Probability Notes

April 19, 2020

Contents

1 Probability Theorems

1.1 Set Theorems

For any three sets, the following hold true

$$
\begin{aligned}
A &= (A \cap B) \cup (A \cap B^c) \; where \; B \; and \; B^c \; are \; disjoint \\
A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\
A \cup (B \cap C) &= (A \cup B) \cap (A \cup C)
\end{aligned}
$$

1.2 Basic Probability Rules

$$
\begin{aligned}
&\text{If } A \cap B = \phi, \text{ then } P(A \cup B) = P(A) + P(B) \\
&P(A|B)P(B) = P(B|A)P(A) = P(A \cap B) && \text{Bayes' Theorem} \\
&P(A) = P(A \cap B) + P(A \cap B^c) = P(A|B)P(B) + P(A|B^c)P(B^c) \\
&P(A \cap B \cap C) = P(A)P(B|A)P(C|B, A) && \text{Chain Rule}
\end{aligned}
$$

1.2.1 Total Probability Theorem

Let A_1, A_2, .. ,A_n be n disjoint events that completely cover the event space, and B be another event, then

$$
\begin{aligned}
P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n) \\
\text{or, } P(B) &= \sum_{i=1}^{n} P(B|A_i)P(A_i)
\end{aligned}
$$

1.3 Independence

Two events A and B are independent iff

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

Note that *independence* is not the same as *disjoint*

$$A \cap B = \phi \Rightarrow P(A \cap B) = 0 \text{ but } P(A) \neq P(B) \neq 0$$

Multiple events $A_1, A_2, \ldots, A_n$ are independent iff

$$P(A_i \cap A_j \cap \ldots \cap A_k) = P(A_i)P(A_j) \; .. \; P(A_k) \;\; \forall \, i, j, \ldots, k \mid i, j, \ldots, k \in 1, 2, \ldots, n$$

Conditional Independence is similar to the above equation. For an event C,

$$P(A_i \cap A_j \cap \ldots \cap A_k|C) = P(A_i|C)P(A_j|C) \; .. \; P(A_k|C) \;\; \forall \, i, j, \ldots, k \mid i, j, \ldots, k \in 1, 2, \ldots, n$$

1.4 Joint Probability Distributions

Joint Probability Distributions are defined for two or more than two variables. In this section, we only consider two variables. The formal definition is

$$P_{XY}(x, y) = P(X = x \; and \; Y = y)$$

Based on this definition, the following theorems follow

$$\sum_x \sum_y P_{XY}(x,y) = 1$$

$$P_X(x) = \sum_y P_{XY}(x,y) \qquad \text{Marginal Probability}$$

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

$$\sum_x P_{X|Y}(x|y) = 1 \qquad \text{Since Y is fixed and we sum over all X's}$$

$$P_{XYZ}(x,y,z) = P_X(x)P_{Y|X}(y|x)P_{Z|X,Y}(z|x,y) \qquad \text{Chain Rule}$$

1.5 Expected Value

Before going to expected value, let's define a Random Variable

Random Variable X is a linear map : $\mathbb{R} \to \mathbb{R}$. The value taken by the variable is denoted by x

X will have an associated probability distribution, i.e., $P_X(X=x)$. Using these quantities, we have

$$E[X] = \sum_x x P_X(X=x) \qquad \text{Expected Value}$$

Based on this definition, the following theorems for expected value follow

$$\begin{aligned}
E[\alpha] &= \alpha E[\alpha X] && = \alpha E[X] \\
E[\alpha X + \beta] &= \alpha E[X] + \beta \\
E[g(X)] &= \sum_x g(x) P_X(X=x) \\
E[X^2] &= \sum_x x^2 P_X(X=x) && \text{Also called } \textit{Second Moment} \\
E[X|A] &= \sum_x x P_{X|A}(X|A) \\
E[g(X)|A] &= \sum_x g(x) P_{X|A}(X|A) \\
E[X+Y+Z] &= E[X] + E[Y] + E[Z] && \text{Linearity of Expectation} \\
E[XY] &= \sum_X \sum_Y xy P_{XY}(x,y) \\
E[g(X,Y)] &= \sum_X \sum_Y g(xy) P_{XY}(x,y) \\
E[XY] &= E[X]E[Y] && \text{if X and Y are } \textit{independent}
\end{aligned}$$

where $\alpha, \beta \in \mathbb{R}$, $g(X) : \mathbb{R} \to \mathbb{R}$, and A is an event, X, Y, Z are Random Variables

1.5.1 Total Expectation Theorem

The *Total Expectation Theorem* is the natural extension of the *Total Probability Theorem* Let $A_1, A_2, \ldots, A_n$ be n disjoint events that completely cover the event space, and X be random variable, then

$$E[X] = E[X|A_1]P(A_1) + E[X|A_2]P(A_2) + \cdots + E[X|A_n]P(A_n)$$

$$\text{or, } E[X)] = \sum_{i=1}^{n} E[X|A_i]P(A_i)$$

1.6 Variance

The formal definition of variance is

$$Var(X) = E[(X - \bar{X})^2] = E[X^2] - E[X]^2$$

Using this definition, the following theorems follow

$$\begin{aligned}
E[X^2] &= E[X]^2 + Var(X) \\
Var(\alpha) &= 0 \\
Var(\alpha X + \beta) &= \alpha^2 Var(X) \\
Var(X + Y) &= Var(X) + Var(Y) \text{ if } X \text{ and } Y \text{ are } \textit{independent} \text{ random variables}
\end{aligned}$$

1.7 Cumulative Probability Distribution

Cumulative probability distribution is defined for both discrete and continuous variables

$$F_x(X) = P(X \leq x) = \begin{cases} \int_{-\inf}^{x} p_X(t)dt & X \text{ is a discrete random variable} \\ \sum_{k<=x} P_X(k) & X \text{ is a continuous random variable} \end{cases}$$

2 Binomial Random Variable

Binomial Random Variable X is defined as the number of successes in an experiment with n independent trials, where each trial can only have two outcomes, *success* or *failure*.
Let X_i denote the Random Variable corresponding to the individual trials, with probability of success p. Then we have the following

$$\begin{aligned}
X_i &= \begin{cases} 1 & \text{if success in trial i} \\ 0 & \text{otherwise} \end{cases} && \text{indicator variable} \\
X &= X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i \\
P(X = k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}$$

2.1 Mean and Variance

First let's calculate the mean and variance for a single trial X_i

$$\begin{aligned}
E[X_i] &= 1 * p + 0 * (1-p) && = p \\
Var(X_i) &= (1-p)^2 p + (0-p)^2 (1-p) &&= p(1-p)
\end{aligned}$$

We know that all $X_i's$ are independent. Hence, the mean and variance for X become

$$\begin{aligned}
E[X] &= E[\sum_{i=1}^{n} X_i] &&= \sum_{i=1}^{n} E[X_i] &&= np \\
Var(X) &= Var(\sum_{i=1}^{n} X_i) &&= \sum_{i=1}^{n} Var(X_i) &&= np(1-p)
\end{aligned}$$

3 Continuous Uniform Random Variable

A uniform random variable is defined as follows

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

3.1 Mean and Variance

$$\begin{aligned} E[X] &= \int_a^b x\frac{1}{b-a}dx = [\frac{x^2}{2(b-a)}]_a^b \\ &= \frac{a+b}{2} \\ Var(X) &= \int_a^b (x - \frac{a+b}{2})^2 \frac{1}{b-a}dx \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

4 Gaussian Distribution

The gaussian distribution (or normal distribution) is defined between $-\inf$ and $\inf$. It is parametrized by mean μ and variance σ, $X \sim \mathcal{N}(\mu, \sigma^2)$

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

As already described,

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

A *Standard Normal* is defined as a normal distribution with $\mu = 0$ and $\sigma^2 = 1$
Any normal distribution can be converted to a standard normal as $X = \frac{X-\mu}{\sigma}$
If $Y = aX + b$, then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

5 Covariance and Correlation

For any two random variables X and Y,

$$\begin{aligned} Cov(X,Y) &= E[(X - \overline{X})(Y - \overline{Y})] = E[XY] - E[X]E[Y] \\ Cov(X,X) &= Var(X) \\ Corr(X,Y) &= E[(\frac{X - \overline{X}}{\sigma_X})(\frac{Y - \overline{Y}}{\sigma_Y})] \\ &= \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \end{aligned}$$

Key points to note

- *Inpdependence* $\Rightarrow Cov(X,Y) = Corr(X,Y) = 0$, but the converse is ***not*** true
- Correlation is dimensionless and $-1 \leq Corr(X,Y) \leq 1$ with value close to 0 implying minimal relation and values close to $-1, 1$ implying perfect relation

6 Iterated Expectation and Variance

The law of iterated expectation tells the following about expectation and variance

$$\begin{aligned} E[E[X|Y]] &= E[X] \\ Var(X) &= E[Var(X|Y)] + Var(E[X|Y]) \end{aligned}$$

Proof for Iterated Expectation

$$\begin{aligned} P(X) &= \sum_y P(X|Y)P(Y) \\ \Rightarrow E[X] &= \sum_x xP(X) = \sum_x \sum yxP(X|Y)P(Y) \\ &= \sum_y P(Y) \sum_x xP(X|Y) = \sum_y P(Y)E[X|Y] \\ \text{or, } E[X] &= E[E[X|Y]] && E[X|Y] \text{ is a function of } X \text{ and not } Y \end{aligned}$$

Proof for Variance

$$\begin{aligned} Var(X) &= E[X^2] - E[X]^2 \\ Var(X|Y) &= E[(X - \overline{X})^2|Y] = E[X^2|Y] - E[X|Y]^2 && 1 \\ Var[E(X|Y)] &= E[E(X|Y)^2] - E[E[X|Y]]^2 \\ &= E[E[(X|Y)]^2] - E[X]^2 && 2 \\ E[Var(X|Y)] &= E[E[X^2|Y]] - E[E[X|Y]^2] && \text{from } 1 \\ &= E[X^2] - E[E[X|Y]^2] && 3 \\ E[Var(X|Y)] + Var(E[X|Y]) &= E[X^2] - E[X]^2 && \text{adding 2 and 3} \\ &= Var(X) \end{aligned}$$

7 Random number of Random Variables

Let X_i be independent identically distributed Random Variables and let $Y = \sum_{i=1}^{N} X_i$ be the sum of N such random variables where N itself is a random variable. Then,

$$\begin{aligned} Y &= X_1 + X_2 + \cdots + X_N \\ E[Y|N=n] &= \sum_{i=1}^{n} E[X_i] \\ &= NE[X] \\ E[Y] &= E[E[Y|N]] = E[NE[X]] \\ &= E[N]E[X] && \text{since } E[X] \text{ will be a number} \\ Var(Y) &= E[Var(Y|N)] + Var(E[Y|N]) \\ &= E[NVar(X)] + Var(NE[X]) \\ &= E[N]Var(X) + E[X]^2 Var(N) \end{aligned}$$

8 Bernoulli Process

Bernoulli process falls under the family of random processes, which are random variables continuously evolving over time. Bernoulli process can be described as a sequence of independent

Bernoulli trials, where each trial has only two outcomes : success with $P(success) = p$ and failure.

$$P_{X_t}(x_t) = \begin{cases} p & \text{if } X_t = 1 \\ 1-p & \text{if } X_t = 0 \end{cases}$$
$$E[X_t] = p$$
$$Var(X_t) = p(1-p)$$

8.1 Mean and Variance

Number of successes S in n time slots

$$P(S = k) = \binom{n}{k} p^k (1-p)^(n-k)$$
$$E[S] = np$$
$$Var(S) = np(1-p)$$

8.2 Interarrival Times

Let T_1 denote the number of trials till the first success

$$P(T_1 = t) \qquad = (1-p)^{t-1}p \qquad t \in 1, 2, \ldots$$
$$E[T_1] = \frac{1}{p}$$
$$Var(T_1) = \frac{1-p}{p^2}$$

This process is memoryless as all future coin flips are independent of whatever has happened till now. Also, the distribution is a *Geometric Random Variable.*

8.3 Sum of Interarrival times

We are interested in the total time till k arrivals. Let this random variable be Y_k

$$\begin{aligned} Y_k &= T_1 + T_2 + \cdots + T_k && \text{where } T_i\text{'s are i.i.d geometric with parameter } p \\ P(Y_k = t) &= P(k-1 \text{ arrivals between } t = 1 \text{ to } t = t \text{ and last arrival at time } t) \\ &= \binom{t-1}{k-1} p^k (1-p)^{t-k} && \forall\, t \geq k \\ E[Y_k] &= \sum_{i=1} kE[T_i] \\ &= \frac{k}{p} \\ Var(Y_k) &= \sum_{i=1}^{k} Var(T_i) \\ &= \frac{k(1-p)}{p^2} \end{aligned}$$

9 Poisson Process

Poisson process also falls in the realm of random processes but is different from Bernoulli process as it is a continuous time process. This process is very commonly used to model arrival times and number of arrivals in a given time interval.

$$\begin{aligned}
P(k,\tau) &= \text{Probability of } k \text{ arrivals in interval of duration } \tau \\
\sum_k P(k,\tau) &= 1 && \text{for a given } \tau
\end{aligned}$$

Assumptions

- The Probability is dependent only on τ and not the *location* of the interval
- Number of arrivals in disjoint time intervals are *independent*

9.1 Derivation from Bernoulli Process

For a very small interval δ,

$$\begin{aligned}
P(k,\delta) &= \begin{cases} 1-\lambda\delta & k=0 \\ \lambda\delta & k=1+O(\delta^2) \\ 0 & k>2 \end{cases} \\
\lambda &= \lim_{\delta\to 0}\frac{P(1,\delta)}{\delta} && \text{arrival rate per unit time} \\
E[k] &= (\lambda\delta)*1+(1-\lambda\delta)*0 \\
&= \lambda\delta \\
\tau &= n\delta
\end{aligned}$$

The last equation clearly implies that we can approximate the whole process as a bernoulli process where we have n miniscule time intervals with at most one arrival per interval.

$$\begin{aligned}
P(k\ arrivals) &= \binom{n}{k}p^k(1-p)^{n-k} \\
&= \binom{n}{k}(\frac{\lambda\delta}{n})^k(1-\frac{\lambda\delta}{n})^{n-k} \\
\lambda\tau &= np && \text{or, arrival rate * time = E[arrivals]} \\
Poisson &= \lim_{\delta\to 0,n\to\inf} Bernoulli \\
or,\ P(k,\tau) &= \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!} && k=0,1,\cdots,\text{ for a given } \tau \\
where,\ \sum_k P(k,\tau) &= 1 && \text{for a given } \tau
\end{aligned}$$

Let N_t denote the no of arrivals till time t, then

$$\begin{aligned}
E[N_t] &= \lambda t \\
Var(N_t) &= \lambda t
\end{aligned}$$

9.2 Time till k^{th} arrival

Suppose the k^{th} arrival happens at a time t. Then we are saying that there have been $k-1$ arrivals till time t and the k^{th} arrival happens at time t (precisely in an interval of $[t, t+\delta]$).

Let Y_k be the required time,

$$\begin{aligned}
f_{Y_k}(t)\delta &= P(t \leq Y_k \leq t+\delta) \\
&= P(k-1 \text{ arrivals in } [0,t])(\lambda\delta) \\
&= \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t}(\lambda\delta) \\
f_{Y_k}(t) &= \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t} \qquad \text{Erland Distribution}
\end{aligned}$$

9.3 Time of 1^{st} arrival

Using the Erlang Distribution described above, we have

$$f_{Y_1}(t) = \lambda e^{-\lambda t}$$

$Y_k = T_1 + T_2 + \cdots + T_k$ where all T_i are independent and exponential distributions.

9.4 Merging of Poisson Processes

Merging of two Poisson processes is also a Poisson process. Consider two flasbulbs of Red and Green colours, flashing as Possion processes with rates λ_1 and λ_2. Then the process denoting the combined flashing of the two bulbs is also Poisson.
Consider a very small interval of time δ. In this small interval, any of the individual bulbs can have at most one flashes (since we ignore higher order terms). Thus, the following four possibilities arise

0	Red	$\overline{Red}$
$Green$	$\lambda_1\delta\lambda_2\delta$	$(1-\lambda_1\delta)\lambda_2\delta$
$\overline{Green}$	$\lambda_1\delta(1-\lambda_2\delta)$	$(1-\lambda_1\delta)(1-\lambda_2\delta)$

Table 1: Base Probabilities for flashes

0	Red	$\overline{Red}$
$Green$	0	$\lambda_2\delta$
$\overline{Green}$	$\lambda_1\delta$	$(1-(\lambda_1+\lambda_2)\delta)$

Table 2: Probabilities after ignoring δ^2 terms

Thus, the combined process is Poisson with parameter $\lambda_1 + \lambda_2$

$$P(\text{arrival happened from first process}) = \frac{\lambda_1\delta}{\lambda_1\delta + \lambda_2\delta} = \frac{\lambda_1}{\lambda_1+\lambda_2}$$

9.5 Splitting of Poisson Process

Suppose we have a Poisson process with parameter λ which we split into two processes up and down, with probabilities p and $1-p$. The two resulting processes are also Poisson with different parameters.
Consider a small time slot of length δ. Then,

$$\begin{aligned}
P(\text{arrival in this time slot}) &= \lambda\delta \\
P(\text{arrival in up slot}) &= \lambda\delta p \\
P(\text{arrival in down slot}) &= \lambda\delta(1-p)
\end{aligned}$$

Thus, up and down are themselves Poisson with parameters λp and $\lambda(1-p)$ respectively.

9.6 Random Indcidence for Poisson

Suppose we have a Poisson process with parameter λ running forever. We show up at a random time instant. What is the length of the chosen interarrival time (the total of the time from the last arrival to the next arrival).
Let $T_1^{'}$ denote the time that has elapsed since the last arrival and T_1 be the time till the next arrival. Note that the reverse process is also Poisson with the same parameter. Thus,

$$E[\text{interarrival time}] = E[T_1^{'} + T_1] = \frac{1}{\lambda} + \frac{1}{\lambda} = \frac{2}{\lambda}$$

This may seem paradoxical since the time difference between any two arrivals in a Poisson process is same and it's expected length is $\frac{1}{\lambda}$, whereas we got an interval twice this length. The paradox is resolved by considering the fact that when we choose a random point in time, it is more likely to fall in an interval of larger size than the smaller ones (since probability will be proportional to the length of the interval).

Consider a separate example where we want to compare two values $E[\text{size of a family}]$ and $E[\text{size of a family of a given person}]$.
The two value will be different due to the underlying nature of the way experiment is conducted. For the first, we randomly choose families and average their sizes. Here, family of any size is equally likely to be picked. In the second case, we first pick a person from the population, get their family size, and then average the sizes of the families. Note that, this experiment is biased since the we are more likely to select people from larger families (or equivalently, it is more likely that we pick a person from a large family since the probability of picking is proportional to the family size). Hence, the second value will likely be larger and the two quantities are not equal.

10 Markov Process

Markov Process is a discrete time process that is not memoryless. Here the random variable takes several possible states, and the probability distribution is defined in such a way that $P(\text{transition from state 1 to state 2})$ is dependent on state 1.

Let X_n be the random variable denoting the state after n transitions and X_0 will represent the starting state (which can be given or random). Markov assumption states that *Given the current state, past does not matter.* Armed with these,

$$\begin{aligned}
p_{ij} &= P(\text{next state } j \mid \text{current state } i) \\
p_{ij} &= P(X_{n+1} = j | X_n = i) = P(X_{n+1} = j | X_n = i, X_{n-1}, \ldots, X_0) \\
r_{ij}(n) &= P(X_n = j | X_0 = i) && \text{or, in state } j \text{ after } n \text{ steps} \\
r_{ij}(n) &= \sum_{k=1}^{m} r_{ik}(n-1) p_{kj}
\end{aligned}$$

10.1 Recurring and Transient States

A state i is called *recurrent* if, starting from i, and travelling anywhere, it is always possible to return to i. If a state is not recurrent, it is *transient*. States in a recurrent class are periodic if they can be grouped into $d > 1$ groups so that all transitions from one group lead to the next group.

10.2 Steady State Probabilities

Do $r_{ij}(n)$ converge to some π_j (independent of i) ?
Yes if,

- recurrent states are all in a single class
- single recurrent class is not periodic (otherwise oscillations are possible)

Assuming yes,

$$r_{ij}(n) = \sum_k r_{ik}(n-1)p_{kj}$$

$$\lim_{n\to\inf} r_{ij}(n) = \sum_k r_{ik}(n-1)p_{kj}$$

$$\pi_{ij} = \sum_k \pi_{ik}p_{kj} \qquad \text{balance equations}$$

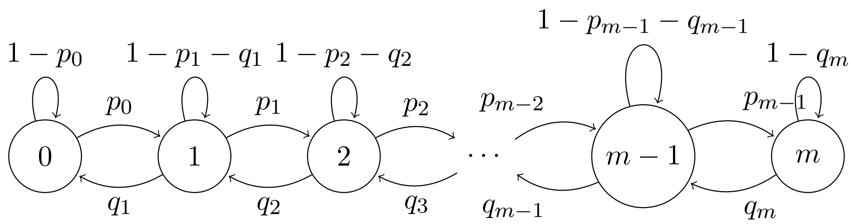
$$\sum_i \pi_i = 1$$

$$\text{frequency of transitions } k \to j = \pi_k p_{kj} \qquad \text{in one step}$$

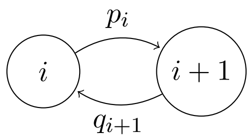
$$\text{frequency of transitions into } j = \sum_k \pi_k p_{kj} e \qquad \text{influx from all connected states}$$

10.3 Birth Death Process

Consider the checkout counter example. The states are represented by the number of people currently being processed, and we always move n to $[n-1, n, n+1]$, i.e., either the people in the queue decrease by one, remain same or increase by one. Let the probability for moving up be p and moving down be q.



Let's estimate the steady state probabilities. Consider the following diagram splitting the chain into two parts through the two adjacent states



In this case, to maintain steady state, long term frequency of left-right transition should be same as right left transition, i.e., $\pi_i p_i = \pi_{i+1} q_i$

In the special case of $p_i = p$ and $q_i = q \; \forall \; i$,

$$\rho = \frac{p}{q} \qquad \text{load factor}$$
$$\pi_{i+1} = \pi_i \frac{p}{q} = \pi_i \rho$$
$$\pi_i = \pi_0 \rho^i \qquad i = 0, \ldots, m$$
$$\text{Using } \sum_{i=0}^{m} \pi_0 \rho^i = 1,$$
$$\pi_0 = \frac{1}{\sum_{i=0}^{m} \rho^i}$$

if $p < q$ and $m \to \inf$,

$$\pi_0 = 1 - \rho$$
$$\pi_i = (1 - \rho)\rho^i$$
$$E[X_n] = \frac{\rho}{1 - \rho} \qquad \text{Exponential Distribution}$$

When $\rho = 1$ or $p = q$, then all states are equally likely - symmetric random walk.

10.4 Absorption Probabilities

let a_i denote the probability of absorption and μ_i denote the expected no of steps until absorption starting from state i. Then,

$$a_i = \sum_j a_j p_{ij} \qquad \text{outflux to the possible states}$$
$$\mu_i = 1 + \sum_j \mu_j p_{ij}$$

For multipe absorption states, we can possibly consider them together as a group and calculate the relevant quantities.
For a given state s,

$$E[\text{steps to first time reach } s \text{ from } i] = t_i$$
$$t_i = E[min\{n \geq 0 \text{ such that } X_n = s\}]$$
$$t_s = 0$$
$$t_i = 1 + \sum_j t_j p_{ij} \qquad \text{outflux to all possible states}$$

Mean recurrence time (mean time to reach back a state) for s

$$t_s^* = E[min\{n \geq 1 \text{ such that } X_n = s\} | X_0 = s]$$
$$t_s^* = 1 + \sum_j t_j p_{ij}$$

10.5 Weak Law of Large Numbers

Suppose we want to know the mean height of penguins in the world. The absolutely correct answer can be obtained by taking the average of the entire population. But ths is not practical, and often we will have to resort to estimating the quantity through a sample. Let there be

n penguins in the sample and $X_1, X_2, \ldots, X_n$ be the random variables denoting their heights. Then,

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$
$$\lim_{n \to \inf} E[M_n] = E[X] = \text{The true mean}$$

10.6 Markov Inequality/Chebychev Inequality

For nonnegative random variable X,

$$E[X] = \sum_x x p_X(x) \geq \sum_{x \geq a} x p_X(x) \qquad \text{discrete case}$$
$$= \int_x x p_X(x) \ \geq \int_{x \geq a} x p_X(x) \qquad \text{continuous case}$$

Applying the above set of inequalities to the variable $X - \mu$

$$E[(X-\mu)^2] \geq a^2 P((X-\mu)^2 \geq a^2)$$
$$\text{or,} \ \ Var(X) \geq a^2 P(\mid X - \mu \mid \geq a)$$

For continuous case,

$$\sigma^2 = \int_{-\inf}^{\inf} (x-\mu)^2 f_X(x) dx$$
$$\geq \int_{-\inf}^{\mu - c} (x-\mu)^2 f_X(x) dx + \int_{\mu + c}^{\inf} (x-\mu)^2 f_X(x) dx$$
$$\geq c^2 P(\mid X - \mu \mid \geq c)$$

Hence,

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

or,

$$\boxed{P(\mid X - \mu \mid \geq k\sigma) \leq \frac{1}{k^2}} \quad \text{where } c = k\sigma$$

Going back to the problem of estimating the mean,

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$
$$E[M_n] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \mu \quad \text{expectation of expectation}$$
$$Var(M_n) = \sum_{i=1}^{n} Var(\frac{X_i}{n}) = \frac{\sigma^2}{n} \quad \text{since } X_i \text{ are independent}$$
$$\boxed{P(\mid M_n - \mu \mid \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}}$$

or, as $n \to \inf$, $M_n - \mu \to 0$, ϵ is the error bound/confidence.

11 Central Limit Theorem

Chebychev's inequality gives a loose bound. We can do better with CLT. Let X be a random variable with mean μ and variance σ^2, and let X_i be independent identically distributed random variables with the same distribution as X. Then,

$$
\begin{aligned}
S_n &= X_1 + X_2 + \cdots + X_n \\
Z_n &= \frac{S_n - E[S_n]}{\sigma_n} \quad \text{random variable with mean 0 and variance 1} \\
&= \frac{S_n - nE[X]}{\sqrt{n}\sigma} \\
\text{or, } S_n &= \sqrt{n}\sigma Z_n + nE[X] \\
\text{In } \lim_{n\to\inf} Z_n &\to Z \text{ (standard normal)} \\
\text{or, } &\boxed{Z = \frac{S_n - nE[X]}{\sqrt{n}\sigma}} \quad \text{only for CDF (no comment on PDF/PMF)} \\
\text{Thus, } &\boxed{P(Z > c) = P(\frac{S_n - nE[X]}{\sqrt{n}\sigma} > c)}
\end{aligned}
$$

By defining the confidence on how close we desire S_n to the actual mean, we can calculate the required value of the n using standard normal CDF tables. However, we need to have an estimate of variance of the distribution in order to do the estimate of n.

12 Problems

12.1 Cumulative Distribution Function

A random variable X is a combination of a continuous and discrete distribution as follows

$$
f_X(x) = \begin{cases} 0.5 & a \le x \le b \\ 0.5 & \text{x} = 0.5 \\ 0 & \text{otherwise} \end{cases}
$$

Find the Cumulative Distribution of X.

Ans. Cumulative Distribution of X can be found by integration and is as follows

$$
f_X(x) = \begin{cases} 0 & x < 0 \\ 0.5x & 0 \le x < 0.5 \\ 0.75 & \text{x} = 0.5 \\ 0.75 + 0.5(x - 0.5) & 0.5 < x \le 1 \\ 1 & 1 < x \end{cases}
$$

12.2 Number of tosses till first head

When tossing a fair coin, what is the $E[\# \; tosses \; till \; the \; first \; H]$?

Ans. Let X be the # of tosses till first H. Then, $(X = 1) \cap (X > 1) = \phi$.
Using *Total Expectation Theorem*

$$
\begin{aligned}
E[X] &= P(X = 1)E[X|X = 1] + P(X > 1)E[X|X > 1] \\
&= 0.5 * 1 + 0.5E[X] \\
\Rightarrow E[X] &= 2
\end{aligned}
$$

$P(X = 1) = 0.5$ because then we get the head in the first toss itself. Since $P(X = 1) + P(X > 1) = 1$, we have $P(X > 1) = 0.5$. $E[X] = E[X|X > 1]$ because the tosses are *independent* and thus memoryless.

12.3 Iterated Expectation practice

A class has two sections denoted by the random variable Y. Let X denote the quiz score of a student. Given that section 1 has 10 students, section 2 has 20 students, $E[X|Y = 1] = 90E[X|Y = 2] = 60, Var(X|Y = 1) = 10, Var(X|Y = 2) = 20$, find $E[X]$ and $Var(X)$.

Ans. We use the formulae from iterated expectation to calculate these.

$$\begin{aligned}
P_Y(y) &= \begin{cases} \frac{1}{3} & y = 1 \\ \frac{2}{3} & y = 2 \end{cases} \\
E[X] &= E[E[X|Y]] = \sum_y E[X|Y]P(Y) \\
&= 90 * \frac{1}{3} + 60 * \frac{2}{3} \\
Var(X) &= E[Var(X|Y)] + Var(E[X|Y]) \\
&= \sum_y Var(X|Y)P(Y) + ((90 - E[E[X|Y])^2\frac{1}{3} + (60 - E[E[X|Y]])^2\frac{2}{3}) \\
&= \frac{650}{3}
\end{aligned}$$

12.4 Hat Problem

n people throw their hats in a box and then pick a hat at random. What is the expected number of people who pick their own hat ?

Ans. Let X denote the number of people who pick their own hat. We have been asked $E[X]$.
Let X_i be a binary random variable denoting whether the i^{th} person picked their own hat, i.e.,

$$\begin{aligned}
X_i &= \begin{cases} 1 & \text{if } i^{th} \text{ person picks their own hat} \\ 0 & \text{otherwise} \end{cases} \\
P(X_i = 1) &= \frac{1}{n} \\
E[X_i] &= 1 * \frac{1}{n} + 0 * (1 - \frac{1}{n}) = \frac{1}{n}
\end{aligned}$$

Consequently

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

It is interesting to see the variance of X. Note that the formula for variance is $E[X^2] - E[X]^2$.

Thus,

$$X^2 = (\sum_{i=1}^{n} X_i)^2 = \sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} X_i X_j$$
$$E[X^2] = \sum_{i=1}^{n} E[X_i^2] + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} E[X_i X_j]$$

Note that X_i and X_j are not independent since after the first person has picked the hat, only $n-1$ hats remain

$$X_i X_j = \begin{cases} 1 & \text{if } X_i = X_j = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned}
P(X_i X_j = 1) = P(X_i = 1)P(X_j = 1|X_i = 1) &= \frac{1}{n} * \frac{1}{n-1} \\
E[X_i X_j] = 1 * (\frac{1}{n} * \frac{1}{n-1}) + 0 * (1 - \frac{1}{n} * \frac{1}{n-1}) &= \frac{1}{n(n-1)} \\
E[X_i^2] = 1^2 \frac{1}{n} + 0^2 (1 - \frac{1}{n}) &= \frac{1}{n}
\end{aligned}$$

Putting these values in the original equation for variance

$$\begin{aligned}
E[X_2] = n\frac{1}{n} + \frac{1}{n}\frac{1}{n-1}(\frac{n(n-1)}{2} * 2) &= 2 \\
Var(X) = 2 - 1^2 &= 1
\end{aligned}$$

12.5 Breaking a stick

A stick of length l is broken first at X uniformly chosen between $[0, l]$, and then at Y, uniformly chosen between $[0, X]$. Find the expected length of the shorter part.

Ans. The following is the joint probability distribution of X and Y

$$f_{XY}(x, y) = f_X(x) f_{Y|X}(y|x) = \frac{1}{l}\frac{1}{x} = \frac{1}{xl} \; \forall \; 0 \leq y \leq x \leq 1$$

Using marginal probabilities, we can calculate $f_Y(y) and E[Y] as$

$$f_Y(y) = \int f_{XY}(x, y)dx = \int_y^l \frac{1}{xl}dx = \frac{1}{l}\log\frac{l}{y} \qquad \text{Note that for any } y,\; y \leq x \leq l$$
$$E[Y] = \int y f_Y(y) = \int_0 ly\frac{1}{l}\log\frac{l}{y} = \frac{l}{4}$$

This problem can also be approched using iterated expectation

$$\begin{aligned}
E[Y] &= E[E[Y|X]] = E[\text{uniform random variable between 0 and } x] \\
&= E[\frac{X}{2}] = \frac{1}{2}E[X] \\
&= \frac{l}{4}
\end{aligned}$$

12.6 PMF of g(X)

Let X be uniform in $[0, 2]$, then find the PMF of $Y = X^3$.

Ans. Always solve such questions using the cumulative distribution approach.

$$P(X \leq x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2}x & 0 \leq x \leq 2 \\ 1 & 2 < x \end{cases}$$

$$P(Y \leq y) = P(X^3 \leq y) = P(X \leq y^{\frac{1}{3}})$$

$$= \begin{cases} 0 & y < 0 \\ \frac{1}{2}y^{\frac{1}{3}} & 0 \leq y^{\frac{1}{3}} \leq 2 \\ 1 & 2 < y^{\frac{1}{3}} \end{cases}$$

$$f_Y(y) = \frac{dP(Y <= y)}{dy}(y)$$

$$= \begin{cases} 0 & y < 0 \\ \frac{1}{6}y^{\frac{-2}{3}} & 0 \leq y \leq 8 \\ 0 & 8 < y \end{cases}$$

12.7 Poisson Emails

You get emails according to a Poisson process at the rate of 5 messages/hour. You check email every 30 minutes. Find

- P(no new message)
- P(one new message)

Ans. We can model the arrival process like a Poisson process. $\lambda = 5$ and $\tau = \frac{1}{2}$

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

$$P(5, \frac{1}{2}, 0) = \frac{(5 * \frac{1}{2})^0 e^{-5*\frac{1}{2}}}{0!}$$

$$P(5, \frac{1}{2}, 1) = \frac{(5 * \frac{1}{2})^1 e^{-5*\frac{1}{2}}}{1!}$$

12.8 Poisson Fishing

We go fishing where we catch fishes at the rate of $0.6/hour$. We fish for two hours. If we do not catch a fish in the first two hours, we fist until the first catch. Find the following

- P(fish for > 2 hours)
- P(fish for > 2 but < 5 hours)
- P(catch at least two fish)
- E[fish]
- E[Total fishing time]

- E[future fishing time—fished for two hours]

Ans.

- P(fish for > 2 hours) = $P(k = 0, \tau = 2) = e^{-0.6*2}$
- P(fish for > 2 but < 5 hours) = P(first catch in $[2, 5]$ hours) = $P(k = 0, \tau = 2)(1 - P(k = 0, \tau = 3)$ which is no fish in $[0, 2]$ but at least 1 fish in the next 3 hours (which will be independent of first 2 hours)
- P(catch at least two fish) = P(at least 2 catches before 2 hours) = $1 - P(k = 0, \tau = 2) - P(k = 1, \tau = 2)$
- E[fish] has two possibilities, either single fish after 2 hours, or many fist before 2 hours. $E[fish] = E[fish|\tau \leq 2](1 - P(\tau > 2)) + E[fish|\tau > 2]P(\tau > 2) = (0.6 * 2) * (1 - P(k = 0, \tau = 2)) + 1 * P(k = 0, \tau = 2)$
- E[Total fishing time] = $2 + P(k = 0, \tau = 2)\frac{1}{\lambda}$, since we fish for atlest 2 hours
- E[future fishing time—fished for two hours] can be obtained using the memoryless property of Poisson process. The expected time till first arrival is independent of what has happened till now. Thus, $E[T_1] = \frac{1}{\lambda}$

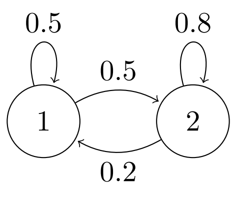
12.9 Poisson Lightbulbs

We have three identical but independent lightbulbs whose lifetimes are modelled by a Poisson process with parameter λ. Given that we start all the three bulbs together, find the E[time until last bulb dies out]

Ans. Start with the merged Poisson process which will denote the time till the first bulb will fail. For this process, $\lambda^{'} = 3\lambda$. Hence, E[first bulb fails] = $\frac{1}{3\lambda}$.
After the first bulb dies out, we are left with a process with $\lambda^{'} = 3\lambda$. Due to memoryless property, E[second bulb fails] = $\frac{1}{2\lambda}$ and consequently E[last bulb fails] = $\frac{1}{\lambda}$.
Note the above two times denote the time difference, i.e. the time taken for the bulb to die out after the last bulb died out. Thus, E[time until last bulb dies out] = $\frac{1}{3\lambda} + \frac{1}{2\lambda} + \frac{1}{\lambda}$

12.10 Steady State Markov Process

Find the steady state probabilites of the following Markov Process

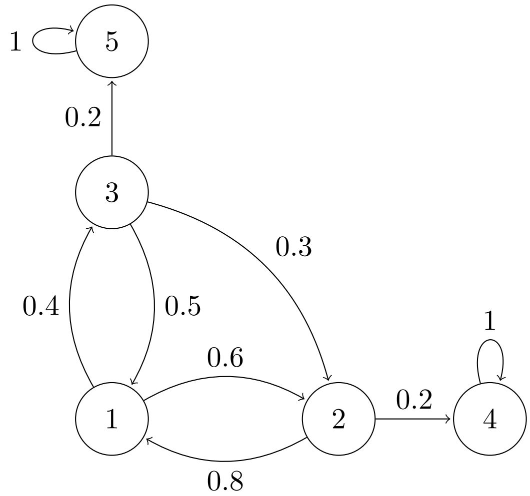


Ans. Using balance equations, we have

$$\pi_1 = \pi_1 p_{11} + \pi_2 p_{21}$$
$$\pi_2 = \pi_1 p_{12} + \pi_2 p_{22}$$
$$\pi_1 + \pi_2 = 1$$

Solving, $\pi_1 = \frac{2}{7}$ and $\pi_2 = \frac{5}{7}$

12.11 Absorption Probabilities



Calculate the absorption probabilites for state 4 and expected time to absortion from all states. (for absorption time, assume $p_{35} = 0$ and $p_{32} = 0.5$

Ans. Let a_i denote the abosorption probabilites into state 4 starting from i

$$
\begin{aligned}
a_5 &= 0, a4 = 1 \\
a_i &= \sum_j a_j p_{ij} \\
a_2 &= a_1 p_{21} + a_4 p_{24} \\
a_3 &= a_1 p_{31} + a_2 p_{32} + a_5 p_{35} \\
a_1 &= a_2 p_{12} + a_3 p_{13}
\end{aligned}
$$

Solving, $a_1 = \frac{9}{14}, a_2 = \frac{5}{7}$ and $a_3 = \frac{15}{28}$

Let μ_i denote the expected time till absorption starting from i, then

$$
\begin{aligned}
\mu_4 &= 0 \\
\mu_1 &= 1 + \mu_2 p_{12} + \mu_3 p_{13} \\
\mu_2 &= 1 + \mu_1 p_{21} + \mu_4 p_{24} \\
\mu_3 &= 1 + \mu_1 p_{31} + \mu_2 p_{32}
\end{aligned}
$$

Solving, $\mu_1 = \frac{55}{4}, \mu_2 = 12$ and $\mu_3 = \frac{111}{8}$