STAT 333 notes: Applied Probability

Johnew Zhang

September 22, 2018

Contents

1 Introduction

In this section we are going to mainly cover two things, review of probability theory and review of random variable.

1.1 Basic Concepts of Probability

1.1.1 What is a probability model?

Three components form the probability model, sample space, event and probability function.

Sample Space For example, toss a coin or toss a die, the outcome is not predictable; all possible outcomes are known; the set of all possible outcomes form a sample space (Ω). Tossing a coin has 6 possible outcomes, denoted as $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Event Roughly speaking, an event is a subset of the sample space (Ω). For example, towing a die has $E = \{2, 4, 6\}$. Here E is an event.

Probability Function We use P to denote it. It satisfies three conditions

1. for any even E, $0 \leq P(E) \leq 1$.
2. $P(\Omega) = 1$.
3. "Additivity": If $E_1, E_2, \cdots$ are disjoint or mutually exclusive (i.e. $E_i \cap E_j = \emptyset$), then

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$

In other words, the probability of the union of disjoint events is equal to the sum of probability of disjoin events.

Let's consider the following example. Toss a die, for an event. Let

$$P(E) = \frac{\text{number of outcomes in E}}{6}$$

For example, $E_1 = \{1, 2\}, P(E_i) = \frac{2}{6}$. Claim that this P is a probability function.

Some important properties

1. If $E_1 \subset E_2$, then $P(E_1) \leq P(E_2)$. In other words, the larger set has larger probability than smaller set. Note: the meaning of subset is that if E_1 occurs, we must have E_2 occur, then $E_1 \subset E_2$.
2. $P(\emptyset) = 0$
3. $P(E) + P(E^c) = 1$ where E^c is the complementary set of E.
4. $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ if $E_1 \cap E_2 = \emptyset$.

5. In general, $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

Conditional Probability

Definition. *Let $E\&F$ be two events and $P(F) > 0$, then conditional probability of E given F is*

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

One useful result:

$$P(E \cap F) = P(E|F)P(F)$$

Independence Two events are independent if $P(E \cap F) = P(E)P(F)$. Extension to multiple events are not required. (Message: Probability of intersection of independent events = product of probability of each event.

Simple& Useful result: Suppose we have a sequence of independent trails & a sequence of events :$E_1, \cdots, E_n, \cdots$. Further: E_i only depends on the trial.

Conclusion:

1. $E_1, \cdots, E_n, \cdots$ are independent
2. $P(\cap_{i=1}^{n} E_i) = \prod_{i=1}^{n} P(E_i)$ & $P(\cap_{i=1}^{\infty} E_i) = \prod_{i=1}^{\infty} P(E_i)$

For example, suppose we toss a fair die repeatedly & independently, we get a sequence of number. Find P(observe 333 in the sequence). Let E = observe "333" in the sequence, E^c = Not observe "333" in the sequence. E^c implies that first 3 numbers is not "333", $\cdots$

$P(E^c) \leq P$(1st 3 numbers is not "333" $\cdots$)

Therefore, $P(E^c) = P$(1st 3 number is not "333") $\cdots = (1 - \frac{1}{6}^3)(1 - \frac{1}{6}^3) \cdots = 0$. As a result, $P(E^c) \leq 0$ & $P(E^c) = 0 \implies P(E) = 1$

Summary:

1. partition the sequence into no-overlap blocks
2. $P(E_1) \leq P(E_2)$ if $E_1 \subset E_2$
3. $P(E) + P(E^c) \leq 1$
4. independence property

Here is another example:

toss a fair coin repeatedly & independently $P(H) = P(T) = \frac{1}{2}$

1. P(1st 2 tosses gives HH) $= \frac{1}{4}$
2. P(1st 2 tosses gives TH) $= \frac{1}{4}$

3. $P(\text{TH occurs before HH in the sequence}) = \frac{3}{4}$.

   Case I: 1st outcome is a tail and TH occurs before HH. If you have T already, only need one H to get TH but HH to get HH.

   Case II: sat is head and second is tail. Similar as case I, TH occurs 1st.

   Case II: sat is a head and second is also a head. Then HH occurs before TH.

Bayes Formula Suppose: $F_1, F_2, \cdots$ are disjoint events such that $\cup F_i =$ Sample space.

Result 1: consider an event E, $P(E) = \sum P(E \cap F_i) = \sum P(E|F_i)P(F_i)$

Result 2:

$$P(F_k|E) = \frac{P(E \cap F_k)}{P(E)} = \frac{P(E|F_k)P(F_k)}{\sum P(E|F_i)P(F_i)}$$

Here is another example (Monty Hall problem). There are three doors A, B, C and two goats & a car. Monty knows the position of car, but you don't know. Now, let's randomly select a door and the probability of choosing a car is a third. Monty opens a door to reveal a goat. You can have choice to switch the door. if you choose the door has car, you can win the car.

Solution: I suppose you always choose door A and then you will have 2 thirds chance to win. In general, P is two thirds. Now let's try to apply Bayes formula to solve this question. Suppose you choose door A and Monty opens door B. $E_i = \text{car behind door}, i \in \{A, B, C\}$ and $E =$ Monty open door B).

$$P(\text{win if switch}) = P(E_C|E) = \frac{P(E|E_C)P(E_C)}{P(E|E_A)P(E_A) + P(E|E_B)P(E_B) + P(E|E_C)P(E_C)}$$

Hence $P(\text{win if switch}) = \frac{2}{3}$

1.2 Review of Random Variable

A random variable is a function defined from sample space to real line. X : sample space $\to \mathbb{R}$. Two types of random variables:

- discrete: all possible values is finite or countable such as binomial and poisson.
- continuous: all possible values contain an interval, such as normal

1.3 Some important distributions

1.3.1 Binomial Trials

Each trial has 2 outcomes: success or failure. All trails are independent. Probability of success will be the same for all trials: $P(S) = P$.

$$\text{Let } I = \begin{cases} 1 & \text{if S on ith trial} \\ 0 & \text{OW} \end{cases}$$

$I_1, I_2, \cdots$ are a sequence of Bernoulli trials.

Comments:

1. $I_1, I_2, \cdots$ are independent and identically distributed (iid)
2. $P(I_i = 1) = P, P(I_i = 0) = 1 - P$

1.3.2 Binomial Random Variable

denoted by $BIN(n,p)$

Let $X = \#$ of S in n Bernoulli Trials $X \sim BIN(n,p)$

Range of $X = \{0, 1, 2, \cdots, n\}$ Probability mass function

$$P(X = k) = \binom{n}{r} P^k (1 - P^{n-k}, k = 0, 1, \cdots, n$$

Result I:

If $X \sim BIN(n,p)$, then $X = I_1 + I_2 + \cdots + I_n$ [connection between binomial and Bernoulli RVs]

Result II:

If $X_1 \sim BIN(n,p)$, $X_2 \sim BIN(n,p)$ and X_1, X_2 are independent, then $X_1 + X_2 \sim BIN(n,p)$

1.3.3 Geometric Random Variable

denoted by $GEO(P)$ Let $X = \#$ of trials to see the first $S \sim GEO(p)$

- $GEO(p)$ is the first waiting time
- $x \in \{1, 2, 3, \cdots\}$
- p.m.f.: $P(X = k) = (1-p)^{k-1} p, k = 1, \cdots$

Property: Non-memory property (i.e. $P(X > n + m | X > m) = P(X > n)$

Meaning of No-memory property

No matter how long you spent, as long as you do not observe S, the remaining time $\sim GEO(P)$

"Toy Example"

Toss a fair coin. Observe 5 T already. What is the probability of requiring 10 trials in total to see the first H.

Solution:

Remaining time $\sim GEO(P = 1/2)$. Here $P(\text{remaining time} = 5) = (1-P)^{5-1} P = \frac{1}{2^5}$

1.3.4 Negative Binomial Random Variable

denoted by $NEGBIN(r, P)$

Let $X = \#$ of trials to see r S in the sequence $X \sim NEGBIN(r, p)$

Support of this distribution is $\{r, r+1, r+2, \cdots\}$

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

first r trials of r S of $k - r$ F has $r - 1$ S.

Property: $NegBin \& Geo(P)$

Let X_1 = wiring time for the first S, X_2 = waiting time for the second S after the first S

X_r = waiting time for rth S after (r -1)th S.

1. $X_1, X_2, \cdots, X_r \sim_{iid} Geo(p)$
2. $X = \sum_{i=1}^{r} X_i$

1.3.5 Poisson Random Variable

denoted by $Pois(\lambda)$

If the customers come to T.H. uniform and randomly over the time with rate λ per unit.

Let

$$X(t) = \# \text{ of customers coming to TH in } [0, t]$$

Then $X(t) \sim Pois(\lambda t)$

$X(t) \in \{0, \cdots\}$

$$P(X(t) = k) = \frac{(\lambda t)^t e^{-\lambda t}}{k!}, k = 0, 1, 2, \cdots$$

If $X_1 \sim POIS(\lambda_1)$ and $X_2 \sim POIS(\lambda_2)$ and independent, then

$$X_1 + X_2 \sim POIS(\lambda_1 + \lambda_2)$$

1.3.6 Exponential Random Variable

denoted by $EXP(\lambda)$

Let X = waiting time to see a customer in TH

$X \sim EXP(\lambda)$ $f(x) = \lambda e^{-\lambda x}, x \geq 0.$

Properties: $P(X > x) = e^{-\lambda x}$ (tail probability); No-memory property $P(X > t+s | X > s) = P(X > t)$

1.3.7 Summary for Rate λ

1. λ is the unit rate for poisson rv
2. λ is also the rate for exponential waiting time

1.4 Expectation & Variance

1.4.1 Expectation

1. Linearity:

$$E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i E[X_i]$$

e.g. $E[X_1 + X_2] = E[X_1] + E[X_2]$

2.

1.4.2 Variance

$$Var(X) = E[(X - \mu)^2] = E[X^2] - E[X]^2$$

1. $$Var(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i)$$

   if $X_1, \cdots, X_n$ are independent.

   e.g. If X_1 and X_2 are independent, then $Var(X_1 \pm X_2) = Var(X_1) \pm Var(X_2)$

2. In general

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

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

   $Var(X_1+X_2+X_3) = Var(X_1)+Var(X_2)+Var(X_3)+2Cov(X_1, X_2)+2Cov(X_1, X_3)+2Cov(X_2, X_3)$

1.4.3 Covariance

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

If X and Y are independent, then $Cov(X, Y) = 0$.

1.5 Indicator Variable

Indicator variables have only two values: 0&1. Bernoulli RV is indicator variable.

For an event A define

$$I_A = \begin{cases} 1 & \text{A occurs} \\ 0 & O.W. \end{cases}$$

Suppose $P(A) = P$. What is $E[I_A]$ & $Var(I_A)$?

By definition $E[I_A] = 1 \times P(I_A = 1) + 0 \times P(I_A = 0) = P(A) = P$

By definition $Var(I_A) = E[I_A^2] - E[I_A]^2 = P - P^2 = P(1 - P)$

e.g. $X \sim BIN(n; p)$ Find $E[X]$ and $Var(X)$

$X = \sum_{i=1}^n I_i$ where $I_1, \cdots, I_n$ are iid Bernoulli RVs so

$$E[X] = E[\sum_{i=1}^n I_i] = np$$

$$Var[X] = Var[\sum_{i=1}^n I_i] = \sum_{i=1}^n Var(I_i) = np(1 - p)$$

1.5.1 Example

We have two boxes: red and black. Red box has 4 red balls and 6 black balls. Black box has 6 red balls and 4 black balls. Step 1: toss a coin. H from red box and T from black box. Choose the ball and record colour and place it back. Step 2: Choose a ball from the box has the same colour as the ball you selected in step 1. For example, step1 gets a red ball and step 2 selects from Red box. Step 1 gets a blck ball and step 2 selected from black box.

$X = \#$ of red balls selected in the first 2 steps. Find $E[X]$ and $Var(X)$.

Solution: Let $I_1 = \begin{cases} 1 & \text{the ball in step 1 is red} \\ 0 & O.W \end{cases}$ and $I_2 = \begin{cases} 1 & \text{the ball in step 2 is red} \\ 0 & O.W \end{cases}$

Then $X = I_1 + I_2$.

$$E[X] = E[I_1] + E[I_2] = P(I_1 = 1) + P(I_2 = 1)$$

$$\begin{aligned} P(I_1 = 1) &= P(\text{first ball is red}) = P(\text{first ball is red}|H)P(H) + P(\text{the first ball is red}|T)P(T) \\ &= (0.4 + 0.6) \times 0.5 \end{aligned}$$

Hence $P(I_1) = 0.5$

$$\begin{aligned} P(I_2 = 1) = P(\text{second ball is red}) &= P(\text{second ball is red}|\text{first is black})P(\text{first is black}) \\ &+ P(\text{second ball is red}|\text{first is red})P(\text{first is red}) = 0.5 \end{aligned}$$

Therefore, $E[X] = 1$

$$Var(X) = Var(I_1 + I_2) = Var(I_1) + Var(I_2) + 2Cov(I_1, I_2)$$
$$Var(I_1) = P(I_1 = 1)(1 - P(I_1 = 1)) = 0.5^2$$
$$Var(I_2) = P(I_2)(1 - P(I_2 = 1)) = 0.5^2$$
$$Cov(I_1, I_2) = E[I_1 I_2] - E[I_1]E[I_2] = E[I_1 I_2] - 0.5^2$$

Note $I_1 I_2 = \begin{cases} 1 & I_1 = I_2 = 1 \\ 0 & O.W. \end{cases}$

Now

$$\begin{aligned} E[I_1 I_2] &= P(I_1 = I_2 = 1) = P(I_1 = 1 \& I_2 = 1) = P(I_2 = 1 | I_1 = 1)P(I_1 = 1) \\ &= P(\text{second is red}|\text{first is red})P(\text{first is red}) = 0.4 \times 0.5 = 0.2 \end{aligned}$$

Hence $Cov(I_1, I_2) = 0.2 - 0.5^2 = -0.05$

$$Var(X) = 0.5^2 + 0.5^2 - 2 \times 0.05 = 0.4$$

2 Waiting Time Random Variable

2.1 Background

Suppose we have a sequence of independent trials and we would like to observe E.

Let $T_E = \#$ of trials required to see first E (waiting time for first E)

Range of $T_E = \{1, 2, 3, \cdots\} \cup \{\infty\}$. We are interested in

1. Can we observe E? $P(T_E < \infty) = 1$ or $P(T_E = \infty) = 0$?

2. How long will it take? $E[T_E]]$

2.2 Classification of T_E

$$P(T_E < \infty) = 1 \text{ or } P(T_E = \infty) = 0$$

If above is true, T_E is called proper; otherwise, it is improper. If $E[T_E] < \infty$, then it is called short proper. If $E[T_E] = \infty$, it is null proper. For improper, it is automatically $E(T_E) = \infty$

If $P(T_E < \infty) < 1$ or > 0, then T_E is improper. Note: here $E(T_E) = \infty$.

Aside

$\sum_{i=1}^{\infty} a_i$ is just a limit summation does not include "∞". $\sum_{i=1}^{\infty} a_i = \lim_{n\to\infty} \sum_{i=1}^{n} a_i$. $[\sum_{i=1}^{\infty}]$ is the summation over all positive integers but not "∞".

Example 1 (Short proper) Consider the Bernoulli trials $P(S) = p$. E_1 = waiting time for the first "S" $\sim GEO(P)$ and E_2 = waiting time for r "S" $\sim NegBin(r, p)$

Claim: $T_{E_1}\&T_{E_2}$ are short proper. Solution: for E_1: check

1. $P(T_{E_1} < \infty) = 1$ or $P(T_{E_1} = \infty) = 0$.
2. $E(T_{E_1}) < \infty$

For

$$\begin{aligned} P(T_{E_1} < \infty) &= \sum_{i=1}^{\infty} P(T_{E_1} = k) \\ &= \sum_{i=1}^{\infty} (1-p)^{-1} p \\ &= \frac{p}{1-(1-p)} = 1 \end{aligned}$$

Equivalently,

$$\begin{aligned} P(T_{E_1} = \infty) &= P(\text{first} = F, \text{second} = F, \text{third} = F, \cdots) \\ &= P(\text{first} = F)P(\text{second} = F) \cdots = (1-p)^{\infty} = 0 \end{aligned}$$

For $E(T_{E_1})$. Recall $E(T_{E_1}) = \frac{1}{p} < \infty \implies T_{E_1}$ is short proper

For E_2: Note $T_{E_2} = \sum_{i=1}^{r} X_i$. Well it is just that a sum of r geometric distributions will become a negative binomial with r successes.

$$\begin{aligned} P(T_{E_2} < \infty) &= P(\sum_{i=1}^{r} X_i < \infty) \\ &= P(X_1 < \infty, X_2 < \infty, \cdots, X_r < \infty) \\ &= \prod_{i=1}^{r} P(X_i < \infty) = 1 \end{aligned}$$

$$E(T_{E_2}) = E(\sum_{i=1}^{r} X_i) = \sum_{i=1}^{r} E(X_i) = \sum_{i=1}^{r} \frac{1}{p} = r/p < \infty$$

Hence T_{E_2} is short proper.

Example 2 (null proper) Toss a coin independently. Suppose that nth toss probability of $H = P_n = \frac{1}{n+1}$. That is probability of H at first toss is $\frac{1}{2}$. $\cdots$ at second toss is $\frac{1}{3}$. $T_H =$ waiting time for the first H

Claim: T_H is null proper. Solution: Check

1. $P(T_H < \infty) = 1 = \sum_{i=1}^{\infty} P(T_H = i)$
2. $E(T_H) = \infty = \sum_{i=1}^{\infty} i \times P(T_H = i)$

Find $P(T_H = n)$,

$$\begin{aligned} P(T_H = n) &= P(\text{first} = T, \cdots, (n-1)th = T, nth = T) \\ &= P(\text{first} = T) \cdots P((n-1)th = T) P(nth = T) \\ &= (1 - \frac{1}{1+1}) \times (1 - \frac{1}{1+2}) \cdots \times (1 - \frac{1}{n}) \frac{1}{n+1} \\ &= \frac{1}{n(1+n)} = \frac{1}{n} - \frac{1}{n+1} \end{aligned}$$

$$P(T_H < \infty) = \sum_{i=1}^{\infty} P(T_H = i) = \sum_{i=1}^{\infty} (\frac{1}{i} - \frac{1}{i+1}) = 1$$

$$E[T_H] = \sum_{n=1}^{\infty} n \times P(T_H = n) = \sum_{n=1}^{\infty} n \times (\frac{1}{n \times (n+1)}) = \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty$$

3 Conditional Probability & Conditional Expectation

In this chapter, we consider two or more RVs and need the concept of joint RVs.

3.1 Joint Discrete Random Variables

- How to characterize the joint random variables? By joint c.d.f.
- Joint c.d.f: two rvs $X \& Y$. Joint c.d.f. of $X \& Y$ s defined as

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

  Joint c.d.f uniquely define joint RVs.
- Joint discrete RVs: Roughly speaking, if $X \& Y$ are both discrete, then $X \& Y$ are called joint discrete.

- Joint p.m.f. for discrete case: Suppose $X\&Y$ are joint discrete then

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

is joint p.m.f of $X\&Y$. Properties

  1. $f_{XY}(x, y) \geq 0$.
  2. $\sum_x \sum_y f_{XY}(x, y) = 1$

- Marginal p.m.f. from joint p.m.f.

$$f_X(x) := P(X = x) = \sum_y f_{XY}(x, y)$$

$$f_Y(y) := P(Y = y) = \sum_x f_{XY}(x, y)$$

- Joint expectation: Suppose $h(x, y)$ is a bivariate function, then

$$E[h(X, Y)] = \sum_x \sum_y h(x, y) f_{XY}(x, y)$$

For example, $E[XY] = \sum_x \sum_y xy f_{XY}(x, y)$ and $E[X] = \sum_x \sum_y x f_{XY}(x, y)$

3.1.1 Independence

If $f_{XY}(x, y) = f_X(x) f_Y(y)$, then $X\&Y$ are independent. Properties: Expectation and independence: If $X\&Y$ are independent, then

1. $g_1(X)\&g_2(Y)$ are independent;
2. $E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$ For example, if $X\&Y$ are independent then $E[XY] = E[X]E[Y]\&Cov(X, Y) = 0$.

3.2 Joint Continuous Random Variables

If the joint c.d.f of $X\&Y$ can be written as

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{\infty}^{y} f_{XY}(s, t) ds dt$$

then

1. $X\&Y$ are joint continuous
2. $f_{XY}(x, y)$ is called joint p.d.f

Properties:

1. $f_{XY}(x, y) \geq 0$
2. The sum of all possibilities is 1
3. Marginal p.d.f.
$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)dy$$

3.3 Conditional Distribution & Conditional Expectation

3.3.1 Discrete Case

Definition. *Suppose $X\&Y$ has joint p.m.f $f_{XY}(xy)$, marginal p.m.f. $f_X(x), f_Y(y)$. The conditional p.m.f. of X given $Y = y$ is*

$$f_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x|Y = y)}{P(Y = y)} = \frac{f_{XY}(xy)}{f_Y(y)}$$

or

$$Conditional\ pmf = \frac{Joint\ pmf}{marginal\ pmf}$$

Properties:$f_{X|Y}(x|y)$ is a pmf. That is

1. $f_{X|Y}(x|y) \geq 0$
2. $\sum_x f_{X|Y}(x|y) = 1$

Proof. 1.
$$f_{X|Y}(xy) = \frac{f_{XY}(x, y) \geq 0}{f_Y(y) > 0} \geq 0$$

2.
$$\sum_x f_{X|Y}(x|y) = \sum_x \frac{f_{XY}(xy)}{f_Y(y)}$$

so
$$\sum_x f_{X|Y}(x|y) = \frac{\sum_x f_{XY}(xy)}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1$$

□

Definition. *Conditional expectation: Since $f_{X|Y}(x|y)$ is a pmf, we can define expectation based on it. Condition expectation of X given $Y = y$*

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

Conditional expectation of $g(X)$ given $Y = y$.

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

Comment: if $X\&Y$ are independent. Here, $f_{XY}(xy) = f_X(x)f_Y(y)$ and $f_{X|Y}(x|y) = \frac{f_{XY}(xy)}{f_Y(y)} = f_X(x)$ so the conditional pmf is just the marginal pmf.

Note: $E[X|Y = y] = \sum_x x f_{X|Y}(x|y) = \sum_x x f_X(x) = E[X]$ so if $X\&Y$ are independent

$$E[X|Y = y] = E[X]$$

Example

Suppose $X_1 \sim POIS(\lambda_1)$, $X_2 \sim POIS(\lambda_2)$ are independent. Let $X = X_1, Y = X_1 + X_2 \sim POIS(\lambda_1 + \lambda_2)$ and Find $f_{X|Y}(x|y) \& E[X|Y = y]$

Solution: By definition

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

For

$$P(X = x, Y = y) = P(X_1 = x, X_1+X_2 = y) = P(X_1 = x, X_2 = y-x) = P(X_1 = x)P(X_2 = y-x)$$

Hence we just need to plug the poisson distribution formula into the expression and we are done.

$$f_{X|Y}(x|y) = \frac{y!}{x!(y-x)!} \frac{\lambda_1^x \lambda^{y-x}}{(\lambda_1 + \lambda_2)^y} = \binom{y}{x} (\frac{\lambda_1}{\lambda_1 + \lambda_2})^x (\frac{\lambda_2}{\lambda_1 + \lambda_2})^{y-x}$$

so $X|Y = y \sim BIN(y, P = \frac{\lambda_1}{\lambda_1+\lambda_2})$

Therefore $E(X|Y = y) = y\frac{\lambda_1}{\lambda_1+\lambda_2}$

3.3.2 Continuous Case

Suppose $X\&Y$ are joint continuous with joint df $f_{X|Y}(xy)$. Marginal pdf $f_X(x), f_Y(y)$.

Definition. *Conditional pdf: The conditional pdf pf X given $Y = y$ is*

$$f_{X|Y}(x|y) = \frac{f_{XY}(xy)}{f_Y(y)} = \frac{\textit{joint pdf}}{\textit{Marginal pdf}}$$

Claim: $f_{X|Y}(x|y)$ is a pdf. That is

1. $f_{X|Y}(x|y) \geq 0$.
2. $\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = 1$

Definition. *Conditional Expectation: Conditional expectation of X given $Y = y$ is*

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

conditional expectation of $g(X)$ given $Y = y$ is

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

Comment: If $X\&Y$ are independent $f(X|Y = y) = f_X(x)$, $E[X|Y = y] = E[X]$

Example

$$f_{XY}(xy) = \begin{cases} xe^{-xy} & x > 0, y > 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $f_{X|Y}(x|y) \& E[X|Y = y]$

First of all

$$f_Y(y) = \frac{1}{y^2}\Gamma(2) = \frac{1}{y^2}$$

Conditional pdf

$$f_{X|Y}(x|y) = \frac{xe^{-xy}}{1/y^2} = xy^2 e^{-xy}$$

$$E[X|Y = y] = \int_0^{\infty} x f_{X|Y}(x|y) dx = \int_0^{\infty} x^2 y^2 e^{-xy} dx$$

Let $t = xy$. Hence

$$E[X|Y = y] = \frac{1}{y} \int_0^{\infty} t^2 e^{-t} dt$$

so

$$E[X|Y = y] = \frac{\Gamma(3)}{y} = \frac{2}{y}$$

Summary of Properties for conditional expectation: For both continuous and discrete

1. Conditional expectation has all properties of expectation
2. Substitution rule:

$$E[Xg(Y)|Y = y] = E[Xg(y)|Y = y] = g(y)E[X|Y = y]$$

   In general $E[h(XY)|Y = y] = E[h(XY)|Y = y]$

3. If X&Y are independent, $E[X|Y = y] = E[X]$ and $E[g(X)|Y = y] = E[g(X)]$

3.3.3 Expectation Theorem

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

By double expectation theorem,

$$E[X] = E[X|Y=s]P(y=s) + E[X|Y=F]P(Y=F) = 1 \times P + [1+E[X]] \times (1-p)$$

$$E[X] = p + (1-p) + (1-p)E[X] \implies E[X] = 1/p$$

Example

A miner is crapped. There are 3 doors to go out side. Door 1: leads to safety after 2 hours. Door 2: return the miner to starting point in 3 hours. Door 3: return the miner to starting point in 4 hours. Assume miner randomly choose a door at each time

$$X = \text{length of time until miner go out}$$

Find $E[X]$

Solution: Let Y denote the door number $P(Y=1) = P(Y=2) = P(Y=3) = \frac{1}{3}$

$X|Y=1=2 \implies E[X|Y=1]=2$, $X|Y=2=3+$ remaining time, $X|Y=3=4+$remaining time. No memory: remaining time $\&X$ have same distribution. so

$$E[X|Y=1]=2$$

$$E[X|Y=2] = 3 + E[\text{remaining time}] = 3 + E[X]$$

$$E[X|Y=3] = 4 + E[\text{remaining time}] = 4 + E[X]$$

By double expectation theorem, $E[X] = \sum_{y=1}^{3} E[X|Y=y]P(Y=y) = \frac{1}{3} \times 2 + \frac{1}{3} \times [3+E[X]] + \frac{1}{3} \times [3+E[X]] \implies E[X] = 2+3+4 = 9$

Think: If miner will not choose the door that he choose before what is $E[X]$.

3.4 Calculating probability by conditioning

Suppose we have an event A, we are interested in $P(A)$. Let $I_A = \begin{cases} 1 & \text{if A occurs} \\ 0 & \text{O.W.} \end{cases}$

$$P(A) = E[I_A] = E[E[I_A|Y]] = \begin{cases} \sum_y E[I_A|Y=y]f_Y(y) & \text{Y is discrete} \\ \int_{-\infty}^{\infty} E[I_A|Y=y]f_Y(y)dy & \text{Y is continuous} \end{cases}$$

Example 3.6

$X_1, X_2, X_3 \sim_{iid} Uniform[0,1]$

1. $P(X_1 < X_2) = \frac{1}{2}$
2. P($X_1 < X_2 < X_3) = \frac{1}{6}$

Solution:

1. $P(X_1 < X_2) = E[I_{X_1<X_2}]$ Conditioning on $Y = X_2$ $f_Y(y) = \begin{cases} 1 & 0<y<1 \\ 0 & \text{OW} \end{cases} = \int_0^1 E[I[X_1 < X_2]|Y = X_2 = y]f_Y(y)dy$ so $P(X_1 < X_2) = \int_0^1 E[I_{X_1<X_2}|Y = X_2 = y] \times 1dy = \int_0^1 E[I_{X_1<y}|Y = X_2 = y]dy = \int_0^1 E[I_{X_1<y}]dy$
   Note $E[I_{X_1<y}] = P(X_1 < y) = \int_0^y 1 \times \text{pdf of } X_1 dy = y$
   $\implies P(X_1 < X_2) = \int_0^1 E[I_{X_1<y}]dy = \int_0^1 ydy = \frac{1}{2}$

2. $$P(X_1 < X_2 < X_3) = E[I_{(X_1<X_2<X_3)}] = E[E[I_{X_1<X_2<X_3}|X_2]]$$
   so
   $$\begin{aligned} P(X_1 < X_2 < X_3) &= \int_0^1 E[I_{X_1<X_2<X_3}|X_2 = y]f_{X_2}(y)dy \\ &= \int_0^1 E[I_{X1<y<X_3}|X_2 = y]dy \end{aligned}$$
   X_1, X_3 are independent from $X_2 \implies P(X_1 < X_2 < X_3) = \int_0^1 E[I_{X_1<y<X_2}]dy$
   $E[I_{X_1<y<X_3}] = P(X_1 < y < X_3)$ Then $P(X_1 < y < X_3) = P(X_1 < y \& y < X_3) = P(X_1 < y)P(y < X_3)$
   $$P(X_1 < y) = \int_0^y f_{X_1}(x_1)dx_1 = \int_0^y 1dx_1 = y$$
   $$P(X_3 > y) = \int_y^1 f_{X_3}(x_3)dx_3 = (1 - y)$$
   so
   $$P(X_1 < X_2 < X_3) = \int_0^1 P(X_1 < y < X_3)dy = \int_0^1 y \times (1 - y)dy = \frac{1}{6}$$

Example

An insurance company suppose number of accidents each customer $\sim$ Poisson. And the poisson mean for a randomly selected customer is a RV & has p.d.f $f(y) = ye^{-y}, y > 0$. Find probability of a random selected customer has n accident.

Solutions: X = number of accidents for a randomly selected customer, Y = poisson mean of a randomly s

Conditions tell us

$$\begin{cases} Y & \text{has pdf } ye^{-y}, y > 0 \\ X|Y = y & \sim Poisson(y) \end{cases}$$

Find $P(X = n)$.

Solution: $P(X = n) = E[I_{x=n}] = E[E[I_{X=n}|Y]] = \int_0^\infty E[I_{X=n}|Y = y]f_Y(y)dy$.

Now

$$E[I_{X=n}|Y = y] = P(X = n|Y = y) = \frac{y^n e^{-y}}{n!}$$

so

$$P(X = n) = \int_0^\infty \frac{y^n e^{-y}}{n!} ye^{-y} dy$$

You substitute by $y = \frac{t}{2}$. We get the following $= \int_0^\infty t^{n+1}e^{-t}dt/(2^{n+2\times n!}) = \Gamma(n + 2)$ so

$$P(X = n) = \frac{\Gamma(n + 2)}{2^{n+2}n!} = \frac{n + 1}{2^{n+2}}$$

3.5 Calculate Variance by Conditioning

Given X calculate $Var(X)$.

Method 1: definition of variance and double-expectation.

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

Example (miner problem from previous lecture)

Let R = Remaining time. R and X have same distribution because of no memory property. Now

$$E[X^2] = E[E[X^2|Y]] = \sum_{y=1}^{3} E[X^2|Y = y]P(Y = y)$$

so

$$E[X^2] = \frac{1}{3}\times 2^2 + \frac{1}{3}\times E[(3+R)^2] + \frac{1}{3}\times E[(4+R)^2] = \frac{1}{3} + \frac{1}{3}E[9+6R+R^2] + \frac{1}{3}E[16+8R+R^2]$$

Note $E[X^2] = E[R^2]$

Expand them, we can get $E[X^2] = 155$ and $Var(X) = E[X^2] - E[X]^2 = 74$.

Method 2: Def Conditional variance. Given $Y = y$, the conditional variance of X is $Var(X|Y = y) = E[X|Y = y] - [E[X|Y = y]]^2$.

Note that $Var(X|Y = y)$ is a function of y and w use $h(y) = Var(X|Y = y)$
Another definition: Conditional variance of X given Y. $Var(X|Y) = h(Y)$
Comments

1. Two steps to calculate $Var(X|Y)$. Step 1: Calculate $h(y) = Var(X|Y = y)$. Step 2: $Var(X|Y) = h(Y)$

2. If $X\&Y$ are independent, $Var(X|Y) = Var(X$.

3. Substitution rule can still be applied. A coin is weighted such that $P(H) = \frac{1}{4}$.

Let N = number of tosses required to get 3 Hs by the weighted coin. $N \sim NegBin(3, \frac{1}{4})$
We toss another fair coin N times. Let X = number of H is in the N tosses Find $Var(X)$.
solution: $X|N = n \sim Bin(n\frac{1}{2})$ and $N \sim NegBin(3, \frac{1}{4})$
$E[X|N = n] = \frac{n}{2}, Var(X|N = n) = n \times \frac{1}{2}(1 - \frac{1}{2}) = \frac{n}{4} \implies Var(X|N) = \frac{N}{4}$
so

$$Var(X) = Var(E[X|N]) + E[Var(X|N)] = Var(\frac{N}{2}) + E(\frac{N}{4}) = \frac{1}{4}Var(N) + \frac{1}{4}E[N]$$

Hence $Var(X) = 12$

3.6 Application to compound RVs to a random summation of iid RVs

Suppose $X_1, X_2, \cdots$ are a sequence of iid RVs. N is a RV which only take non-negative integers. Further $N\&X_1, X_2, \cdots$ are independent Let $W = \sum_{i=1}^{N} X_i$ Comound RV.
Our interest: Find $E[W]$ and $Var(W)$. (Aggregate Claim example).

Theorem.

$$E[W] = E[N]E[X_i]$$

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

Proof.

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

Here $X = W$ and $Y = N$.
$E[X] = E[E[X|Y]]$
$E[X|N]$:
Two steps:

Step 1 $E[W|N = n] = E[\sum_{i=1}^{N} X_i|N = n] = E[\sum_{i=1}^{n} X_i|N = n]$ so $E[W|N = n] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = nE[X_i]$.

Hence $E[W|N = n] = nE[X_i]$

Step 2 $E[W|N] = NE[X_i]$ so $E[W] = E[E[W|N]] = E[NE[X_i]] = E[N]E[X_i]$

$Var(W|N)$:

Step 1 $Var(W|N=n) = Var(\sum_{i=1}^{n} X_i|N=n) = Var(\sum_{i=1}^{n} X_i|N=n) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$. $X_1, \cdots, X_n$, independent. Since $X_1, \cdots, X_n$ are iid, $Var(W|N = n) = Var(\sum_{i=1}^{n} X_i) = nVar(X_i)$. Then $Var(W|N=n) = nVar(X_i)$

Step 2 $Var(W|N) = NVar(X_i)$. By conditional variance formula $Var(W) = E[Var[W|N]] + Var(E[W|N]) = E[NVar(X_i)] + Var(NE[X_i]) = E[N]Var(X_i) + Var(N)[E[X_i]]^2$

□

Example

Let N = number of customers marking claim in 2012 $\sim Poi(200)$. Suppose the amount claimed by each customer $\sim EXP(rate = \frac{1}{100})$.

Let W = total amount paid to all customers in 2012. Find $E[W]\&Var(W)$.

Solution: Let X_i = amount claimed by ith customer

$W = \sum_{i=1}^{n} X_i$. $E[W] = E[N]E[X_i] = 20000$. $Var(W) = E[N]Var(X_i) + Var(N)E[X_i]^2 = 4 \times 10^6$.

Example

$N \sim NegBin(3, \frac{1}{4})$ X = number of "H", in N tosses with a fair coin. Find $E[X]\&Var(X)$.

Solution: Let $xX = \begin{cases} 1 & \text{if ith tosses is "H"} \\ 0 & \text{OW} \end{cases}$

$E[X] = E[N]E[X_i] = 6$ $Var(X) = 12$

4 Probability Generating Function [pgf]

4.1 Generating Function [gf]

Definition. *Given a sequence of real numbers* $\{a_0, a_1, a_2, \cdots\} = \{a_n\}_{n=0}^{\infty}$. *Definte* $A(S) = \sum_{n=0}^{\infty} a_n S^n \rightarrow$ *power series*

According to values of $\{a_n|_{n=0}^{\infty}$. We have

1. $A(S)$ only converges at $S = 0$
2. $A(S)$ converges when $|S| < R$ and diverges when $|S| > R$.
3. $A(S)$ converges when $|S| < \infty = \mathbb{R}$.

Example for case 2, $A(S) = \sum_{n=0}^{\infty} S^n = \frac{1}{1-S}$. $A(S)$ converges when $|S| < 1$ and diverges when $|S| > 1$ so $R = 1$.

Example for case 3, $A(S) = \sum_{n=0}^{\infty} \frac{S^n}{n!} = e^s$ for all $|S| < \infty$. Hence $R \to \infty$.

Our interest: Case 2 and Case 3 and $A(S)$ is called gf of $\{a_n\}_{n=0}^{\infty}$ and R is called convergence radius.

Theorem. This is one-to-one correspondence between $\{a_n\}_{n=0}^{\infty}$ and $A(S)$. That is

1. Given $\{a_n\}_{n=0}^{\infty}$, $A(S)$ is unique
2. Given $A(S)$, $\{a_n\}_{n=0}^{\infty}$ is unique

Given $A(S)$, we can find $\{a_n\}_{n=0}^{\infty}$ by Talor expansion.

$$A(S) = \sum_{n=0}^{\infty} \frac{A^{(n)}(0)}{n!} S^n$$

Next: Review for power series and two properties of gf. This will help us to get

1. Given $\{a_n\}_{n=0}^{\infty}$, find $A(S)$
2. Given $A(S)$, get $\{a_n\}_{n=0}^{\infty}$

4.2 Four Power Series

1. Geometric
$$A(S) = \sum_{n=0}^{\infty} s^n = \frac{1}{1-S} \& R = 1$$
$a_n = 1$ for $n \geq 0$.
2. Alternate Geometric
$$A(S) = \sum_{n=0}^{\infty} (-1)^n S^n = \frac{1}{1+S} \& R = 1$$
$a_n = (-1)^n$ for $n \geq 0$
3. Exponential
$$A(S) = \sum_{n=0}^{\infty} \frac{S^n}{n!} = e^S \& R = \infty$$
$a_n = \frac{1}{n!}$ for $n \geq 0$.

4. Binomial
$$A(S) = \sum_{k=0}^{n} \binom{n}{k} S^k = (1+S)^n \& R = \infty$$
since it is finite summation $a_k = \binom{n}{k}$, $k = 0, 1, 2, \cdots n$
In general
$$A(S) = \sum_{k=0}^{\infty} \binom{\alpha}{k} S^k = (1+S)^\alpha \& R = \infty$$
where $a_n = \binom{\alpha}{n} = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}$ for $n \geq 0$.

For example:
$$\binom{-\frac{1}{2}}{n} = (-\frac{1}{4})^n \binom{2n}{n}$$
A useful result for random walk.
Solution:
$$\binom{-\frac{1}{2}}{n} = (-\frac{1}{2})^n \frac{2n!}{n! \times 2^n n!} = 2^n n!$$
Simplify the form we have
$$\binom{-\frac{1}{2}}{n} = (-\frac{1}{4})^n \binom{2n}{n}$$

4.3 Properties of generating function

Suppose
$$A(S) = \sum_{n=0}^{\infty} a_n S^n \text{ and Radius } = R_A$$
$$B(S) = \sum_{n=0}^{\infty} b_n S^n \text{ and Radius } = R_B$$

1. Addition:
$$C(S) = A(S) + B(S) = \sum_{n=0}^{\infty} c_n S^n$$
$$c_n = a_n + b_n \text{ for } n \geq 0 \text{ Radius } = \min(R_A, R_B)$$

2.
$$C(S) = A(S)B(S) = (\sum_{n=0}^{\infty} a_n S^n)(\sum_{n=0}^{\infty} b_n S^n) = \sum_{n=0}^{\infty} c_n S^n$$
What is c_n?
$$c_n \neq a_n b_n$$

but

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}$$

Well c_n is the convolution of a_n and b_n.

Example

1. $C(S) = \frac{1}{(1-S)(1+S)}$,find $R_c \& c_n$
2. $C(S) = \frac{1}{(1-S)^2}$, find $R_c \& c_n$.

Solution:

1.

$$C(S) = \frac{1}{(1-S)(1+S)} = \frac{1}{S}(\frac{1}{1-S} + \frac{1}{1+S})$$

$$A(S) = \frac{1}{1-S} = \sum_{n=0}^{\infty} S^n \text{ and } R_A = 1$$

$$B(S) = \frac{1}{1+S} = \sum_{n=0}^{\infty} (-1)^n S^n \text{ and } R_B = 1$$

so

$$C(S) = \frac{1}{2}(A(S) + B(S)) = \frac{1}{2}\sum_{n=0}^{\infty}(1 + (-1)^n)S^n = \sum_{n=0}^{\infty}\frac{1}{2}(1 + (-1)^n)S^n$$

so

$$C_n = \frac{1}{2}(1 + (-1)^n) \text{ and } R_c = \min(R_A, R_B) = 1$$

2.

$$C(S) = \frac{1}{1-S}\frac{1}{1-S}$$

so

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} = \sum_{k=0}^{n}(1 \times 1) = n + 1, \text{ for } n \geq 0$$

and

$$R_c = \min(R_A, R_B) = 1$$

4.4 Probability generating function (pgf)

Definition. *Suppose X has range* $\{0, 1, 2, \cdots\} \cup \{\infty\}$

Let

$$P_n = P(X = n) \text{ for } n \geq 0$$

the pdf of X is defined as

$$G_X(S) = \sum_{n=0}^{\infty} P(X = n)S^n = \sum_{n=0}^{\infty} P_n S^n$$

Comment: If X is proper, i.e., $P(X = \infty) = 0$,

$$G_X(S) = \sum_{n=0}^{\infty} P(X = n)S^n = E[S^X]$$

Most situation, we have proper RV and need this formula.

4.4.1 Applications

1. If $G_X(S)$ is known or easy to find, we can find $P_n = P(X = n)$ from $G_X(S)$

   Two ways here

   (a) Taylor expansion

   $$G_X(S) = \sum_{n=0}^{\infty} \frac{G_X(0)^{(0)}}{n!} S^n \implies P_0 = P(X = 0) = G_X(0)$$

   and

   $$P_n = P(X = n) = \frac{G_0^{(n)}(0)}{n!} n \geq 1$$

   (b) B_1 fair power series and tow properties.

2. Check if X is proper or not based on $G_X(S)$ Note $P(X < \infty) = \sum_{n=0}^{\infty} P(X = n) = \sum_{n=0}^{\infty} P_n = G_X(1) \implies G_X(1) = 1 \implies X$ is proper $and \mathrm{G}_X(1) < 1 \implies X$ is improper

3. Calculate $E[X]$ and $Var(X)$ based on $G_X(S)$ [X is proper]

   Take the first derivative of $G_X(S)$,

   $$[G_X(S)]' = [\sum_{n=0}^{\infty} P_n S^n]' = \sum_{n=0}^{\infty} P_n \times nS^{n-1}$$

Set $S = 1$,

$$G_X(1)' = \sum_{n=0}^{\infty} P_n \times n = \sum_{n=0}^{\infty} P(X = n) \times n = E[X]$$

For $Var(X)$ take the second derivative.

$$G_X(S) = \sum_{n=0} P_0 \times n \times (n-1) \times S^{n-2}$$

Set $S = 1$, Hence

$$G''_X(1) = \sum_{n=0}^{\infty} P_n \times n \times (n-1) = \sum_{n=0}^{\infty} P(X = n) \times n \times (n-1) = E[X(X-1)]$$

so

$$Var(X) = E[X^2] - E[X]^2 = G''_X(1) + G'_X(1) - [G_X(1)']^2$$

Why pgf not moment generating function? Properties 1 and 2 takes that pdf can find $P(X = n)$ and moment of X but moment generating function cannot.

4. Uniquness: If $X \& Y$ have same pgf, they have same distribution, that is pdf determines distribution type.

5. Independence: If $X_1, \cdots, X_n$ are independent, then $G_{X_1+X_2+\cdots+X_n}(S) = \prod_{i=1}^{n} G_{X_i}(S)$

   Solution:

$$G_{X_1+X_2+\cdots+X_3} = E[S^{X_1+\cdots+X_n}] = E[S^{X_1} S^{X_2} \cdots S^{X_n}]$$

   Since $X_1, \cdots, X_n$ are independent,

$$G_{X_1+\cdots,X_n} = \prod_{i=1}^{n} E[S^{X_i}] = \prod_{i=1}^{n} G_{X_i}(S)$$

   Example: find pdf of

   (a) I_A with $P(A) = p$
   (b) $X \sim Bin(n, p)$
   (c) $X \sim Pois(\lambda)$

   Solution:

   (a) $G_{I_A}(S) = E[S^{I_A}] = S^0 P(I_A = 0) + S^1 P(I_A = 1) = PS + 1 - P$. Radius is just ∞ since it is a finite summation.

(b) $X \sim Bin(n, p)$

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

$I_1, \cdots, I_n$ are Bernoulli rvs. so

$$G_X(S) = G_{\sum_{i=1}^{\infty}}(S) = \prod_{i=1}^{n} G_{I_i}(S)$$

so

$$G_X(S) = \prod_{i=1}^{n}(1 - P + PS) = (1 - P + PS)^n$$

Converge Radius is ∞.

(c)

$$P(X = n) = \frac{\lambda^n}{n!}$$

$$G_X(S) = E[S^X] = \sum_{n=0}^{\infty} P(X = n)S^n = \sum_{n=0}^{\infty} \lambda^n e^{-\lambda} \frac{S^n}{n!}$$

so

$$G_X(S) = \sum_{n=0}^{\infty} \frac{(\lambda S)^n}{n!} e^{-\lambda} = e^{\lambda s - \lambda}$$

Converge Radius is ∞.

Example: Suppose $G(S) = \frac{1}{4}S^{-1}/(1 - \frac{3}{4}S)$

1. $G(S)$ is pdf of a proper RV X.
2. Find $E[X]$.

Solution:

1. Need to show

$$G(S) = \sum_{n=0}^{\infty} P_n S^n$$

and $P_n \geq 0$ and $\sum_{n=0}^{\infty} P_n = G(1) = 1$.. It is easy to check

$$\sum_{n=0}^{\infty} P_n = G(1) = \frac{\frac{1}{4}}{1 - \frac{3}{4}} = 1$$

done.

Now,

$$G(S) = \frac{1}{4}S^4 \times \frac{1}{1-\frac{3}{4}S} = \sum_{n=0}^{\infty}(\frac{3}{4}S)^n$$

so

$$G(S) = \frac{1}{4}S^4 \sum_{n=0}^{\infty}(\frac{3}{4})^n S^n = \sum_{n=0}^{\infty}\frac{1}{4}(\frac{3}{4})^n S^{n+4}$$

We change index

$$m = n+4 \implies n = m-4 = \sum_{m=4}^{\infty}\frac{1}{4}(\frac{3}{4})^{m-4}S^m$$

.

Therefore

$$P_0 = P_1 = P_2 = P_3 = 0$$

$$P_n = \text{coefficient of } S^n = \frac{1}{4}(\frac{3}{4})^{n-4} \text{ for } n \geq 4$$

Hence $G(S)$ is pdf of proper RV X.

2. $E[X] = G'(1) = 7$ (it is easy)

5 Renewal Process

Background: Suppose we have a sequence of RVs. $\{X_1, X-2, \cdots\} = \{X_n\}_{n=1}^{\infty}$ called a stochastic process. For example, let

$$\lambda = \text{event based on } \{X_n\}_{n=1}^{\infty}$$

$$\lambda = \text{“SF” and } X_n = \text{outcome on the nth trial}$$

λ occurs on the nth trial if $X_{n-1} = S$ and $X_n = F$.

Our interest

1. Can we observe it or not?
2. How long on average will it take?

Let

$$T_\lambda = \text{Waiting time for first } \lambda$$

so we need to find $P(T_\lambda < \infty)$ and $F(T_\lambda)$

Example we covered before

1. P(observe “333” in the sequence) = 1.
2. $E[T_{SF}] = \frac{1}{P(1-P)}$, $E[T_{SS}] = \frac{1}{p^2} + \frac{1}{p}$

   Expectation of T_{SF}, T_{SS} have different forms; hat are the difference between SF and SS?

5.1 Classification of Events

Notation:

$$T_{\lambda}^{(K,K+1)} = \text{waiting time for kth } \lambda \text{ given that we have } \lambda k \text{ times}, k = 0, 1, 2, \cdots$$

5.1.1 Definition of Renewal Event

If all waiting time RVs:

$$T_{\lambda}, T_{\lambda}^{(1,2)}, T_{\lambda}^{(2,3)}, \ldots$$

are iid, then λ is called a renewal events.

Information here

The occurrence of an event λ does not help the second event; No-memory property is here.

5.1.2 Definition of delayed renewal event

1. $T_{\lambda}^{(1,2)}, T_{\lambda}^{(2,3)}, \cdots$ are iid.

2. but T_{λ} and $T_{\lambda}^{(1,2)}, T_{\lambda}^{(2,3)}, \cdots$ have different distribution.

Comments

1. first event and second event, third event and so on are different.

2. Once we observe first event, other events (second event, third event) are the same. The waiting time for other event have same distributions.

5.1.3 Associate Renewal Event of a delayed renewal event

Once we observe first λ the waiting times have same distributions and λ becomes a renewal event. This renewal event is called associated renewal event and denote it by $\tilde{\lambda}$.

Example, Consider a sequence of Bernoulli Trials $\lambda_1 =$ "SF" and $\lambda_2 =$ "SS".

Claim:λ_1 is renewal event, and λ_2 is delayers renewal event.

Solution: first argue λ_2 is delayed renewal.

$$P(T_{\lambda_2} = 1) = P(\text{observe "SS" in the first trial}) = 0$$

$$P(T_{\lambda_2}^{(1,2)}) = P(S) = P$$

$$P(T_{\lambda_2}^{(2,3)} = 1) = P(S) = P$$

Only ones to get the third.

Next argue λ, is renewal event.

$$P(T_{\lambda_1} = 1) = 0, P(T_{\lambda_1} = 2) = P(\text{first is S and second is F}) = P(1 - P)$$

$$P(T_{\lambda_1}^{(1,2)} = 1) = 0, P(T_{\lambda_1}^{(1,2)} = 2) = P(S)P(F) = P(1 - P)$$

Looks that: all waiting times have same distribution and λ_1 is renewal event. For "SF", the all occurrence of SF does not affect second event (No memory) $\implies$ renewal.

For "SS", the occurrence of "SS" affects second SS so delayed renewal.

General Rule for determining if λ is renewal or delayed renewal

Consider two consecutive events or two events in a row:

1. If there is no overlap between two events, λ is renewal
2. I there is some overlap, then λ is delayed renewal.

5.1.4 Example

$\lambda_1 = "SF"$ two consecutive events: firstλ_1 is "SF" and second λ_2 is "SF". There is no overlap so λ_1 is renewal events. $\lambda_2 = "SS"$, two consecutive events: the first λ_2 is SS and second λ_2 is also SS. Well there will be an overlap between these two events so λ_2 is delayed renewal; overlap "S" helps the second λ_2.

5.1.5 Example

Toss a fair die

$$X_n = \text{ the number shown on the nth toss}$$

λ_1 = "123"; λ_2 = "123", λ_3 = "1212". Classify all three events:

Solution:

λ_1 : two events in a row 123123 No overlap, λ_1 is renewal.

λ_2 : two events in a row 121212 $\implies$ λ_2 is delayed renewal.

λ_3 : overlap= "12"; λ_3 is delayed renewal.

Next we study renewal events

1. Classify renewal event, Let $f_\lambda = P(T_\lambda < \infty)$ where f_λ is the probability of finally observing λ.
   - If $f_\lambda < 1$, λ is called transient
   - If $f_\lambda = 1$, λ is called recurrent.

- If $f_\lambda = 1$ and $E[T_\lambda] = \infty$, λ is called null recurrent.
- If $f_\lambda = 1$ and $E[T_\lambda] < \infty$, λ is called positive recurrent.

2. Comments

   (a) If λ is transient, $E[T_\lambda] = \infty$ since $P(T_\lambda = \infty) > 0$.

   (b) If $\mathrm{E}[\mathrm{T}_\lambda] < \infty$, λ is positive recurrent.

   For example: "SF" is renewal event. $E[T_{SF}] = \frac{1}{p(1-p)} < \infty, [0 < p < 1]$. This implies "SF" is positive recurrent.

3. Consider v_λ = number of events observed in the sequence and range of this is $\{0, 1, \cdots, \} \cup \{\infty\}$.

   Theorem. Distribution of v_λ

$$\begin{cases} P(v_\lambda = k) = f_\lambda^k(1 - f_\lambda), k = 0, 1, 2, \cdots \\ P(v_\lambda = \infty) = f_\lambda^\infty \end{cases}$$

   Proof. $P(v_\lambda = k) = P(T_\lambda < \infty, T_\lambda^{(1,2)} < \infty, \cdots, T_\lambda^{(k,k+1)} < \infty) =_{iid} P(T_\lambda < \infty) \cdots P(T_\lambda^{(k,k+1)} < \infty) = f_\lambda^k(1 - f_\lambda)$

   $P(v_\lambda = \infty) = P(T_\lambda < \infty, T_\lambda^{(1,2)} < \infty, \cdots) = f_\lambda^\infty$

   □

   Comments: If $f_\lambda = 1, P(v_\lambda) = \infty) = 1$ and $f_\lambda < 1, P(v_\lambda) = \infty) = 0$. For the expectation of v_λ

$$E[v_\lambda] = \frac{f_\lambda}{1 - f_\lambda}$$

   [only fr renewal event]. Note this is only for renewal event but not for delayed renewal event.

   Proof. If $f_\lambda < 1, P(v_\lambda = \infty) = 0$ and $E[v_\lambda] = \sum_{k=0}^\infty P(v_\lambda = k) \times k$ so $E[v_\lambda] = \sum_{k=0}^\infty f_\lambda^k(1 - f_\lambda) \times k = f_\lambda \frac{1}{1-f_\lambda}$

   If $f_\lambda = 1$, then automatically

$$E[v_\lambda] = \frac{f_\lambda}{1 - f_\lambda} = \infty$$

   Therefore summarize two parts,

$$E[v_\lambda] = \frac{f_\lambda}{1 - f_\lambda}$$

   □

4. Quick summary:

   (a) $f_\lambda = 1 \iff \lambda$ is recurrent $\iff P(v_\lambda = \infty) = 1$ and $E[v_\lambda] = \infty$.
   (b) $f_\lambda < 1 \iff \lambda$ is transient $\iff E[v_\lambda] = \frac{f_\lambda}{1-f_\lambda} < \infty \& P(v_\infty) = 0$

5.2 Renewal Sequence [for renewal event]

Definition. *The renewal sequence associated with a renewal event λ is defined as $\gamma_0 = 1$ and $\gamma_n = P($observe λ at the nth trial$)$ for $n \geq 1$.*

For example, "SF". $\gamma_0 = 1, \gamma_n = P($observe "SF" at the nth trial$) = P(X_{n-1} = S, X_n = F) = 0$ for $n \geq 1$. For $n \geq 2$, $\gamma_n = p(1-p)$

Comment: $\{\gamma_n\}_{n=0}^{\infty}$ in general is easy to find and it has a lot of application.

Theorem. $E[v_\lambda] = \sum_{n=1}^{\infty} \gamma_n = \begin{cases} \infty & \lambda \text{ is recurrent} \\ < \infty & \lambda \text{ is transient} \end{cases}$

SF example $\gamma_0 = 1, \gamma_1 = 0, \gamma_n = p(1-p)$ for $n \geq 2$.

$$E[v_\lambda] = \sum_{n=1}^{\infty} \gamma_n = \sum_{n=2}^{\infty} p(1-p) = \infty$$

SF is recurrent

Proof. Let $I_n = \begin{cases} 1 & \text{if } \lambda \text{ occurs at the nth trial} \\ 0 & OW \end{cases}$. Then $v_\lambda = \sum_{i=1}^{\infty} I_n$. $\therefore E[v_\lambda] = E[\sum_{n=1}^{\infty} I_n] = \sum_{n=1}^{\infty} E[I_n] = \sum_{n=1}^{\infty} r_n$. □

5.2.1 Example

Toss a fair die. X_n = number of shown on the nth toss, λ = "123", λ occurs on the nth toss if $x_{n-2} = 1$ and $x_{n-1} = 2, x_n = 3$ (i.e. in a sequence). Claim: λ is recurrent.

Solution: $r_0 = 1, r_1 = r_2 = 0$. (we need at least 3 tosses to get "123")

$f_3 = P(x_1 = 1, x_2 = 2, x_3 = 3) = (\frac{1}{6})^3$

In general, for $n \geq 3$. $r_n = P(x_{n-2} = 1, x_{n-1} = 2, x_n = 3) = \frac{1}{6}^3$.

$E[v_\lambda] = \sum_{i=3}^{\infty} r_n = \infty$. Therefore λ is recurrent.

5.3 Renewal Relationship

Still for renewal event. In this section, calculate pgf of T_λ based on $\{r_n\}_{n=0}^{\infty}$.

Definition. *First waiting time probabilities* $\begin{cases} f_n = P(\textit{first observed } \lambda \textit{ at the nth time }) = P(T_\lambda = n) & n \geq 1 \\ f_0 = 0 \end{cases}.$

Let $F_\lambda(s) = \sum_{n=0}^{\infty} P(T_\lambda = n)s^n = \sum_{n=0}^{\infty} f_n s^n =$ *pgf of* T_λ

Let $R_\lambda(s) = \sum_{n=0}^{\infty} r_n s^n \implies$ *gf of* $\{r_n\}_{n=0}^{\infty}$

It is easy to find since $\{r_n\}_{n=0}$ *is easy to find.*

Theorem. Renewal relation: $F_\lambda(s) = 1 - \frac{1}{R_\lambda(s)}$

Proof. Step 1: find relationship between $\{r_n\}_{n=0}^{\infty}$ and $\{f_n\}_{n=0}^{\infty}$

For $n \geq 1$, $r_n = P(\text{observe } \lambda \text{ at nth trial})$. Condition on the first occurrence time where $T_\lambda = 1, 2, \cdots, n$.

Then $r_n = \sum_{k=1}^{n} P(\text{observe } \lambda \text{ at nth trial} | T_\lambda = k)P(T_\lambda = k)$.

$P(\text{observe } \lambda \text{ at nth trial} | T_\lambda = k) = P(\text{observe } \lambda \text{ at (n-k)th trial} | \lambda \text{ is observed at time } 0)$

$\therefore$ λ is renewal event.

$\therefore$ the occurrence of λ does not affect next λ. thus $P(\text{observe } \lambda \text{ at nth trial} | T_\lambda = k) = P(\text{observe } \lambda \text{ at n -k}) = r_{n-k}$ Hence $r_n = \sum_{k=1}^{n} r_{n-k} f_k$. when $k = 0, r_{n-k} f_k = r_n f_0 = 0$. Hence $r_n = \sum_{k=0}^{n} r_{n-k} f_k$ for $n = 1$ and $r_0 = 1$.

Aside from the proof: how to use $r_n = \sum_{k=0}^{n} r_{n-k} f_k, n \geq 1$.

When $n = 1$, $r_1 = r_1 f_0 + r_0 f_1 = r_0 f_1$. When $n = 2$, $r_2 = r_2 f_0 + r_1 f_1 + r_0 f_2 = r_1 f_1 + r_0 f_2$ given r_0, r_1, r_2 to get f_1, f_2.

If we continue, we can find $\{f_n\}_{n=0}^{\infty}$ by $\{r_n\}_{n=0}^{\infty}$ why do we define $r_0 = 1$? It has been used in $r_n = \sum_{k=0}^{n} r_{n-k} f_k$.

When $n = 1$, $r_1 = r_0 f_1$. Note that $r_1 = f_1$ by definition. Therefore $r_0 = 1$

Step 2: find relationship between $F_\lambda(s)$ and $R_\lambda(s)$.

From $r_n = \sum_{k=0}^{n} r_{n-k} f_k, n \geq 1, r_0 = 1$.

$$R_\lambda(s) = \sum_{n=0}^{\infty} c_n s^n = 1 + \sum_{n=1}^{\infty} r_n s^n$$

Plug in, we get $R_\lambda(s) = 1 + \sum_{n=1}^{\infty} (\sum_{k=0}^{n} r_{n-k} f_k) s^n$ for $n = 0, \sum_{k=0}^{n} r_{n-k} f_k = \sum_{k=0}^{0} r_{0-k} f_k = r_0 f_0 = 0$

$\implies$ $R_\lambda(s) = 1 + \sum_{n=0}^{\infty} f_n s^n c_n \sum_{n=0}^{\infty} r_n s^n$. Here $a_n = f_n$ and $b_n = r_n$. Hence $R_\lambda(s) = 1 + F_\lambda(s) R_\lambda(s)$.

Therefore we prove the relation. □

From $F_\lambda(s)$, we can

1. Check if n is recurrent or not. Or check $f_\lambda = F_\lambda(1) = \sum_{n=0}^{\infty} f_n = 1$ or not
2. $E[T_\lambda] = F'_\lambda(1)$.
3. $f_n = P(T_\lambda = n) = \frac{F_\lambda^{(n)}(0)}{n!}$ for $n \geq 1$.

5.3.1 Example

Toss a fair die, $\lambda =$ "123". Show

1. λ is recurrent

   Renewal sequence

   $r_0 = 1, r_1, r_2 = 0$ since we need at least 3 trials to get 123.

   $r_n = P(X_{n-2} = 1, X_{n-1} = 2, X_{n-3}) = (\frac{1}{6})^3$ for $n \geq 3$.

   Next: $f_\lambda(s) = \sum_{n=0}^{\infty} r_n s^n = 1 + \sum_{n=3}^{\infty} (\frac{1}{6})^3 S^n = 1 + \frac{(\frac{1}{6})^3 S^3}{1-S}$

   $F_\lambda(S) = 1 - \frac{1}{R_\lambda(S)} = 1 - \frac{1-S^2}{1-S+(\frac{1}{6})^3 S^3}$.

   $f_\lambda = F_\lambda(1) = 1$

2. λ is positive recurrent

$$F_\lambda(S) = \frac{1 - S + (\frac{1}{6})^3 S^3 - (S-1)[1 - S + \frac{1}{6}^3 S^3]'}{[1 - S + \frac{1}{6}^3 S^3]^2}$$

   Therefore $F(T_\lambda) = F_\lambda(1) = \frac{(\frac{1}{6})^3 S^3}{[(\frac{1}{6})^3 S^3]^2} = 6^3$.

   Hence $F(T_\lambda) = \frac{1}{P(1)P(2)P(3)} < \infty$. Therefore λ is positive recurrent

3. Find f_5, f_6, f_7.

   In general, $f_n = \frac{F_\lambda^{(n)}(0)}{n!}, n \geq 1$.

   Find f_5, f_6, f_7 by definition.

   $f_5 = P(T_\lambda) = 5 = P($ first observation of 5th trial$) = \frac{1}{6}^3$

   $f_6 = P(T_\lambda = 6) = P(X_4 = 1, X_5 = 2, X_6 = 3 \& X_1, X_2, X_3) = \frac{1}{6}^3[1 - \frac{1}{6}^3]$.

   $P_7 = P(T_\lambda = \frac{1}{6}^3[1 - 2\frac{1}{6}^3]$.

5.4 Delayed Renewal Relation

1. $T_\lambda^{(1,2)}, T_\lambda^{(2,3)}, \cdots$ are iid.

2. T_λ and $T_\lambda^{(1,2)}, T_\lambda^{(2,3)}$ have different distribution.

This means: once we observe first λ, λ becomes a renewal event. This renewal event is called associate renewal event of λ. Further we denote this renewal event by $\tilde{\lambda}$. purpose here is to find the pgf of T_λ.

Delayed renewal sequence : $d_0 = 0$, $d_n = P(\lambda$ occurs at the nth trial $)$ for $n \geq 1$.

Associated renewal sequence or renewal sequence of $\tilde{\lambda}$ $\tilde{r}_0 = 1, \tilde{r}_1 = P($observe $\tilde{\lambda}$ at nth trial$)$.
since when we observe first λ is not important for $\tilde{r}_n$, we assume that we observe first λ at zeroth trial. By argument, $\tilde{r}_n$ =P(λ occurs at nth trial$|\lambda$ is observed at zeroth trial).

gf of $\{d_n\}_{n=0}^{\infty}$ $D_\lambda(s) = \sum_{n=0}^{\infty} d_n S^n$.

gf of $\{\tilde{r}_n\}_{n=0}^{\infty}$ $R_{\tilde{\lambda}}(S) = \sum_{n=0}^{\infty} \tilde{r}_n S^n$.

pgf of T_λ $f_n = P(T_\lambda = n), n \geq 1$ and $f_0 = 0$.
$F_\lambda(S) = \sum_{n=0}^{\infty} f_n S^n$: pgf same as before of T_λ.

Theorem. Delayed renewal relation: $F_\lambda(S) = \frac{D_\lambda(S)}{R_{\tilde{\lambda}}(S)}$.

Once we have $F_\lambda(S)$, we can

1. $f_\lambda = P(T_\lambda = \infty) = F_\lambda(1)$.
2. $F(T_\lambda) = F_\lambda(1)$
3. $f_n = \frac{F_\lambda^{(n)}(0)}{n!}$.

How to find $F(T_\lambda^{(1,2)})$?
Hint: $T_\lambda^{(1,2)} = T_{\tilde{\lambda}}$ waiting time for first $\tilde{\lambda}$.

1. Find $R_{\tilde{\lambda}}(S)$: gf of $\{\tilde{S}_n\}_{n=0}^{\infty}$.
2. $F_{\tilde{\lambda}}(S) = 1 - \frac{1}{R_{\tilde{\lambda}}(S)}$: pgf of $T_{\tilde{\lambda}}$.
3. $F(T_\lambda) = F_{\tilde{\lambda}}(1)$.

Example: toss a count, $P(H) = P$, $\lambda =$ "HH". $d_0 = 0, d_1, d_n = P(X_{n-1} = H, X_n = H) = P^2$ for $n \geq 2$.
$\tilde{r}_n$:
$\tilde{r}_0 = 1, \tilde{r}_1 = P = P(X_1 = H)$. $\tilde{r}_n = P(X_{n-1} = H, X_n = H) = P^2$.

5.4.1 Example

Toss a fair coin. $\lambda =$ "121"

1. $F_\lambda(s)$
2. $E[T_\lambda]$ and $E[T_\lambda^{(1,2)}]$.

3. average number of trials to see 5 λs.

4. Find f_5 and f_6

5. P(number of time observing $\lambda = \infty$)

Solution: λ is delayed renewal and overlap is "1". Delayed renewal sequence $d_0 = 0, d_1 = d_2 = o$[Need at least 3 trials], $d_n = P(X_{n-2} = 1, X_{n-1} = 1, X_n = 1) = \frac{1}{6}^3$ for $\lambda \geq 3$.

$$D_\lambda(S) = \sum_{n=0}^{\infty} d_n S^n = \sum_{n=3}^{\infty} \frac{1}{6}^3 S^n = \frac{1}{6}^3 \frac{S^3}{1-S}$$

Associated renewal sequence, $\tilde{r}_0 = 1, \tilde{r}_1 = 0, \tilde{r}_2 = P(\text{observe } \lambda \text{ at the second trial}) = P(X_1 = 2, X_2 = 1) = \frac{1}{6}^2$

$$\tilde{r}_3 = P(X_1 = 1, X_2 = 2, X_3 = 1) = \frac{1}{6}^3$$

In general $\tilde{r}_n = P(X_{n-2} = 1, X_{n-1} = 2, X_n = 1) = \frac{1}{6}^3$ so

$$R_{\tilde{\lambda}}(S) = \sum_{n=0}^{\infty} \tilde{r}_n S^n = 1 + 0 + \frac{1}{6}^2 S^2 + \sum_{n=3}^{\infty} \frac{1}{6}^3 S^n = 1 + \frac{1}{6}^2 S^2 + \frac{\frac{1}{6}^3 S^3}{1-S}$$

Therefore

$$D_\lambda(S) = \frac{1}{6}^3 \frac{S^3}{1-S}$$

Hence

$$F_\lambda(S) = \frac{D_\lambda(S)}{F_{\tilde{\lambda}}(S)} = \frac{\frac{1}{6}^3 S^3}{[1 + \frac{1}{6}^2 S^2](1-S) + \frac{1}{6}^3 S^3}$$

$$F_\lambda(1) = P(T_\lambda < \infty) = 1$$

$\implies$ we can finally observe λ.

$$E[T_\lambda] = F'_\lambda(1)$$

Check

$$F'_\lambda(1) = 6 + 6^3 = \frac{1}{P(\text{overlap})} + \frac{1}{P(\text{"121"})}$$

Now $E[T_\lambda^{(1,2)}]$ and pgf of $T_{\tilde{\lambda}} = T_\lambda^{(1,2)}$

$$F_{\tilde{\lambda}}(S) = 1 - \frac{1}{R_{\tilde{\lambda}}(S)} = 1 - \frac{1-S}{(1 + \frac{1}{6}^3 S^2)(1-S) + \frac{1}{6}^3 S^3}$$

$$E[T_{\tilde{\lambda}}] = E[T_\lambda^{(1,2)}] = F'_{\tilde{\lambda}}(1)$$

Check

$$F'_\lambda(1) = 6^3 = \frac{1}{P(121)}$$

(3) Need to find

$$E[T_\lambda + T_\lambda^{(1,2)} + T_\lambda^{(2,3)} + T_\lambda^{(3,4)} + T_\lambda^{(4,5)}] = E[T_\lambda] + 4E[T_\lambda^{(1,2)}] = 6 + 6^3 + 3 \times 6^3$$

(4)

$$\begin{aligned} f_5 &= P(T_\lambda = 5) = P(X_3 = 1, X_4 = 2, X_5 = 1, (X_1, X_2) \neq (1,2)) \\ &= P(X_3 = 1, X_4 = 2, X_5 = 1)P((X_1, X_2) \neq (1,2)) \\ &= \frac{1}{6}^3 [1 - \frac{1}{6}^2] \end{aligned}$$

$$\begin{aligned} f_6 &= P([X_4, X_5, X_6] = 121) \times P((X_1, X_2, X_3) \neq 121 \qquad\qquad (X_2, X_3) \neq 12) \\ &= \frac{1}{6}^3 [1 - P((X_1, X_2, X_3) = 121) - P((X_2, X_3) = 12)] \\ &= \frac{1}{6}^3 [1 - \frac{1}{6}^3 - \frac{1}{6}^2] \end{aligned}$$

(5)

$$\begin{aligned} P(\text{observe } \lambda \text{ infinite times}) &= P(\text{observe } \lambda \text{ infinite times}|\text{observe } \lambda \text{ once}) \times P(\text{observe } \lambda \text{ once}) \\ &= P(V_{\tilde{\lambda}} = \infty) = f_{\tilde{\lambda}}^\infty = [F_{\tilde{\lambda}}(1)]^\infty = 1 \end{aligned}$$

so $P($ observe λ infinite times$) = 1$.

5.5 Renewal Theorem

Calculate $E[T_\lambda] = \begin{cases} \lambda \text{ is renewal} \\ \lambda \text{ is delayed renewal} \end{cases}$

Some notations:

Renewal event

- Period: let $d = gcd\{n | Y_n > 0, n \geq 1\}$. d is called period. If $d = 1$, aperiodic; if $d > 1$, periodic.

  E.G., "123" example: $r_0 = 1, r_1 = r_2 = 0, r_n = \frac{1}{6}^3$ for $n \geq 3$. $d = gcd\{n | r_n > 0, n \geq 1\} = gcd\{3, 4, 5, \cdots\} = 1$, aperiodic. Check: any fixed pattern, period= 1.

  E.G. (periodic), We cover one example in 5.6.

  Two comments:

1. aperiodic: means there is a warm-up period and after this period, $r_n > 0$. "123" warm up period $r_1 = r_2 = 0$, $r_n > 0$ for $n \geq 3$.
2. periodic: only when index is a multiple of d, we can have positive probabilities.

Theorem.

$$E[T_\lambda] = \begin{cases} \frac{1}{\lim_{n\to\infty} r_n} & d = 1 \\ \frac{d}{\lim_{n\to\infty} r_{nd}} & d > 1 \end{cases}$$

Example 5.6:

1. Toss a coin $\lambda_1 =$ "123" and $\lambda_2 =$ "123456". $E[T_{\lambda_1}]$ and $E[T_{\lambda_2}]$

   Solution: $\lambda_1 : d = 1$ and $r_n = \frac{1}{6}^3$ for $n \geq 3$, so $E[T_{\lambda_1}] = \frac{1}{\lim_{n\to\infty} r_n} = \frac{1}{\frac{1}{6}^3} = 6^3$

   λ_2: $r_0 = 1, r_1 = r_2 = \cdots, r_5 = 0$ and $r_n = \frac{1}{6}^6$ for $n \geq 6$. $d = gcd\{6, 7, 8, 9, \cdots\} = 1$ and $E[T_{\lambda_2}] = \frac{1}{\lim_{n\to\infty} r_n} = 6^6$

   Comments:

   (a) $d = 1$ for fixed pattern

   (b) only need to have r_n for $n \geq$ number of letters in pattern.
2. Toss a coin, $P(H) = p, 0 < p < 1$. $\lambda_3 =$ "H" and $\lambda_4 =$ "HHT"

   Solution: $\lambda_3 =$ "HT" : $d = 1$. Hence $r_n = P(1-P)$ for $n \geq 2$.

   $E[T_\lambda] = \frac{1}{P(1-P)}$

   $\lambda_4 =$ "HHT", $d = 1, r_n = P^2(1-P)$ for $n \geq 3$. $E[T_{\lambda_4}] = \frac{1}{P^2(1-P)}$

Delayed Renewal Event λ is a displayed renewal event.

Theorem.

$$E[T_\lambda] = E[T_{overlap}] + E[T_{\tilde{\lambda}}]$$

$$\text{Overlap} = \text{Overlap between two consecutive events}$$

$$T_{overlap} = \text{waiting time for first overlap}$$

$$T_{\tilde{\lambda}} = \text{waiting time for the first } \tilde{\lambda}$$

Understanding

1. To observe λ, we need to observe overlap E.G. $\lambda =$ "HH", overlap $=$ "H" to observe "HH", we need to observe "H".

$$T_\lambda = T_{\text{overlap}} + T_{\lambda|\text{overlap is observed}}$$

   Hence
$$E[T_\lambda] = E[T_{overlap}] + E[T_\lambda|\text{overlap is observed}]$$

2. argue $T_{\tilde{\lambda}}$ and $T_{\lambda|\text{overlap is observed}}$ have same distribution. Therefore $E[T_{\tilde{\lambda}}] = E[T_\lambda|\text{overlap is observed}]$
   Use "HH" as example: "H" is overlap: 1st $= H$, $T_{\lambda|\text{overlap is observed}} = 1$ is observed. $1st = T$, $T_{\lambda|\text{overlap is observed}} = 1 + T_{HH}$.
   $T_{\tilde{\lambda}}$: here "HH" is observed. $1st = H$ and $T_{\tilde{\lambda}} = 1$,
   $1st = T$ and $T_{\tilde{\lambda}} = 1 + T_{HH}$, so $T_{\tilde{\lambda}}$ and $T_{\lambda|\text{overlap is observed}}$ have same distribution and same expectation.
   Intuition:
   (a) Only overlap will be useful and other parts are useless.
   (b) because (1), we observe other parts or not does not help. This implies observe "overlap" and observe whole effect have same effect on waiting tim for λ.
   Therefore $T_{\tilde{\lambda}}$ and $T_{\lambda|\text{overlap is observed}}$ have same distribution and expectation so $E[T_\lambda] = E[T_{\text{overlap}}] + E[T_\lambda|\text{overlap is observed}] = E[T_{\text{overlap}}] + E[T_{\tilde{\lambda}}]$
   How to apply?
   (a) $E[T_{\tilde{\lambda}}] = \frac{1}{\lim_{n\to\infty} \tilde{r}_n}$, $d = 1$ apply renewal theorem for renewal event
   (b) $E[T_{overlap}]$
      i. overlap is renewal apply renewal theorem for renewal event.
      ii. overlap is delayed renewal. Continue partition.

Example 5.7: Toss a die $\lambda_1 =$ "121", $\lambda_2 =$ "12121", find $E[T_{\lambda_1}]$ and $E[T_{\lambda_2}]$.
Solution: λ_1 : overlap = 1, $E[T_{\lambda_1}] = E[T_{\text{“}1\text{”}}] + E[T_{1\tilde{2}1}] = \frac{1}{\frac{1}{6}} + \frac{1}{\frac{1}{6}^3}$

λ_2 : overlap = "121", $E[T_{\lambda_2}] = E[T_{121}] + E[T_{12\tilde{1}21}] = E[T_1] + E[T_{1\tilde{2}1}] + E[T_{12\tilde{1}21}]$ so $E[T_{\lambda_2}] = \frac{1}{\frac{1}{6}} + \frac{1}{\frac{1}{6}^3} + \frac{1}{\frac{1}{6}^5}$

5.6 Random Walk

Background suppose we have particle starting from 0. Each step, it can move to the right by 1 unit with probability p, and can move to the left by 1 unit with probability $(1-p)$.

For example, toss a coin such that H goes to right and T goes to the left. If we toss a coin 5 times and get HHTTT.

Definition. *Let* $x_0 = 0$ *(starting point) and* x_n *is the position of particle after n steps, then* $\{x_n\}_{n=0}^{\infty}$ *is called simple or ordinary random walk.*

We are interested in

1. λ_{00} is the returning to 0 given the process starts from 0. Let $T_{\lambda_{00}}$ be the waiting time for the first λ_{00} We could like to figure out $f_{\lambda_{00}} = P(T_{\lambda_{00}} < \infty)$ and $E[T_{\lambda_{00}}]$.

2. λ_{0k} is the visiting k given that process starts from 0. Let $T_{\lambda_{0k}}$ is the waiting time for visiting k given the process starts from 0. We would like to figure out $f_{\lambda_{0k}} = P(T_{\lambda_{0k}} < \infty)$ and $E[T_{\lambda_{0k}}]$.

Well, λ_{00} is returning to 0 given the process starts from 0.

Step 1 Is λ_{00} a renewal or delayed renewal event? It is pretty intuitive that T_λ and $T_\lambda^{(1,2)}$ have the same distribution. Hence, it is a renewal event.

Step 2 Use pgf of $T_{\lambda_{00}}$ to find $f_{\lambda_{00}}$ and $E[T_{\lambda_{00}}]$

$r_0 = 1$, $r_n = P(\text{return to 0 after n steps given the process starts from 0})$,

$r_{2n-1} = 0$: since process cannot return to 0 after odd number of steps.

$r_{2n} = P(\text{number of moments to right} = \text{number of moments to the right} = n)$

r_{2n} is just a $Bin(2n, p)$ r.v. equal to n. Therefore $r_{2n} = \binom{2n}{n} p^n q^n$

Aside result: $d = gcd\{n | r_n > 0 \& n > 0\} = gcd(2, 4, 6, \cdots) = 2$ (periodic). Move back to gf of $\{r_n\}_{n=0}^{\infty}$, $r_0 = 1$ and $r_{2n-1} = 0$. Therefore

$$R_{\lambda_{00}}(s) = \sum_{n=0}^{\infty} r_n s^n = 1 + \sum_{n=1}^{\infty} \binom{2n}{n} p^n q^n s^{2n}$$

Note: $1 = \binom{2n}{n} p^n q^n s^{2n}$ for $n = 0$. Hence

$$R_{\lambda_{00}}(s) = \sum_{n=0}^{\infty} \binom{2n}{n} p^n q^n s^{2n}$$

Recall $\binom{-\frac{1}{2}}{n} = (-\frac{1}{4})^n \binom{2n}{n}$ and $\binom{2n}{n} = (-4)^n \binom{-\frac{1}{2}}{n}$. Then

$$R_{\lambda_{00}}(s) = \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} (-4)^n p^n q^n s^{2n} = (1 - 4pqs^2)^{-\frac{1}{2}}$$

Therefore $T_{\lambda_{00}}$'s pgf is

$$F_{\lambda_{00}} = 1 - \frac{1}{R_{\lambda_{00}}(s)} = 1 - \sqrt{1 - 4pqs^2}$$

1. $f_{\lambda_{00}} = P(T_{\lambda_{00}} < \infty) = F_{\lambda_{00}}(1)$.
2. $E[T_{\lambda_{00}}] = F'_{00}(1)$
   $f_{\lambda_{00}} = F'_{00}(1) = 1 - \sqrt{1-4pq} = 1 - |p-q|$
   Then $f_{\lambda_{00}} = \begin{cases} 1 & p = q = 1/2 \\ < 1 & p \neq q \text{ or } p \neq 1/2 \end{cases}$. Therefore $p = 1/2$, λ is recurrent; otherwise, transient.
   Question: $p = 1/2$ is λ positive recurrent or null recurrent?
   Find $E[T_{\lambda_{00}}]$ for $p = 1/2$. $F_{\lambda_{00}}(s)$ when $p = 1/2 = 1 - \sqrt{1-s^2}$
   $F'_{\lambda_{00}}(1) = E[T_{\lambda_{00}}] = \infty$. Hence λ is null recurrent.

One more concept: $v_{\lambda_{00}}$ is the number of times returning to 0 given the process starts from 0. For renewal event

$$E[v_{\lambda_{00}}] = \frac{f_{\lambda_{00}}}{1 - f_{\lambda_{00}}} = \frac{1-|p-q|}{|p-q|}$$

λ_{0k}: visiting k given the process start from 0 start with λ_{01}

Step 1 Is this a renewal or delayed renewal? delayed renewal.

Step 2 We use the delayed renewal relationship to get the pgf.

λ_{00}: returning to 0 and starting from 0.

1. It is renewal event.

2. period $d = 2$.

3. $p = q = \frac{1}{2}$, λ_{00} is null recurrent where $E[T_{\lambda_{00}}] = \infty$ and $E[v_{\lambda_{00}}] = \infty$ and $f_{\lambda_{00}} = 1$.
   $p \neq q$, λ_{00} is transient, $f_{\lambda_{00}} = 1 - |p-q|$
   or $p \neq \frac{1}{2}$, $E[T_{\lambda_{00}}] = \infty$, $E[v_{\lambda_{00}}] = \frac{1-|p-q|}{|p-q|}$

We argue

$$R_{\tilde{\lambda}_{01}} s) = R_{\lambda_{11}}(s) = R_{\lambda_{00}}(s) = (1-4pqs^2)^{-\frac{1}{2}}$$

You read

$$D_{\lambda_{01}}(s) = \frac{1}{2qs}[(1-4pqs^2)^{-\frac{1}{2}} - 1]$$

Therefore

$$F_{\lambda_0}(s) = \frac{\frac{1}{2qs}[(1-4pqs^2)^{-\frac{1}{2}}]}{(1-4pqs^2)^{-\frac{1}{2}}} = \frac{1}{2qs}[1-(1-4pqs^2)^{\frac{1}{2}}$$

Result 1:

$$f_{\lambda_{01}} = P(T_{\lambda_{01}} < \infty) = F_{\lambda_{01}}(1) = \frac{1}{2q}[1-(1-4pqs^2)^{\frac{1}{2}}] = \frac{1}{2q}[1-|p-q|]$$

Hence $f_{\lambda_{01}} = \begin{cases} 1 & p \geq q \\ p/q < 1 & p < q \end{cases}$

Result 2: $E[T_{\lambda_{01}}] = F_{\lambda_{01}}(1)$ for $P \geq 1$.

Check: $F'_{\lambda_{01}}(s)$

$E[T_{\lambda_{01}}] = \frac{1-|p-q|}{2q|p-q|} = \begin{cases} \frac{1}{p-q} & p > q \\ \infty & p = q \end{cases} = \frac{1}{p-q}$, for $p \geq q$

5.6.1 Summary

1. $p \geq q$ or $p \geq \frac{1}{2}$, $f_{\lambda_{01}} = 1 \& E[T_{\lambda_{01}}] = \frac{1}{p-q}$
2. $p < q$ or $p < \frac{1}{2}$, $f_{\lambda_{01}} = p/q \& E[T_{\lambda_{01}}] = \infty$.

λ_{0k}: visiting k ($k > 0$)given the process starts from 0.

Note: $T_{\lambda_{0k}} = T_{\lambda_{01}} + \cdots + T_{\lambda_{k-1,k}}$ and they are iid.

$T_{\lambda_{01}}$ have the same distribution as $T_{\lambda_{12}}$ since both meaning to right by 1 unit.

Result 1:

$$f_{\lambda_{0k}} = P(T_{\lambda_{ok}} < \infty) = [f_{\lambda_{01}}]^k = \begin{cases} 1 & p \geq q \\ \frac{p}{q}k & p < q \end{cases}$$

Result 2:

$E[T_{\lambda_{0k}}] = E[T_{\lambda_{01}}] + \cdots E[T_{\lambda_{k-1,k}}] = kE[T_{\lambda_{01}}] = \begin{cases} k/(p-q) & p \geq q \\ \infty & p < q \end{cases}$

Q1: How about $k < 0$. change positions of p and q. For example

$$f_{\lambda_{0,-1}} = \begin{cases} 1 & q \geq p \\ \frac{q}{p} & q < p \end{cases}$$

$$E[T_{\lambda_{0-1}}] = \begin{cases} \frac{1}{q-p} & q \geq p \\ \infty & q < p \end{cases}$$

Q2: what is $E[v_{\lambda_{0,k}}]$ $v_{\lambda_{0,k}}$ is the number of times visiting k given process starts from 0 $\neq f_{\lambda_{0,k}}/(1-f_{\lambda_{0,k}})$

Idea condition on $T_{\lambda_{0k}} < \infty$ or not.

$$T_{\lambda_{0,k}} = \infty, v_{\lambda_{0,k}} = 0$$

$$T_{\lambda_{0,k}} < \infty, v_{\lambda_{0k}} = 1 + v_{\lambda_{k,k}}$$

Hence

$$E[v_{\lambda_{0k}}] = P(T_{\lambda_{0,k}} < \infty) * E[1 + v_{\lambda_{kk}}] = f_{\lambda_{0,k}}/|p-q|$$

.

5.7 Gambler's ruin model

Consider a random walk starting from i, $0 < i < k$. Our interest is what is the probability hitting k before hitting 0 starting from i?

Example, gambler have 100 dollars to gamble in a casino. Each game: win $10 or lost $10. What is the probability of doubling your money before getting 0?

Solution: probability of hitting 20 before hitting 0 starting from 10, 1 unit is $10. then let

$$p_i = P(\text{hit k before hitting 0 starting from i}) = \frac{1 + \cdots + (\frac{q}{p})^{i-1}}{1 + \cdots + (\frac{q}{p})^{k-1}} = \begin{cases} \frac{1-(\frac{q}{p})^i}{1-(\frac{q}{p})^k} & \text{if } p \neq q \\ \frac{1}{k} & p = q \end{cases}$$

Conditional on the first movement and difference the equations.

$p_i = P(\text{hit k before 0}|\text{ first is right})P(\text{first is right}) + P(\text{hit k before 0}|\text{ first is left})P(\text{first is left}) = qp_{i-1} + pp_{i+1}$, $0 < i < k$ and $p_0 = 0, p_k = 1$.

Firstly, $p_i = qp_{i-1} + pp_{i+1} \implies p_{i+1} - p_i = (\frac{q}{p})(p_i - p_{i-1})$

Secondly, $p_k - p_0 = [1 + \cdots + \frac{q}{p}^{k-1}](p_1 - p_0)$

Therefore, $1 = [1 + \cdots + \frac{q}{p}^{k-1}]p_1$. Hence $p_1 = \frac{1}{[1+\cdots+\frac{q}{p}^{k-1}]}$

It is easy to find out that

$$p_i = \frac{1 + \cdots + \frac{q}{p}^{i-1}}{1 + \cdots + \frac{q}{p}^{k-1}} = \begin{cases} \frac{1-\frac{q}{p}^i}{1-\frac{q}{p}^k} & p \neq q \\ \frac{i}{k} & p = q \end{cases}$$

6 Discrete Markov Process

6.1 Definitions

Suppose we have a sequence of rvs $\{X_n\}_{n=0}^{\infty}$ (stochastic process).

State space denoted by S and state space is all possible values of $\{X_n\}_{n=0}^{\infty}$.

For example, simple random walk, $S = \{0, \pm1, \pm2, \pm3, \cdots\}$. All integers.

A star is, for $i \in S$, we call it state i. For example, simple random walk, 0 is called state 0. In this lecture we study properties of $\{X_n\}_{n=0}^{\infty}$ from Markov process.

Definition. *Discrete Markov process $\{X_n\}_{n=0}^{\infty}$ is called a Markov process if*

1. *S is countable or discrete*

2. $P(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = P(X_{n+1} = j|X_n = i) = P(X_i = j|X_0 = i)$. *The first equality is called Markov property and the second equality is called time homogeneous.*

Comment:

1. Markov property: X_{n+1} refers to future and X_n refers o current information. $X_0, \cdots, X_{n-1}$ refers to the history. Markov process means that the future only depends on current information X_n but not history $X_0, \cdots, X_{n-1}$.

2. time homogeneous: As long as it is one step. For example, from $X_n \to X_{n+1}$ and $X_n \to X_1$, they are same and does not depend on starting time.

3. One-step transition Matrix: [denoted by **P**] Let $\mathbf{P}_{ij} = P(X_i = j|X_0 = i)$. $\mathbf{P}_{ij}$: one-step transition probability from state i to state j.If we put all probabilities into a matrix, we get $P = (P_{ij})_{i\in S, j\in S}$

6.1.1 Example: Random Walk on the circle

3 positions (0, 1, 2). Each step, process can move clock wise 1 unit with probability p. It can move counter clockwise with 1 unit probability q. X_n is position after n steps. Here:

1. $S = \{0, 1, 2\}$

2. $\{X_n\}_{n=0}^{\infty}$ is a Markov process

   (a) X_{n+1} only depends on X_n

   (b) $P(X_{n+1} = j|X_n = i)$ does not depend on n.

3. One-step transition probability.

$$\mathbf{P} = \begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix}$$

6.1.2 Gambler's Ruin Model

Random walk with 2 absorbing boundaries 0 and k.

1. If the process is in i, $0 < i < k$. it moves to right with p and to the left with q.

2. If the process is in 0 or k, the next step the process will stay in 0 or k.

3. state space $S = \{0, 1, 2, \cdots, k\}$,

4. $\{X_n\}_{n=0}^{\infty}$ is a Markov process.

5. One-step transition matrix: $P_{i,i+1} = p$,$\mathrm{P}_{i,i-1} = q, 0 < i < k$ and $P_{ij} = 0, |j - i| \geq 2$.

6. $P_{00} = 1, P_{0j} = 0,$ for $j \neq 0$.

7. For example, $k = 3$

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Properties of **P**

1. first and important concept in Markov process; just like $\{r_n\}_{n=0}^{\infty}$ for renewal process

2. It is a stochastic Matrix (i.e. $P_{ij} \geq 0, \sum_{j \in S} P_{ij} = 1$ [summation of each row is 1])

6.1.3 Example (Random Walk with reflecting boundary)

1. If the walk is in i, $i > 0$, it moves to right with p and moves to left with q.

2. If it is in 0, the next step the process will be in 1. X_n is the position of process after n steps. State space is $S = \{0, 1, 2, \cdots\}$.

3. $\{X_n\}_{n=0}^{\infty}$ is a Markov process.

4. Transition probability is $P_{i,i+1} = p$ and $P_{i,i-1} = q, i > 0$, $P_{ij} = 0$, $|j - i| > 2$. $P_{0,1} = 1$, $P_{0j} = 0, j \neq 1$.

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ q & 0 & p & 0 & \cdots \\ 0 & q & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

6.2 C-K equation

Two questions:

1. given **P** find $P(X_n = j | X_0 = i)$

2. given **P** and $P(X_0 = i)$, $i \in S$, find $P(X_n = j)$.

- n- step transition matrix

  Let $P_{ij}^{(n)} = P(X_n = j|X_0 = i) = P(X_{n+m} = j|X_m = i)$.

  $P_j^{(n)}$: n-step transition probability from i to j.

  $P^{(n)}$: n-step transition matrix.

Theorem. (C-K equation I)

$$P_{ij}^{(n+m)} = \sum_{k \in S} P_{ik}^{(n)} P_{kj}^{(m)}$$

as matrix form

$$P^{(m+n)} = P^{(n)} P^{(m)}$$

Further:

$$P_{(n)} = P^n$$

Proof.

$$P_{ij}^{(n+m)} = P(X_{n+m} = j|X_0 = i) = \sum_{k \in S} P(X_{n+m} = j|X_n = k, X_0 = i) \times P(X_n = k|X_0 = i)$$

$$P(X_n = k|X_0 = i) = P_{ik}(n)$$

$$P(X_{n+m} = j|X_n = k, X_0 = i) = P(X_{n+m} = j|X_n = k) = P_{kj}^{(m)}$$

so

$$P_{ij}^{(m+n)} = \sum_{i \in S} P_{ik}^{(n)} P_{kj}^{(m)}$$

Intuition: from i to j after $n + m$ steps is the same as from i to k after n steps times from k to j after m steps.

Put the point wise form into a Matrix, then

$$P^{(m+n)} = P^{(n)} P^{(m)}$$

From first step to the nth step.

result: $P^{(n)} = P^n$. Reason $P^{(2)} = P^{(1)} P^{(1)} = P^2$ and so on; therefore, $P^{(n)} = P^n$. □

Suppose $\mathbf{P}$ and $P(X_0 = i), i \in S$ are given what is $P(X_n = j), j \in S$.

Notation: $\pi_i^{(0)} = P(X_0 = i)$ and $\pi^{(0)} = (\pi_i^{(0)})_{i \in S}$.

$\pi_j^{(n)} = P(X_n = j)$, $\pi^{(n)} = (\pi_j^{(n)})_{j \in S}$

Theorem. C-K equation 2

$$\pi_j^{(n)} = \sum_{i \in S} \pi_i^{(0)} P_{ij}^{(n)}$$

$$\text{or } \pi^{(n)} = \pi^{(0)} \times P^n$$

Proof.

$$\pi_j^{(n)} = P(X_n = j) = \sum_{i \in S} P(X_n = j | X_0 = i) \times P(X_0 = i) = \sum_{i \in S} \pi_i^{(0)} P_{ij}^{(n)}$$

Put it in Matrix form:

$$\pi^{(n)} = \pi^{(0)} P^{(n)}$$

□

Example 6.6: Random walk on the circle, $S = \{0, 1, 2\}$.

$$P^{(2)} = \begin{pmatrix} 2pq & q^2 & p^2 \\ p^2 & 2pq & q^2 \\ q^2 & p^2 & 2pq \end{pmatrix}$$

and suppose $P(X_0 = 0) = 1/4, P(X_0 = 1) = 1/2, P(X_0 = 2) = 1/4$.

Find $P(X_2 = n), P(X_2 = 1), P(X_2 = 2)$

Solution:

$$\pi^{(0)} = (1/4, 1/2, 1/4)$$

so

$$\pi^{(2)} = \pi^{(0)} P^{(2)}$$

6.3 Classification of States

Purpose: classify all states in S. Let $\lambda_{ii} =$

- λ_{ii} is a renewal event once the process returns to state i, Markov property tells us process only depends on current state i, not history. This is no-memory property so λ_{ii} is a renewal event.
- Renewal sequence $r_0 = P_{ii}^{(0)} = 1)$, $r_n = P(X_n = i | X_0 = i) = P_{ii}^{(n)}$ for $n \geq 1$.
- Def: classification of states state i is transient if λ_{ii} is transient. State i is null recurrent if λ_{ii} is null recurrent. Otherwise, it is positive recurrent.
- Def: State i has period d if λ_{ii} has period d and state i is called aperiodic if $d = 1$.

6.4 Class in Markov Process

6.4.1 Some Definitions

- Accessible: We say state j is accessible from i. [denote by $i \to j$]

  If $\exists n \geq 0$, such that $P_{ij}^{(n)} > 0$. From i to j has positive probability

- Communication if $i \to j$ and $j \to i$ then we say $i \& j$ can communicate.

  Property:

  1. $i \iff i$ $[P_{ii}^{(0)} = 1]$
  2. If $i \iff j$, then $j \iff i$
  3. If $i \iff j$ and $j \iff k$, then $i \iff k$.

- "Class": communication helps us to divide the state space S into disjoint classes or sets. If two states can communicate, then they are in same class j. If not, then they are in different classes.

  Definition: (irreducible) If we only have one class in sate space, then Markov process is called irreducible.

6.4.2 Irreducible: Only one class

Suppose there is a Markov process with $S = \{0, 1, 2, 3, 4\}$

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 0 & 0 & 1/2 \end{pmatrix}$$

Find Classes.
Solution: $0 \to^{P_{01}>0} 1 \to^{P_{10}>0} \to 0$. Therefore $0 \iff 1$.
$2 \to^{P_{23}>0} 3 \to^{P_{32}>0} 2 \iff 2 \iff 3$
$4 \to 0$ or $\to 1$ or 4.
Therefore, we have 3 classes.

Theorem. Let c be a class

1. All states in C should have same type. all states in same class should be
   (a) positive recurrent at same time; or,
   (b) null recurrent at same time; or,
   (c) transient at same time

2. All states in same class C should have same period.

Theorem. (easy way to find period) Let C be a class. If $\exists i \in C$ such that $P_{ii} > 0$ then all states in C have period 1.

Comprehensive version:

1. All states in C should have same period.

2. state i has period 1.

For stae i; $r_n = P_{ii}^{(n)}$ and $r_1 = P_{ii} > 0$.
Period: $d = gcd\{n | r_n > 0, n \geq 1\} = gcd\{1, \text{ other numbers}\} = 1$.

6.4.3 Concept of States

So far, we do not know how to classify states. Need some concepts: $\begin{cases} \text{open class} \\ \text{closed class} \end{cases}$

Closed Class A class C is called closed. If process cannot leave the class, i.e., for $i \in C, j \notin C, P_{ij} = 0$.

Open Class A class C is called open, if it is possible for process to leave the class, i.e. $\exists i \in C$ and $j \notin c$ such that $P_{ij} > 0$.

Comment: Once the process leaves the open class, it will not return to that open class.

Theorem. 1. All states in open class are transient.

2. For closed class and number of states is finite; then all states must be positive recurrent.

How about closed class with infinite number of states?
Example: $S = \{0, 1, 2, 3, 4\}$.

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 0 & 0 & 1/2 \end{pmatrix}$$

1. Which class is cloud or open? Three classes $C_1 = \{0, 1\}$ is closed. $C_2 = \{2, 3\}$ is closed. $C_3 = \{4\}$ is open (transient).

2. period of each state. All $P_{ii} > 0$ for $i \in \{0, 1, 2, 3, 4\}$. Therefore periods of all states are 1 and all star are aperiodic.

6.5 Stationary distribution

So far, we don't know how to calculate $E[T_{ii}]$.

$$\begin{cases} \pi P = \pi \\ \sum_{i \in S} \pi_i = 1 \end{cases}$$

The meaning of π_i: proportion of process in state i.

6.6 Absorption Probability

Q 4:

Open class $\{1, 2\}$. Let $A_{1,C_1} = P(\text{ending in } C_1 | X_0 = 1)$ and $A_{1,C_2} = P(\text{ending in } C_2 | X_0 = 1)$. $A_{1,C_1} + A_{1,C_2} = 1$.

$$A_{2,C_1} = P(\text{ending in } C_1 | X_0 = 2)$$

$$A_{2,C_2} = P(\text{ending in } C_2 | X_0 = 2)$$

$$A_{2,C_1} + A_{2,C_2} = 1$$

Others can be similarly defined. Idea:Conditional on X_1. start with A_{1,C_1}.

$$A_{1,C_1} = \sum_i P(\text{ending in } C_1 | X_1 = i) \times P(X_1 = i | X_0 = 1) = P_{1,4} \times A_{4,C_1} + P_{1,1} \times A_{1,C_1} + P_{1,2} \times A_{2,C_1}$$

Therefore, $A_{1,C_1} = \frac{1}{3} A_{1,C_1} + \frac{1}{3} A_{2,C_1} \implies A_{2,C_1} \implies A_{1,C_1}$ For A_{2,C_1}, we will then list all the possibility that goes to state 3 (C_1), 6 (C_2), 1, 2.

Conditional on X_1, then

$$A_{2,C_1} = \sum_i P(\text{ending in } C_1 | X_1 = i) \times P(X_1 = i | X_n = 2) = P_{1,3} \times A_{3,C_1} + P_{2,6} \times A_{6,C_1} + P_{2,1} \times A_{1,C_1} + P_{2,2} \times A_{2,4}$$

Therefore, $A_{2,C_1} = \frac{1}{4} + \frac{1}{4} A_{1,C_1} + \frac{1}{4} A_{2,C_1}$ combined with $A_{2,C_1} = 2A_{1,C_1}$. Hence $A_{2,C_1} = 2/5$ and $A_{1,C_1} = \frac{1}{5}$. So $A_{2,C_1} = 2/5$ and $A_{2,C_2} = 1 - 2/5 = 3/5$. $A_{1,C_1} = 1/5$ and $A_{1,C_2} = 4/5$.

Q 5:

$X_0 = 0$: stationary distribution? 0 in C_1 (closed class);: process stays in C_1 and proportion of visiting C_2 and C_3 is zero. For C_1: stationary distribution $\pi_{C_1} = (1/3, 1/3, 1/3)$. For whole process stationary distribution is $(1/3, 0, 0, 1/3, 0, 1/3, 0)$. Similarly: if $X_0 = 4, 6$ in C_2, $\pi_{C_2} = (5/7, 2/7)$. For whole process: $\pi = (0, 0, 0, 0, 5/7, 0, 2/7)$.

Q 6:

$X_0 = 1$ in open class. What is stationary distribution? Recall: if it is in C_1, it is $1/5$; if it is in C_2, it is $4/5$.

Stationary distribution:

$$A_{1,C_1} \times (1/3, 0, 0, 1/3, 0, 1/3, 0) + A_{1,C_2} \times (0, 0, 0, 0, 5/7, 0, 2/7)$$

Q 7:
Let T_1 = waiting time to leave class C_3 given process starts from 1.
T_2 = waiting time for process to leave C_3 given the process starts from 2.
Find $E[T_1]$ and $E[T_2]$.
Idea: condition on X_1. state 1 to 4 (in C_2 leave C_3, 2 (in C_3), 1 (in C_3). Therefore

$$T_1 = \begin{cases} 1 & 1/3(4) \\ 1+T_2 & 1/3(2) \\ 1+T_1 & 1/3(1) \end{cases}$$

Double-expectation theorem implies that $E[T_1] = 1\times\frac{1}{3}+E[1+T_2]\times1/3+E[1+T_1]\times1/3$. Therefore, $E[T_1] = 1+1/3E[T_2]+1/3E[T_1]$. For T_2, state 2 to 3 (leaveC_3), 6 (leave C_3), 1 (in C_3), 2 (in C_3). Values of T_2 is $1, 1, 1+T_1, 1+T_2$ so $E[T_2] = \frac{1}{4}\times1+\frac{1}{4}\times1+\frac{1}{4}E[1+T_1]+\frac{1}{4}E[1+T_2]$. Therefore $E[T_2] = 1+1/4E[T_1]+1/4E[T_2]$ combined with $E[T_1] = 1+1/3E[T_1]+1/3E[T_2]$. We have $E[T_1] = 13/5$ and $E[T_2] = 11/5$.

7 Poisson Process

7.1 Exponential distribution

Exponential distribution: continuous waiting time R.V. Summary: $exp(\lambda)$

1. pdf: $f(x) = \lambda e^{-\lambda}x, x>0, \lambda>0$ where λ is called the rate
2. Tail probability: $P(X>x) = e^{-\lambda x}$ for $x>0$. $X \sim Exp(\lambda)$.
3. $E[X] = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$. $X \sim exp(\lambda)$.
4. No-memory property: $P(X>s+t|X>s) = P(X>t)$. Intuitively, No-matter how long we spent as long as we don't observe the event, the remaining time still follows $exp(\lambda)$.
5. Alarm clock lemma: If $X_i \sim exp(\lambda_i), i = 1, 2, \cdots, \lambda$. and $X_1, \cdots, X_n$ are independent, then
   (a) $\min\{X_1, \cdots, X_n\} \sim Exp(\sum_{i=1}^n X_i)$.
   (b) $P(X_i = \min\{X_1, \cdots, X_n\}) = \frac{\lambda_i}{\sum_{i=1}^n \lambda_k}$.

Example: $T_1 \sim Exp(\lambda_1)$ and $T_2 \sim Exp(\lambda_2)$ are independent. $T = \max(T_1, T_2)$, find $E[T]$. Solution: Method I: $\max(T_1, T_2) + \min(T_1, T_2) = T_1 + T_2$. Well, $\min(T_1, T_2) \sim Exp(\lambda_1, \lambda_2)$. Hence $E[\max(T_1, T_2)] = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1+\lambda_2}$.

Method II: $T = \min(T_1, T_2)$ + remaining time. $E[\max(T_1, T_2)] = \frac{1}{\lambda_1+\lambda_2}$. If $T_1 < T_2$: prob =P($T_1 = \min(T_1, T_2)) = \frac{\lambda_1}{\lambda_1+\lambda_2}$. Also the remaining time of T_2 is $exp(\lambda_2)$. If $T_1 > T_2$, $P(T_2 = \min(T_1, T_2) = \lambda_2\frac{}{\lambda_1+\lambda_2}$. Remaining time of $T_1 \sim Exp(\lambda_1)$. Therefore,

E[Remaining time) $= E[exp(\lambda_2)] \times P(\text{case I}, T_1 < T_2) + E[exp(\lambda_1)] \times P(\text{cases II}, T_1 > T_2) = \frac{1}{\lambda_2} \times \frac{\lambda_1}{\lambda_1+\lambda_2} + \frac{1}{\lambda_1} \times \frac{\lambda_2}{\lambda_1+\lambda_2}$

7.2 Poisson Process

This section:

Interested in $\{X(t), t \geq 0\}$ (continuous). This is called continuous process.
Continuous process: time is continuous.
Discrete process: time is discrete.
For both: state space is discrete.

Definition. *Counting process:* $\{X(t) : t \geq 0\}$ *is called a counting process if*

1. $X(t) \geq 0$.
2. $X(t)$ *can only be non-negative integers.*
3. $X(t)$ *is an increasing function of* t.

For counting process: $X(t) = i$ means we observe i events in $(0, t]$. Therefore, condition 3 makes sense [more time $\implies$ more events]. Plot of $\{t, X(t)\}$ for counting process. (only increase but not decrease)

Definition. *Poisson process: a counting process is called a poisson process if*

$X(0) = 0$*;*

For $0 < S_1 < S_2 \leq t_1 < t_2$, $X(t_2) - X(t_1)$ *and* $X(S_2) - X(S_1)$ *are independent.*

$X(t+s) - X(s) \sim Pois\lambda t)$ *where the first part is the number of events in* $(S, t+S]$ *and* λ *is the unit rate and* t *is the length of interval.*

Summary of 3 conditions

1. starts from 0
2. in two non-overlapped intervals, the number of events are independent.
3. The number of events in an interval is poisson distributed with mean = unit rate $\times$ length of interval.

7.2.1 Property 1

In a very small interval, we can only observe 0 or 1 event.

Mathematically, $\lim_{h \to 0^+} \frac{P(X(t+h) - P(X(t))}{h} = 0$.

Proof. $X(t+h) - X(t) \sim Pois(\lambda h)$ Therefore,

$$\lim_{h \to 0^+} \frac{P(X(t+h) - X(t) \geq 2)}{h} = \lim_{h \to 0^+} \frac{1 - \frac{(\lambda h)^0 e^{-\lambda h}}{0!} - \frac{(\lambda h)^1 e^{-\lambda h}}{1!}}{h}$$

Apply L'Hôpital's Rule,

$$\lim_{h \to 0^+} \frac{P(X(t+h) - X(t) \geq 0)}{h} = \lim_{h \to 0^+} \lambda^2 h e^{\lambda h} = 0$$

□

Property 1 tells us we cannot jump from 0 to 2.

7.2.2 Property 2

$$T_1 = \text{waiting time for the first event}$$

$$T_2 = \text{waiting time for the second event after the first event}$$

In general,

$$T_i = \text{waiting time for ith event after ith event}$$

Conclusion: $T_1, T_2, \cdots$ are iid and follow $exp(\lambda)$ where λ is the unit rate of poisson process.

Proof. Only for T_1, $P(T_1 > t) = P(X(t) = 0)$ (tail probability of T_1 is the probability of number of events in $(0, t]$. Therefore, it is just $\frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$ so $P(T_1 > t) = e^{-\lambda t}$ tail probability of $exp(\lambda)$. Hence $T_1 \sim exp(\lambda)$. □

7.2.3 Property 3

Suppose $S < t$ given $X(t) = n$. $X(S)|X(t) = n \sim BIN(n, \frac{S}{t})$.

Proof. $P(X(S) = k|X(t) = n) = \frac{P(X(S)=1, X(t)=n)}{P(X(t)=n)} = \frac{P(X(t)-X(S)=n-k)}{P(X(t)=n)}$

Note $X(S)$ and $X(t) - X(S)$ are independent, $X(S) \sim Pois(\lambda) \& X(t)$ and $X(t) - X(S) \sim Pois(\lambda(t-s))$ so

$$P(X(S) = k|X(t) = n) = \frac{(\lambda S)^k e^{-\lambda S}/k! \times \frac{[\lambda(t-S)]^{n-k} e^{-\lambda(t-S)}}{(n-k)!}}{\frac{(\lambda t)^n e^{-\lambda t}}{n!}}$$

Simplify it, we get $P(X(S) = k|X(t) = n) \sim Bin(n, S/t)$ □

7.2.4 Example

Customer arrive at CIBC according to a Poisson process at the rate 90 per hour. If in the first half hour, we see 15 customers

1. What is the probability that 5 customers arrive in next half hour?

$$X(t) = \text{number of customers in (0, t] unit is per hour}$$

$$X(t) \sim \text{Poisson process unit rate } \lambda = 20/hr$$

$$X(1/2) = 15$$

$$P(X(1){-}X(1/2) = 5|X(1/2) = 15) = P(X(1){-}X(1/2) = 5) \sim Poisson(10) = \frac{10^5 e^{-10}}{5!}$$

2. Of 15 people who arrived in the first half hour, what is the probability that 10 arrive in the first 10 minutes?

$$P(X(1/6) = 10|X(1/2) = 15) = \binom{15}{10}(\frac{1}{3})^{10}(\frac{2}{3})^5$$

7.2.5 Property 4

$$X(t) = \text{number of events in (0, t]\&it follows poisson process with unit rate } \lambda$$

The events can be classified as type I event (prob= p) or type II event (prob= q)

$$X_1(t) = \text{number of type I events in (0, t]}$$

$$X_2(t) = \text{number of type II events in (0, t]}$$

$$X(t) = X_1(t) + X_2(t)$$

Conclusion

1. $X_1(t)$ is a poisson process with unit rate λp
2. $X_2(t)$ is a poisson process with unit rate λq.
3. $X_1(t)$ and $X_2(t)$ are independent.

How to proceed?

1. $X(t) \sim Pois(\lambda t)$
2. $X_1(t)|X(t) = n \sim Bin(n, p)$

8 Midterm Coverage

Chapter 1
1. commonly used distributions: binomial, geometric, negative binomial, poisson. $\star$ when to use them?, pmf, expectation and variance will be given.
2. property of expectation and variance
3. indicator variable

Chapter 2 classification of RVs. Definition: improper, null improper, short proper.

Chapter 3
1. Joint RVs: joint pdf, joint pmf and independence
2. Conditional expectation, distribution
3. property of conditional expectation (linearity, substitution rule, independence)
4. double expectation theorem, variance
5. calculate probability, variance

Chapter 4 emphasize pgf

1. Def of pgf
2. 5 properties of pgf. Note: 4 power series and two properties of gf.

Chapter 5 Delayed renewal is not covered

1. definition of renewal, delayed renewal
2. Rule to determine renewal or delayed renewal
3. Classification of renewal
4. expectation and probability
5. Renewal sequence