 2\text{nd}=R)$
> $=P(2\text{nd}=R|1\text{st}=R)P(1\text{st}=R)=0.4(0.5)=0.2$
> $Cov(I_1,I_2)=E(I_1I_2)-E(I_1)E(I_2)=0.2-0.5^2=-0.05$
> $Var(X)=Var(I_1+I_2)=Var(I_1)+Var(I_2)+Cov(I_1,I_2)$
> $=P(I_1=1)P(I_1=0)+P(I_2=1)P(I_2=0)=2(0.5)^2+2(-0.05)=0.4$

Chapter 2: Waiting time r.v.

- Background: Suppose we have a sequence of trials and we would like to observe an event E based on the sequence of trials.
  Let T_E = # of trials or waiting time to observe first E [including the trial to observe E]
  Range: $\{1,2,3,\dots\}\cup\{\infty\}$ where ∞ means E never occurs.
- Classification of T_E
  i. If $P(T_E<\infty)<1$ or $P(T_E<\infty)=0$, then T_E is improper.
     Note: In this case, $E(T_E)=\infty$ since $P(T_E<\infty)>0$.
  ii. If $P(T_E<\infty)=1$ or $P(T_E<\infty)=0$, then T_E is proper.
     a. If $P(T_E<\infty)=1$ and $E(T_E)=\infty$, then T_E is null proper.
     b. If $P(T_E<\infty)=1$ and $E(T_E)<\infty$, then T_E is short proper.
  - Comments
     a. If T_E is improper, then $E(T_E)=\infty$.
     b. If $E(T_E)<\infty$, then T_E is short proper.
  - Aside: $\sum_{n=1}^{\infty}a_n$ do not include ∞ in the summation.
    Recall $\sum_{n=1}^{\infty}a_n=\lim_{n\to\infty}\sum_{i=1}^{n}a_i$.
    Hence $P(T_E<\infty)=\sum_{n=1}^{\infty}P(T_H=n)$

> Example 2.1 (Short proper waiting time rv)
> Suppose we toss a coin repeatedly and independently. At each toss, the probability of getting "H" is p with $0<p<1$. Let T_H be the waiting time for the first "H".
> Claim: T_H is a short proper waiting time random variable.
>
>> Solution:
>> $T_H\sim \text{Geo}(p)$
>> Then $P(T_H=n)=(1-p)^{n-1}p, n\geq 1$
>> $P(T_H<\infty)=\sum_{n=1}^{\infty}(1-p)^{n-1}p-\frac{p}{1-(1-p)}=1$
>> When $T_H\sim \text{Geo}(p)$, E(T_H)=\frac{1}{p}<\infty$Therefore,$T_H$ is a short proper waiting time random variable.

> Example 2.2 (Null proper waiting time rv)
> Suppose we toss a coin repeatedly and independently. For $n=1,2,3,\cdots$ the probability of getting "H" is $\frac{1}{n+1}$ at the
> nth toss.
> Let T_H be the waiting time for the first "H".
> Claim: T_H is a null proper waiting time random variable.
>
> > Solution:
> > $P(T_H=n)=P(\text{1st time to observe } H \text{ is } n)=P(\text{1st}=T,\text{2nd}=T,\cdots,(n-1)\text{th}=T,n\text{th}=H)$
> > $=P(\text{1st}=T)P(\text{2nd}=T)\cdots P((n-1)\text{th}=T)P(n\text{th}=H)$
> > $=\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4}\cdots\frac{n-1}{n}\cdot\frac{1}{n+1}=\frac{1}{n(n+1)}$
> > $P(T_H<\infty)=\sum_{n=1}^{\infty}\frac{1}{n(n+1)}=\sum_{n=1}^{\infty}\left(\frac{1}{n}-\frac{1}{n+1}\right)=(1-\frac{1}{2})+(\frac{1}{2}-\frac{1}{3})+\cdots=1$
> > $E(T_H)=\sum_{n=1}^{\infty}nP(T_H=n)=\sum_{n=1}^{\infty}n\frac{1}{n(n+1)}=\infty$

> Example 2.3 (Improper waiting time rv)
> Suppose we toss a coin repeatedly and independently. For $n=1,2,3,\cdots$, the probability of getting "H" is 2^{-n} at the nth toss. Let T_H be the waiting time for the first "H".
> Claim: T_H is an improper waiting time random variable.
>
> > Solution:
> > $P(T_H=1)=P(\text{1st}=H)=\frac{1}{2}$
> > $P(T_H=2)=P(\text{1st}=T,\text{2nd}=H)=P(\text{1st}=T)P(\text{2nd}=H)\frac{1}{2^3}=\frac{1}{8}$
> > In general,
> > $P(T_H=n)=P(\text{1st}=T,\text{2nd}=T,\cdots,(n-1)\text{th}=T,n\text{th}=H)$
> > $=P(\text{1st}=T)P(\text{2nd}=T)\cdots P((n-1)\text{th}=T)P(n\text{th}=H)\leq\frac{1}{2^{n+1}}$ for $n>2$
> > $P(T_H<\infty)=\sum_{n=1}^{\infty}P(T_H=n)=P(T_H=1)+\sum_{n=2}\infty P(T_H=n)\leq 2^{-(n+1)}\leq\frac{1}{2}+\sum_{n=2}^{\infty}\frac{1}{2^{n+1}}=\frac{1}{2}+\frac{1/8}{1-\frac{1}{2}}=\frac{1}{2}+\frac{1}{4}=\frac{3}{4}<1$

Chapter 3: Conditional Expectation

3.1 Joint rvs:

We consider 2 r.v.s

- Joint discrete r.v.s
  If X and Y are discrete r.v.s, then (X,Y) are joint discrete.
- Joint pmf:
  $f_{X,Y}(x,y)=P(X=x,Y=y)$
- Properties:
  i. Joint pmf is a pmf: $f_{X,Y}(x,y)\geq 0$ and $\sum_x\sum_y f_{X,Y}(x,y)=1$

ii. Marginal pmf of X: $f_X(x) = P(X = x) = \sum_y f_{X,Y}(x,y)$
Marginal pmf of Y: $f_Y(y) = P(Y = y) = \sum_x f_{X,Y}(x,y)$
iii. Expectation: $h(x,y)$ is a bivariate function
$E(h(X,Y)) = \sum_x \sum_y h(x,y) f_{X,Y}(x,y)$

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

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

- Joint continuous r.v.s
  If X and Y are continuous r.v.s and $P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)dtds$, then (X,Y) are joint continuous and $f_{X,Y}(x,y)$ is called joint pdf.
- Properties:
  i. $f_{X,Y}(x,y) \ge 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$
  ii. Mariginal pdf of X: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
  Mariginal pdf of Y: $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$
  iii. Expectation: $h(x,y)$ is a bivariate function
  $E(h(X,Y)) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(x,y) f_{X,Y}(x,y)dy\right] dx$

> E.g. $E(XY) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)dy\right] dx$

> E.g. $E(X) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dy\right] dx = \int_{-\infty}^{\infty} x f_X(x)dx$

- Independence: Both continuous and discrete
  If $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, then X and Y are independent.
  i.e. Joint probability = product of marginal probabilities
- Properties of Independence:
  Ifd X and Y are independent, then $h(X)$ and $g(Y)$ are independent.
  $E(h(X)g(Y)) = E(h(X))E(g(Y))$
  $E(XY) = E(X)E(Y)$ and $Cov(X,Y) = 0$
  Note: If $Cov(X,Y) = 0$, X and Y are not necessarily independent.

> E.g. $X \sim \text{Unif}[-1,1]$, $f_X(x) = \begin{cases} \frac{1}{2} & -1 < x < 1 \\ 0 & \text{o.w.} \end{cases}$ and $Y = X^2$
> then $E(X) = \int_{-1}^{1} x f_X(x)dx = \int_{-1}^{1} \frac{1}{2}xdx = 0$
> $E(XY)) = \int_{-1}^{1} xy f_X(x)dx = \int_{-1}^{1} \frac{1}{2}x^3 dx = 0$
> then $E(XY) = E(X)E(Y)$, $Cov(X,Y) = 0$, but X and $Y = X^2$ are not independent.

3.2 Conditional distribution and conditional expectation

- Discrete case: X and Y
  Joint pmf: $f_{X,Y}(x,y)$
  Marginal pmf: $f_X(x)$, $f_Y(y)$
  - Conditional pmf:
    For a given y such that $f_Y(y) > 0$, the conditional pmf of X given $Y = y$ is defined as

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\text{Joint}}{\text{Marginal}}$$

- Property of conditional pmf:
  Conditional pmf is a pmf, i.e.,
  a. $f_{X|Y}(x|y) \geq 0$
  > Proof: Directly follows definition.

  b. $\sum_x f_{X|Y}(x|y) = 1$
  > Proof: $\sum_x f_{X|Y}(x|y) = \sum_x \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{f_Y(y)} \sum_x f_{X,Y}(x,y) = \frac{1}{f_Y(y)} f_Y(y) = 1$

- Conditional expectation:
  Conditional expectation of X given $Y = y$ is defined as

$$E(X|Y=y) = \sum_x x f_{X|Y}(x|y)$$

  Conditional expectation of $g(X)$ given $Y = y$ is defined as

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

- Independence property:
  If X and Y are independent, then $E(X|Y=y) = E(X)$ and $E(g(X)|Y=y) = E(g(X))$
  > Key: If X and Y are independent, then $f_{X|Y}(x|y) = f_X(x)$.

  > Proof: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x)$
  > $\implies E(X|Y=y) = \sum_x x f_{X|Y}(x|y) = \sum_x x f_X(x) = E(X)$

  Intuitively: Independence makes "Conditional" become "Unconditional".

> Example 3.1
> Suppose $X_1 \sim \text{Pois}(\lambda_1)$ and $X_2 \sim \text{Pois}(\lambda_2)$, and X_1 and X_2 are independent.
> Let $X = X_1, Y = X_1 + X_2$.
> Find:
>
> i. $f_{X|Y}(x|y)$
> > Solution: $Y \sim \text{Pois}(\lambda_1 + \lambda_2)$, therefore $f_Y(y) = \frac{(\lambda_1+\lambda_2)^y e^{-(\lambda_1+\lambda_2)}}{y!}$
> > We need $P(X = x, Y = y)$.
> > $P(X=x, Y=y) = P(X=x, X_1+X_2=y) = P(X_1=x, X_2=y-x) =$
> > $P(X_1=x)P(X_2=y-x) = \frac{\lambda_1^x e^{-\lambda_1}}{x!} \cdot \frac{\lambda_2^{y-x} e^{-\lambda_2}}{(y-x)!}$
> > Therefore, $f_{X|Y}(x|y) = \frac{P(X=x,Y=y)}{P(Y=y)} = \frac{\frac{\lambda_1^x e^{-\lambda_1}}{x!} \cdot \frac{\lambda_2^{y-x} e^{-\lambda_2}}{(y-x)!}}{\frac{(\lambda_1+\lambda_2)^y e^{-(\lambda_1+\lambda_2)}}{y!}} = \binom{y}{x} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{y-x}$ where

> > $x = 0, 1, 2, \cdots, y.$
> > $X|Y = y \sim \text{Bin}(y, \frac{\lambda_1}{\lambda_1+\lambda_2})$
>
> ii. $E(X|Y = y)$ for some positive integer y.
>
> > Solution: $E(X|Y = y) = y\frac{\lambda_1}{\lambda_1+\lambda_2}$

- Continuous case: X and Y
  Joint pdf: $f_{X,Y}(x, y)$
  Marginal pdf: $f_X(x)$, $f_Y(y)$
  - Conditional pdf:
    For a given y such that $f_Y(y) > 0$, the conditional pdf of X given $Y = y$ is defined as

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{\text{Joint}}{\text{Marginal}}$$

  - Property of conditional pdf:
    Conditional pdf is a pdf, i.e.,
    a. $f_{X|Y}(x|y) \geq 0$
    b. $\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = 1$
  - Conditional expectation:
    Conditional expectation of X given $Y = y$ is defined as

$$E(X|Y = y) = \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx$$

    Conditional expectation of $g(X)$ given $Y = y$ is defined as

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

  - Independence property:
    If X and Y are independent, then $E(X|Y = y) = E(X)$ and $E(g(X)|Y = y) = E(g(X))$

    > Key: If X and Y are independent, then $f_{X|Y}(x|y) = f_X(x)$.

> Example 3.2
> Suppose X and Y has the joint pdf
>
> $$f_{X,Y}(x, y) = \begin{cases} xe^{-xy} & x > 0, y > 1 \\ 0 & \text{o.w.} \end{cases}$$
>
> Find:
>
> i. $f_{X|Y}(x|y)$

> > Solution:
> > $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ and
> > $\Gamma(n) = (n-1)!$
> > Let $t = xy$, then $x = \frac{t}{y}$ and $dx = \frac{dt}{y}$
> > Then, $f_Y(y) = \int_0^\infty xe^{-xy}dx = \int_0^\infty \frac{t}{y}e^{-t}\frac{dt}{y} = \frac{1}{y^2}\Gamma(2) = \frac{1}{y^2}, y > 1$
> > $f_{X,Y}(x,y) = \frac{xe^{-xy}}{f_Y(y)} = \frac{xe^{-xy}}{\frac{1}{y^2}} = y^2xe^{-xy}$
>
> ii. $E(X|Y=y)$ for $y > 1$.
>
> > Solution: $E(X|Y=y) = \int_0^\infty xf_{X|Y}(x|y)dx = \int_0^\infty xy^2xe^{-xy}dx = \int_0^\infty (xy)^2e^{-xy}dx = \int_0^\infty te^{-t}\frac{dt}{y} = \frac{1}{y}\Gamma(3) = \frac{2}{y}$

- Summary of properties of conditional expectation
  (Apply to both discrete and continuous cases)
  i. Conditional expectation has properties of expectation.
  > e.g. $E\left[\sum_{i=1}^n a_iX_i|Y=y\right] = \sum_{i=1}^n a_iE(X_i|Y=y)$

  ii. Substitution Rule:
  $E[Xg(Y)|Y=y] = E[Xg(y)|Y=y] = g(y)E[X|Y=y]$.
  Here, $g(Y)$ is a r.v., but $g(y)$ is a constant.
  > e.g. $E[XY|Y=y] = yE[X|Y=y]$
  > In general, $E[h(X,Y)|Y=y] = E[h(X,y)|Y=y]$

  iii. Independence Property
  If X and Y are independent, then $E(X|Y=y) = E(X)$ and $E(g(X)|Y=y) = E(g(X))$

3.3 Calculating expectation by conditioning

This section: We cover $E(X) = E[E(X|Y)]$, double expectation theorem and law of total expectation.

- What is $E(X|Y)$?
  It is a r.v., depends on Y or a function of Y, i.e. $E(X|Y) = g(Y)$.
- What is the funtion g(y)?
  Function $g(y)$ is $g(y) = E(X|Y=y)$ covered in 3.2.
  E.g. in Example 3.1, $E(X|Y=y) = y\frac{\lambda_1}{\lambda_1+\lambda_2} : g(y) = y\frac{\lambda_1}{\lambda_1+\lambda_2}$
  E.g. in Example 3.2, $E(X|Y=y) = \frac{2}{y} : g(y) = \frac{2}{y}$
- How to obtain $E(E|Y)$
  - Step 1: Figure out $g(y) = E(X|Y=y)$
    Either by definition or by properties.
  - Step 2: $E(X|Y) = g(Y)$

  > E.g. In example 3.1, $E(X|Y) = Y\frac{\lambda_1}{\lambda_1+\lambda_2} \implies E(X|Y) = g(Y) = Y\frac{\lambda_1}{\lambda_1+\lambda_2}$

  > E.g. In example 3.2, $E(X|Y) = \frac{2}{Y} \implies E(X|Y) = g(Y) = \frac{2}{Y}$

Note: $E(X) = E(g(Y)) \neq E[E(X|Y=y)] = E(g(y))$

- How to apply $E(X) = E[E(X|Y)]$?

$$E(X) = E(g(Y)) = \begin{cases} \sum_y g(y) f_Y(y) & \text{discrete } Y \\ \int_{-\infty}^{\infty} g(y) f_Y(y) dy & \text{continuous } Y \end{cases}$$

$$\implies E(X) = E(g(Y)) = E[E(X|Y)] = \begin{cases} \sum_y E(X|Y=y) f_Y(y) & \text{discrete } Y \\ \int_{-\infty}^{\infty} E(X|Y=y) f_Y(y) dy & \text{continuous } Y \end{cases}$$

Key: $E(X|Y=y) = g(y)$

- Why $E(X) = E(g(Y)) = \begin{cases} \sum_y g(y) f_Y(y) & \text{discrete } Y \\ \int_{-\infty}^{\infty} g(y) f_Y(y) dy & \text{continuous } Y \end{cases}$

We concerntrate on discrete case.
LHS: $E(X) = \sum_x \sum_y x f_{X,Y}(x,y)$
RHS: $\sum_y E(X|Y=y) f_Y(y) = \sum_y \sum_x x f_{X|Y}(x|y) f_Y(y) = \sum_y \sum_x x f_{X,Y}(x|y) = \sum_x \sum_y x f_{X,Y}(x,y)$

> Example 3.3
> Let $Y_1, Y_2, \ldots$ be independently distributed random variables such that for $n \geq 1$,
>
> $$Y_n \sim \text{Pois}(n)$$
>
> That is, $Y_1 \sim \text{Pois}(1)$, $Y_2 \sim \text{Pois}(2)$, $Y_3 \sim \text{Pois}(3)$, and so on.
> Assume that $N \sim \text{Geo}(0.5)$ and it is independent of $Y_1, Y_2, \ldots$.
> Let $X \sim Y_N$. Find $E(X)$.
>
> > Solution: $E(X) = E(E(X|N)) = \sum_{n=1}^{\infty} E(X|N=n) P(N=n)$
> > $P(N=n) = (1-p)^{n-1} p = \frac{1}{2^n}$, $n = 1, 2, \ldots$.
> > $E(X|N=n) = E(Y_N|N=n) = E(Y_n|N=n) = E(Y_n) = n$, since $Y_n \sim \text{Pois}(n)$.
> >
> > > Method 1: $E(X) = \sum_{n=1}^{\infty} n(\frac{1}{2}^n = E(N) = \frac{1}{0.5} = 2$
> >
> > > Method 2: $g(n) = n \implies g(N) = N \implies E(X) = E(N) = 2$

> Example 3.4
> Suppose we toss a coin repeatedly and independently. At each toss, the probability of getting "H" is with $0 < p < 1$. Let X be the waiting time for the first "H".
> Show that $E(X) = 1/p$.
>
> > Solution: $X \sim \text{Geo}(p)$.
> > Let $Y = \begin{cases} 1 & \text{if 1st outcome } = H \\ 0 & \text{if 1st outcome } = T \end{cases}$.
> > Let R be Remainning time to observe 1st H. Then, E(X) = E(R).

> > $E(X)=\sum_y E(X|Y=y)P(Y=y)=E(X|Y=1)p+E(X|Y=0)(1-p)$
> > $\implies E(X)=1p+(E(R)+1)(1-p)=p+(1-p)+(1-p)E(R)=1+(1-p)E(X)$
> > $\implies E(X)=\frac{1}{p}$.

> Example 3.5
>
> - A miner is trapped. There are 3 doors.
>   - Door 1 leads to safety after 2 hrs.
>   - Door 2 returns the miner to the starting point in 3 hrs.
>   - Door 3 leads the miner to starting points after 4 hrs.
> - Assuming the miner randomly chooses a door at each time.
> - Let X be the length of time until the miner gets out.
>   Find $E(X)$.
>
> > Solution: Let Y be the door # miner chosen at the first time.
> > Then, $P(Y=1)=P(Y=2)=P(Y=3)=\frac{1}{3}$.
> > Here, $X|Y=1=2$, $X|Y=2=3+R$, $X|Y=3=4+R$, where R is the remainning time that miner needs to spend before safety after he returns to starting point
> > Note: X and R have the same distribution, i.e. $E(X)=E(R)$.
> > Then, $E(X)=\sum_y E(X|Y=y)P(Y=y)=E(X|Y=1)P(Y=1)+E(X|Y=2)P(Y=2)+E(X|Y=3)P(Y=3)=2\frac{1}{3}+(3+E(R))\frac{1}{3}+(4+E(R))\frac{1}{3}=\frac{1}{3}(2+3+4)+\frac{2}{3}E(R)=\frac{1}{3}(2+3+4)+\frac{2}{3}E(X)$
> > $\implies E(X)=2+3+4$.

3.4 Computing probabilities by conditioning

- Suppose A is an event and we need P(A)
- Let $I_A=\begin{cases}1 & \text{if } A \text{ occurs}\\ 0 & \text{if } A \text{ does not occur}\end{cases}$
  $\implies P(A)=E(I_A)=\begin{cases}\sum_y E(I_A|Y=y)f_Y(y) & \text{discrete } Y\\ \int_{-\infty}^{\infty}E(I_A|Y=y)f_Y(y)dy & \text{continuous } Y\end{cases}$
  Note: $E(I_A|Y=y)=P(I_A=1|Y=y)=P(A|Y=y)$
  $\implies P(A)=\begin{cases}\sum_y P(A|Y=y)f_Y(y) & \text{discrete } Y\\ \int_{-\infty}^{\infty}P(A|Y=y)f_Y(y)dy & \text{continuous } Y\end{cases}$

> Example 3.6 If X_1, X_2, X_3 are iid random variables from uniform distribution on $[0,1]$.
> (a) Find $P(X_1<X_2)=P(X_1=\min(X_1,X_2))$;
>
> > Solution: $Y=X_2$,
> > $P(X_1<X_2)=\int_{-\infty}^{\infty}P(X_1<X_2|Y=X_2=y)f_{X_2}(y)dy=\int_1^0 P(X_1<X_2|Y=X_2=$

> $y)1dy = \int_0^1 P(X_1 < y|Y = X_2 = y)dy = \int_0^1 P(X_1 < y)dy$
> $P(X_1 < y) = \int_0^y f_{X_1}(x)dx = \int_0^y 1dx = y$
> $P(X_1 < X_2) = \int_0^1 ydy = \frac{1}{2}$

(b) Find P(X_1 < X_2 < X_3).

> Solution: We consider conditioning on on X_2.
>
> $$\begin{aligned} P(X_1 < X_2 < X_3) &= \int_{-\infty}^{\infty} P(X_1 < X_2 < X_3|X_2 = y)f_{X_2}(y)dy \\ &= \int_0^1 P(X_1 < X_2 < X_3|X_2 = y)dy \text{ (as } f_{X_2}(y) = 1) \\ &= \int_0^1 P(X_1 < y < X_3|X_2 = y)dy \text{ (by substitution rule)} \\ &= \int_0^1 P(X_1 < y < X_3)dy \text{ (since } X_1, X_3 \text{ and } X_2 \text{ are independent)} \\ &= \int_0^1 P(X_1 < y, y < X_3)dy \\ &= \int_0^1 P(X_1 < y)P(y < X_3)dy \text{ (since } X_1 \text{ and } X_3 \text{ are independent)} \end{aligned}$$
>
> Here, $P(X_1 < y) = \int_0^y f_{X_1}(x)dx = \int_0^y 1dx = y$ and $P(y < X_3) = \int_y^1 f_{X_3}(x)dx = \int_y^1 1dx = 1 - y$
> Hence, $P(X_1 < X_2 < X_3) = \int_0^1 y(1-y)dy = \frac{1}{6}$

> Example 3.7
> Suppose
>
> - Y has pdf $f(y) = \begin{cases} ye^{-y} & y > 0 \\ 0 & y \le 0 \end{cases}$.
> - $X|Y = y \sim \text{Pois}(y)$
>   Find $P(X = n)$.
>
> > Solution: $P(X = n) = \int_{-\infty}^{\infty} P(X = n|Y = y)f_Y(y)dy = \int_0^{\infty} P(X = n|Y = y)f_Y(y)dy$
> > Here, $X|Y = y \sim \text{Poiss}(y)$ \implies $P(X = n|Y = y) = \frac{y^n e^{-y}}{n!}$
> > Thus, $P(X = n) = \int_0^{\infty} \frac{y^n e^{-y}}{n!} ye^{-y}dy = \int_0^{\infty} \frac{y^{n+1}e^{-2y}}{n!}dy$
> > Let $t = 2y$, then $y = \frac{t}{2}$ and $dy = \frac{dt}{2}$
> > Then, $P(X = n) = \int_0^{\infty} \frac{(\frac{t}{2})^{n+1}e^{-t}}{n!}\frac{dt}{2} = \frac{1}{2^{n+2}n!}\int_0^{\infty} t^{n+1}e^{-t}dt$
> > Recall that $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}dx$ and $\Gamma(n) = (n-1)!$. Then $P(X = n) = \frac{1}{2^{n+2}n!}\Gamma(n + 2) = \frac{1}{2^{n+2}n!}(n+1)! = \frac{n+1}{2^{n+2}}$ for $n \ge 0$.

3.5 Calculating variance by Conditioning

- Method 1: by definition
  Recall: $Var(X) = E(X^2) - [E(X)]^2$
  Note: Double expectation theorem or law of total expectation is applicable for any expectation.
  Hence: $E[f(x)] = E[E(f(x)|Y)]$
  Then: $\begin{cases} E(X^2) = E[E(X^2|Y)] \\ E(X) = E[E(X|Y)] \end{cases}$ and $Var(X) = E(X^2) - [E(X)]^2$

> Example 3.8:
>
> - A miner is trapped. There are 3 doors.
>   - Door 1 leads to safety after 2 hrs.
>   - Door 2 returns the miner to the starting point in 3 hrs.
>   - Door 3 leads the miner to starting points after 4 hrs.
> - Assuming the miner randomly chooses a door at each time.
> - Let X be the length of time until the miner gets out.
>   Find $Var(X)$.
>
> > Solution: Recall example 3.5.
> > Let Y be the door # miner chosen at the first time.
> > Then, $P(Y=1) = P(Y=2) = P(Y=3) = \frac{1}{3}$.
> > Here, $X|Y=1=2$, $X|Y=2=3+R$, $X|Y=3=4+R$, where R and X have the same distribution.
> > We found $E(X) = 2+3+4 = 9 = E(R)$ and $E(X^2) = E(R^2)$.
> >
> > $$\begin{aligned} E(X^2) &= \sum_y E(X^2|Y=y)P(Y=y) \\ &= E(X^2|Y=1)P(Y=1) + E(X^2|Y=2)P(Y=2) + E(X^2|Y=3)P(Y=3) \\ &= 2^2\frac{1}{3} + E[(3+R)^2]\frac{1}{3} + E[(4+R)^2]\frac{1}{3} \\ &= \frac{4}{3} + \frac{1}{3}E(9+6R+R^2) + \frac{1}{3}E(16+8R+R^2) \\ &= \frac{1}{3}(4+9+16) + \frac{1}{3}\cdot 14\cdot E(R) + \frac{2}{3}E(R^2) \\ &= \frac{1}{3}(4+9+16) + \frac{1}{3}\cdot 14\cdot 9 + \frac{2}{3}E(X^2) \end{aligned}$$
> >
> > Solve get $E(X^2) = 155$, and $Var(X) = E(X^2) - [E(X)]^2 = 155 - 9^2 = 74$.

- Method 2: Conditional variance formula
  i. Given $Y=y$ the conditional variance of X is defined as

$$Var(X|Y=y) = E[X^2|Y=y] - [E(X|Y=y)]^2$$

ii. $Var(X|Y=y)$ is a function of y, say $h(y) = Var(X|Y=y)$.

> e.g. $X|Y=y \sim \text{Pois}(y)$, then $Var(X|Y=y) = y$.Hence $h(y) = y$.

iii. Apply $h(y)$ to Y, we get $h(Y)$ and we denote $Var(X|Y) = h(Y)$.

> Note: We also have $Var(X|Y) = E[X^2|Y] - [E(X|Y)]^2$.
> To find $Var(X|Y)$:
> Step 1: Find $h(y) = Var(X|Y=y)$.
> Step 2: Apply $h(y)$ to Y, we get $Var(X|Y) = h(Y)$.

iv. Comments on $Var(X|Y=y)$

a. Substitution rule is still apllicable

b. If X abd Y are independent, then $Var(X|Y=y) = Var(X)$

v. Formula to calculate $Var(X)$

Theorem: Var(X) = $E[Var(X|Y)] + Var[E(X|Y)]$

Proof: $LHS = Var(X) = E(X^2) - [E(X)]^2$

$RHS = E[Var(X|Y)] + Var[E(X|Y)] = E[E(X^2|Y)] - E[E(X|Y)^2] + E[g(X)^2] - E[g(X)]^2 = E(X^2) - E(X)^2 = LHS$

> Example 3.9:
>
> - A coin is weighted such that $P(H) = 1/4$.
> - Let N = number of tosses required to get 3 Hs by using the weighted coin.
> - Suppose that we toss a fair coin N times. Let X = number of Hs in the N tosses. Find $E(X)$ and $Var(X)$.
>
> > Solution: $N \sim \text{NegBin}(3, 1/4)$ and $X|N=n \sim \text{Bin}(n, 1/2)$.
> > $E(X) = E[E(X|N)] = E(N/2) = E(N)/2 = 3/1/2 = 6$.
> > $E(X|N=n) = n/2 = g(n) \implies E(X|N) = g(N) = N/2$.
> > $Var(X) = E[Var(X|N)] + Var[E(X|N)]$
> > $Var(X|N=n) = n/4 = h(n) \implies Var(X|N) = h(N) = N/4$
> > Hence, $Var(X) = E[N/4] + Var(N/2) = \frac{E(N)}{4} + \frac{1}{4}Var(N) = \frac{3/\frac{1}{4}}{4} + \frac{1}{4}\frac{3(1-\frac{1}{4})}{(1/4)^2} = 12$.

- Method 3: Compound r.v. formula

i. Setup:

Suppose $X_1, X_2, \ldots$ are a sequence of iid rvs.

N: a non-negative r.v.

Further: $X_1, X_2, \ldots$ and N are independent.

Then $W = \sum_{i=1}^{N} X_i$ is called a compound r.v.

[If $N = 0$, then $W = 0$.]

ii. Result:

$E(W) = E(N)E(X_1)$

$Var(W) = E(N)Var(X_1) + Var(N)[E(X_1)]^2$

> Proof: $E(W) = E[E(W|N)]$
> $E(W|N=n) = E[\sum_{i=1}^{n} X_i|N=n] = \sum_{i=1}^{n} E(X_i|N=n) = \sum_{i=1}^{n} E(X_1|N =$

> $n) = nE(X_1)$
> Hence E(W|N) = $NE(X_1)$
> $\implies E(W) = E[NE(X_1)] = E(N)E(X_1)$
> $Var(W) = E[Var(W|N)] + Var[E(W|N)]$
> $Var(W|N=n) = Var(\sum_{i=1}^n X_i|N=n) = \sum_{i=1}^n Var(X_i|N=n) = \sum_{i=1}^n Var(X_1|N=n) = nVar(X_1)$
> Hence $Var(W) = E[Var(W|N)] + Var[E(W|N)] = E[N]Var(X_1) + Var(N)E(X_1)^2$

Example 3.10

- A coin is weighted such that $P(H) = 1/4$.
- Let N = number of tosses required to get 3 Hs by using the weighted coin.
- Suppose that we toss a fair coin N times. Let X = number of Hs in the N tosses. Find $E(X)$ and $Var(X)$.

> Solution: Let $X_i = \begin{cases} 1 & \text{if } i\text{th toss is H} \\ 0 & \text{if } i\text{th toss is T} \end{cases}$
> $X = \sum_{i=1}^N X_i : X_1, X_2 ... \overset{\text{iid}}{\sim} \text{Berloulli}\left(\frac{1}{2}\right)$ and $N \sim \text{NegBin}(3, 1/4)$.
> Then, $E(X) = E(N)E(X_1) = \frac{3}{1/4} \cdot \frac{1}{2} = 6$
> $Var(X) = E(N)Var(X_1) + Var(N)[E(X_1)]^2 = \frac{3}{1/4} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{3 \cdot (1-\frac{1}{4})}{1/4^2} \cdot \frac{1}{4} = 12.$

Example 3.11 (Example 3.7 continued; Final exam, Spring 2016)

- Y has pdf $f(y) = \begin{cases} ye^{-y} & y > 0 \\ 0 & y \le 0 \end{cases}$.
- $X|Y = y \sim \text{Pois}(y)$
  Find $E(X)$ and $Var(X)$.

> Solution:
> $E(X) = E[E(X|Y)] = E(Y) = \int_0^\infty f(y)dy = \int_0^\infty y^2 e^{-y} dy$
> $E(X|Y=y) = y$ since $X|Y = y \sim \text{Pois}(y)$.
> Therefore $E(X|Y) = Y$, hence $E(X) = 2$.
> $Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$
> $Var(X|Y=y) = y$ since $X|Y = y \sim \text{Pois}(y)$.
> Therefore $Var(X|Y) = Y$.
> Hence, $Var(X) = E(Y) + Var(Y) = E(Y) + E(Y^2) - [E(Y)]^2$,.
> $E(Y^2) = \int_0^\infty y^3 e^{-y} dy = \Gamma(4) = 3! = 6$.
> $Var(X) = 2 + 6 - 2^2 = 4$.

Example 3.12 (Question 3, Midterm 2017)
Suppose $Y \sim N(0,1)$ and $N \sim \text{Bin}(3, 0.5)$. That is,
$P(N=0) = P(N=3) = \frac{1}{8}$, $P(N=1) = P(N=2) = \frac{3}{8}$.
We further assume that Y and N are independent. Define $X = Y^N$.

> Find $E(X)$, $Var(X)$, and $P(X<0)$.
> Note: $E(Y)=E(Y^3)=E(Y^5)=0, E(Y^2)=1, E(Y^4)=3, E(Y^6)=15$, and $P(Y<0)=0.5$.
>
>> Solution:
>> $E(X)=E[E(X|N)]=E(Y^N)=E(Y^N|N=0)P(N=0)+E(Y^N|N=1)P(N=1)+E(Y^N|N=2)P(N=2)+E(Y^N|N=3)P(N=3)=E(Y^0)P(N=0)+E(Y^1)P(N=1)+E(Y^2)P(N=2)+E(Y^3)P(N=3)=1\cdot\frac{1}{8}+0\cdot\frac{3}{8}+1\cdot\frac{3}{8}+0\cdot\frac{1}{8}=\frac{1}{2}$
>> Similarly, $E(X^2)=E[E(X^2|N)]=E(Y^{2N})=E(Y^{2N}|N=0)P(N=0)+E(Y^{2N}|N=1)P(N=1)+E(Y^{2N}|N=2)P(N=2)+E(Y^{2N}|N=3)P(N=3)=E(Y^0)P(N=0)+E(Y^2)P(N=1)+E(Y^4)P(N=2)+E(Y^6)P(N=3)=28/8=\frac{7}{2}$.
>> $Var(X)=E(X^2)-[E(X)]^2=\frac{7}{2}-\frac{1}{2}^2=\frac{13}{4}$
>> P(X<0) = \sum_{n=0}^3 P(X<0|N=n)P(N=n) = \sum_{n=0}^3 P(Y^N<0|N=n)P(N=n) = \sum_{n=0}^3 P(Y^n<0|N=n)P(N=n) = P(Y^1<0)P(N=1) + P(Y^3<0)P(N=3) = 0.5\cdot\frac{3}{8} + 0.5\cdot\frac{1}{8} = \frac{1}{4}$.

> Example 3.13 (Question 3, Midterm 2019)
> Suppose $X_1, X_2, \ldots$ are a sequence of independent and identically distributed uniform random variables with probability density function $f(x)=\begin{cases}1 & 0<x<1\\ 0 & \text{otherwise}\end{cases}$. Let Y be a continuous random variable with probability density function $g(y)=\begin{cases}3y^2 & 0<y<1\\ 0 & \text{otherwise}\end{cases}$. We further assume that $X_1, X_2, \ldots$ and Y are independent. Let $T=\min\{n\geq 1: X_n<Y\}$. That is, T is the first time to get a smaller value than Y in the sequence $\{X_n\}_{n=1}^{\infty}$.
>
> i. Find $P(T=2)=P(X_1\geq Y, X_2<Y)$.
>> Solution: We condition on Y.
>> $P(T=2)=P(X_1\geq Y, X_2<Y)=\int_0^1 P(X_1\geq y, X_2<y|Y=y)f_Y(y)dy=\int_0^1 P(X_1\geq y|Y=y)P(X_2<y|Y=y)3y^2dy=\int_0^1 P(X_1\geq y)P(X_2<y)3y^2dy$
>> Here, $P(X_1\geq y)=\int_y^1 f(x)dx=\int_y^1 1dx=1-y$ and $P(X_2<y)=\int_0^y f(x)dx=\int_0^y 1dx=y$.
>> Hence, $P(T=2)=\int_0^1(1-y)y3y^2dy=\int_0^1 3y^3-3y^4dy=\frac{3}{4}-\frac{3}{5}=\frac{3}{20}$.
>
> ii. Find $E(T)$.
>> Solution: We condition on Y.
>> Given $Y=y$, $T|Y=y$ = 1st time to get a smaller value than y in the sequence $\{X_n\}_{n=1}^{\infty}$.
>> Note $P(X_1<y)=\int_0^y f(x)dx=\int_0^y 1dx=y$.
>> Hence, $T|Y=y\sim\text{Geo}(y)$.

> > $E(T|Y=y) = \frac{1}{y}$.
> > $E(T) = E[E(T|Y)] = \int_0^1 \frac{1}{y} 3y^2 dy = 3\int_0^1 y dy = \frac{3}{2}$.
>
> iii. Find $Var(T)$.
>
> > Solution:
> > $E(T^2) = E[E(T^2|Y)] = \int_0^1 E(T^2|Y=y) f_Y(y) dy$
> > $T|Y=y \sim \text{Geo}(y)$, hence $E(T^2|Y=y) = Var(T|Y=y) + [E(T|Y=y)]^2 = \frac{1-y}{y^2} + \frac{1}{y^2}$.
> > Hence,
> > $E(T^2) = \int_0^1 \frac{1-y}{y^2} + \frac{1}{y^2} 3y^2 dy = \int_0^1 3(2-y) dy = 6 - \frac{3}{2} = \frac{9}{2}$.
> > $Var(T) = E(T^2) - [E(T)]^2 = \frac{9}{2} - (\frac{3}{2})^2 = \frac{9}{4}$.