Probability Notes

May 25, 2020

Contents

1 Probability Theorems

1.1 Set Theorems

For any three sets, the following hold true

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

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

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

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

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

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

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

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 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

3 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

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

Proof for Variance

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

4 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,

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

5 Convolutions

Convolution operations are defined for both CDF and PDF/PMFs. Let X and Y be random independent variables, then

$$F_{X+Y}(x) = F_X * F_Y = \int_{\mathbb{R}} F_X(x-y)dF_Y(y)$$
$$p_{X+Y}(x) = p_X * p_Y = \int_{\mathbb{R}} p_X(x-y)p_Y(y)dy$$

We can extend the idea to n independent variables as

$$F_X^{n*} = F_X * \cdots * F_X \text{ } n \text{ times}$$

It has the following properties for positive random variable X_is

1. $$F_X^{n*}(x) \le F_X^n(x)$$

This can be proven as

$$\begin{aligned} P(X_1 + \cdots + X_n \leq x) &\leq P(X_1 \leq x, \ldots, X_n \leq x) \\ P(X_1 + \cdots + X_n \leq x) &\leq \prod_{i=1}^{n} P(X \leq x) \text{ by independence} \\ \text{or, } F_X^{n*}(x) &\leq F_X^n(x) \end{aligned}$$

2.

$$F_X^{n*}(x) \geq F_X^{n+1}(x)$$

which follows immediately from the fact that

$$P(X_1 + \cdots + X_n \leq x) \geq P(X_1 \leq x, \ldots, X_{n+1} \leq x)$$

since the volume of the regions denoting the sums will be lower in the higer dimensions. This can be quickly verified by considering $X_1 \leq 1$ and $X_1 + X_2 <= 1$.

6 Moment Generating Function

Moment generating function is defined as the following for all values of t

$$\phi(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p_X(x) & \text{for discrete case} \\ \int_{-\inf}^{\inf} e^{tx} f_X(x) & \text{for continuous case} \end{cases}$$

This function is called the moment generating function because all the moments of the random variable X can be obtained by successively differentiating the function $\phi(t)$.

$$\begin{aligned} \phi'(t) &= \frac{d}{dt} E[e^{tX}] \\ &= E[\frac{d}{dt} e^{tX}] \\ &= E[Xe^{tX}] \\ \text{mean} &= E[X] \\ &= \phi'(0) \end{aligned}$$

Continuing in a similar fashion,

$$\begin{aligned} \phi''(t) &= \frac{d}{dt} E[Xe^{tX}] \\ &= E[\frac{d}{dt} Xe^{tX}] \\ &= E[X^2 e^{tX}] \\ \text{variance} &= \phi''(0) \\ &= E[X^2] \end{aligned}$$

In general, for any $n > 0$, the n^{th} derivative will give the n^{th} moment

$$\phi^n(0) = E[X^n]$$

There exists a one to one correspondence between the moment generating function and the distribution function of a random variable, similar to Lagrangian multipliers.

6.1 Moment Generating Function for Sum of Independent RV

An important property is in the context of sum of two or more random variables. The **moment generating of sum of two independent random variables is simply the product of the moment generating functions of the two individual random variables**

$$\begin{aligned}
\phi_{X+Y}(t) &= E[e^{t(X+Y)}] \\
&= E[e^{tX}e^{tY}] \\
&= E[e^{tX}]E[e^{tY}] \\
\boxed{\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)} &\quad \text{for independent random variables}
\end{aligned}$$

7 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

$$X_i = \begin{cases} 1 & \text{if success in trial i} \\ 0 & \text{otherwise} \end{cases} \qquad \text{indicator variable}$$

$$X = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

$$P(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

7.1 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

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

8 Continuous Uniform Random Variable

A uniform random variable is defined as follows

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

8.1 Mean and Variance

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

9 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 $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

10 Counting Process

Counting process is used in scenarios when we want to count the occurrence of a certain event. N_t denotes the number of events till time t starting from 0. It is assumed that $N_0 = 0$. Formal definition is

A random process $\{N_t, t \in [0, \inf)\}$ is said to be a counting process if N_t is the number of events from time $t = 0$ upto time t. For a counting process, we assume

1. $N_0 = 0$
2. $N_t \in \{0, 1, 2, \cdots\}$ for all $t \in [0, \inf)$
3. for $0 \leq s < t, N_t - N_s$ shows the number of events that occur in the interval $(s, t]$

10.1 Independent Increments

We say that a continuous time counting process N_t has independent increments if for all $0 \leq t_1 < t_2 < \cdots < t_n$, the random variables

$$N_{t_2} - N_{t_1},\ N_{t_3} - N_{t_2}, \ldots,\ N_{t_n} - N_{t_{n-1}}$$

are independent.

Note that these differences are nothing but the number of arrivals in a given time interval. Thus, we are equivalently saying that **the number of arrivals in any two disjoint intervals are**

independent.

A very simple consequence of this property is:
Suppose we wise to find the probability of 2 arrivals in the interval $(1,2]$ and 3 arrivals in the interval $(3,5]$. Then,

$$P(2 \text{ arrivals in } (1,2] \text{ and } 3 \text{ arrivals in } (3,5]) = P(2 \text{ arrivals in } (1,2])P(3 \text{ arrivals in } (3,5]))$$

since the arrivals in disjoint intervals are independent.

10.2 Stationary Increments

We say that a continuous time counting process N_t has stationary increments if for all $t_2 > t_1 \geq 0$ and for all $r > 0$, $N_{t_2} - N_{t_1}$ and N_{t_2+r} and N_{t_1+r} are independent.

In other words, **the number of arrivals in a given time interval is invariant to it's location**. Note that the number of arrivals in the time interval between t_1 and t_2 is nothing but $N_{t_2} - N_{t_1}$. By the above statement, if the process has stationary increments, then this quantity is same as $N_{t_2-t_1}$, which is the distribution of the counting process itself.

11 Renewal Process

This is a fundamental stochastic process useful in modelling arrivals and interarrival times. Some definitions will make the usage clear.

Let S_i denote the ith renewal time or the time when the ith arrival takes place. By definition, $S_0 = 0$. We can also define

$$S_n = S_{n-1} + \xi_n$$
$$S_n = \xi_1 + \xi_2 + \cdots + \xi_{n-1}$$

where ξ_i are positive $(P(\xi > 0) = 1)$ independent identically distributed variables representing the interarrival times. We also define

$$N_t = \underset{k}{\operatorname{argmax}}\{S_k \leq t\}$$
$$\{S_n > t\} = \{N_t < n\}$$

or, N_t is simply the number of arrivals till the time t.

Define the following quantity

$$F^{n*} = F_\xi * \ldots * F_\xi \; n \text{ times}$$
$$u(t) = \sum_{i=1}^{\inf} F^{n*}(t)$$

It can be shown that the function $u(t)$ converges. The expectation of N_t then becomes

$$
\begin{aligned}
E[N_t] &= E[\text{number of } n \text{ such that } S_n \leq t] \\
&= E[\sum_{n=1}^{\inf} I(S_n \leq t)] && \text{sum of Indicators will equal } n \\
&= \sum_{n=1}^{\inf} P(S_n \leq t) && \text{since } E[\text{Indicator}] \text{ is just the function inside indicator} \\
&= \sum_{n=1}^{\inf} F^{n*}(t) && \text{by defining cumulative as sum of } \xi\text{s} \\
&= u(t)
\end{aligned}
$$

11.1 Laplace Transform

For a density function f defined from $\mathbb{R}^{\geq 0} \to \mathbb{R}$, Laplace transform is

$$L_f(s) = \int_{\mathbb{R}^{\geq 0}} e^{-sx} f(x) dx$$

The following properties hold for this transform

1. If f is a probability density function, then
$$E[e^{-sx}] = L_f(s)$$
2. if f_1 and f_2 are two probability density functions, then
$$L_{f_1 * f_2}(s) = L_{f_1}(s) L_{f_2}(s)$$
3. If F is the cumulative probability distribution for X and p is the probability density function, then
$$L_{F_X}(s) = \frac{L_{p_X}(s)}{s}$$
which can be proven using integration by parts as follows
$$L_{F_X}(s) = \int_{\mathbb{R}^{\geq 0}} F_X(x) \frac{d(e(-sx))}{s} = 0 + \frac{1}{s} \int_{\mathbb{R}^{\geq 0}} p_X(x) e^{-sx} dx$$

11.2 Calculating the Expectation

Armed with the concept of a Laplace transform, we make the following observation first

$$
\begin{aligned}
u(t) &= \sum_{i=1}^{\inf} F^{n*}(t) = F(t) + \sum_{i=2}^{\inf} F^{n*}(t) \\
&= F(t) + \big(\sum_{i=1}^{\inf} F^{n*}(t) \big) * F(t) \\
&= F(t) + u(t) * F(t) \\
u(t) &= F(t) + u(t) * p(t)
\end{aligned}
$$

where p is the probability density function and the last line stems from the fact that $\int u * F = \int u(x-y) dF(y) = \int u(x-y) p(y) dy$. Taking Laplace transform on both sides,

$$
\begin{aligned}
L_u(s) &= L_F(s) + L_u(s)L_p(s) \\
L_u(s) &= \frac{L_p(s)}{s} + L_u(s)L_p(s) \text{ from } 3 \\
L_{u(s)} &= \frac{L_p(s)}{s(1 - L_p(s))}
\end{aligned}
$$

The last equation can be used to calculate the laplace transform of $u(t)$ and consecutively guess the functional form of $u(t)$.

11.3 Limit Theorems for Renewal Processes

The following two theorems hold true for Renewal processes

1. If $E[\xi] = \mu < \inf$, then

$$\lim_{t \to \inf} \frac{N_t}{t} = \frac{1}{\mu}$$

which is analogous to the strong law of large numbers. This can be proven as follows

$$
\begin{aligned}
S_{N_t} \leq t \leq S_{N_t+1} &\text{ from the definition of } N_t \\
\text{or, } \frac{N_t}{S_{N_t+1}} &\leq \frac{N_t}{t} \leq \frac{N_t}{S_{N_t}}
\end{aligned}
$$

we can calculate the limits on the two bounds as

$$\lim_{t \to \inf} \frac{N_t}{S_{N_t}} = \lim_{n \to \inf} \frac{n}{S_n} = \frac{1}{\mu}$$

from the strong law of large numbers applied to $\lim_{n \to \inf} \frac{S_n}{n}$. Similarly, one can show

$$\lim_{t \to \inf} \frac{N_t}{S_{N_t+1}} = \lim_{t \to \inf} \frac{N_t}{N_t + 1} \lim_{t \to \inf} \frac{N_t + 1}{S_{N_t+1}} = 1 * \frac{1}{\mu}$$

2. If $Var(\xi) = \sigma^2 < \inf$, then

$$\lim_{t \to \inf} \frac{N_t - t/\mu}{\sigma\sqrt{t}/\mu^{3/2}} = \mathcal{N}(0, 1)$$

which is analogous to the central limit theorem. It can be proven by considering the CLT on ξs

$$
\begin{aligned}
&\lim_{n \to \inf} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \text{CDF of } \mathcal{N}(0, 1) \\
\text{or, } &\lim_{n \to \inf} P(S_n \leq n\mu + \sigma\sqrt{n}x) = \text{CDF of } \mathcal{N}(0, 1) \\
\text{or, } &\lim_{n \to \inf} P(N_t \geq n) = \text{CDF of } \mathcal{N}(0, 1) \text{ from definition of } N_t, \text{ where } t = n\mu + \sigma\sqrt{n}x
\end{aligned}
$$

We substitute $n\mu = t$ for very large value of n, since the total time will become total variables into the expected time for one ξ when n is large. Hence,

$$n = \frac{t}{\mu} - \frac{\sigma\sqrt{t}}{\mu^{3/2}}x$$

$$\lim_{n \to \inf} P(N_t \geq n) = \lim_{n \to \inf} P\left(\frac{N_t - t/\mu}{\sigma\sqrt{t}/\mu^{3/2}} \leq x\right) = \text{CDF of } \mathcal{N}(0, 1)$$

12 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.

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

12.1 Mean and Variance

Number of successes S in n time slots

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

12.2 Interarrival Times (Geometric Random Variable)

Let T_1 denote the number of trials till the first success

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

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**.

12.3 Sum of Interarrival times

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

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

13 Exponential Distribution

Exponential distribution is characterized by the parameter λ and has the following probability distribution

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

Exponential distribution is used to represent the interarrival time probability distribution in the context of Poisson Process. The cumulative distribution is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{otherwise} \end{cases}$$

$$\begin{aligned} P(X > x) &= \int_x^{\text{inf}} \lambda e^{-\lambda x} dx \\ &= e^{-\lambda x} \end{aligned}$$

13.1 Mean and Variance

The mean of the distribution is given by

$$\begin{aligned} E[x] &= \int_0^{\text{inf}} \lambda x e^{-\lambda x} dx \\ &= [-xe^{-\lambda x}]_0^{\text{inf}} + \int_0^{\text{inf}} e^{-\lambda x} dx = \frac{1}{\lambda} \\ &\boxed{E[X] = \frac{1}{\lambda}} \end{aligned}$$

where we used integration by parts, $\int uv' = uv - \int u'v$ and substituted $u = x$ and $v = -e^{-\lambda x}/\lambda$.

For variance, we first calculate the value of $E[x^2]$

$$\begin{aligned} E[x^2] &= \int_0^{\text{inf}} \lambda x^2 e^{-\lambda x} dx \\ &= [-x^2 e^{-\lambda x}]_0^{\text{inf}} + \int_0^{\text{inf}} 2xe^{-\lambda x} dx \\ &= [\frac{-2xe^{-\lambda x}}{\lambda}]_0^{\text{inf}} - [\frac{2e^{-\lambda x}}{\lambda^2}]_0^{\text{inf}} \\ &= \frac{2}{\lambda^2} \\ Var(X) &= E[X^2] - E[X]^2 \\ &\boxed{Var(X) = \frac{1}{\lambda^2}} \end{aligned}$$

The above property can be generalized for the nth power as well

$$E[X^n] = \frac{n!}{\lambda^n}$$

13.2 Memoryless Property

A fundamental mathematical property of the exponential distribution is the memoryless property. In summary, this means that whatever has transpired till now will not affect the future distribution. Mathematically $P(T > t + s) is independent of t$

$$\begin{aligned} P(T > t+s|T > t) &= \frac{P(T > t+s \text{ and } T > t)}{P(T > t)} \\ &= \frac{P(T > t+s)}{P(T > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \end{aligned}$$

$$\boxed{P(T > t+s|T > t) = P(T > s)}$$

14 Poisson Process

14.1 Poisson Random Variable

A random variable X is said to be $Poisson(\mu)$ if it has the following probability distribution

$$p_X(x = k) = \begin{cases} e^{-\mu}\frac{\mu^k}{k!} & \text{for all } x = \{0,\ 1,\ 2, \cdots\} \\ 0 & \text{otherwise} \end{cases}$$

The sum of n independent Poisson variables is also Poisson

$$X_1 + X_2 + \cdots + X_n \sim Poisson(\mu_1 + \mu_2 + \cdots + \mu_n)$$

14.2 Mean and Variance

Expected value is calculated as follows

$$\begin{aligned} E[X] &= \sum_{k=0}^{\inf} k e^{-\mu}\frac{\mu^k}{k!} = \mu e^{-\mu}\sum_{k=1}^{\inf}\frac{\mu^{k-1}}{(k-1)!} \\ &= \mu e^{-\mu}\sum_{k=0}^{\inf}\frac{\mu^k}{k!} \end{aligned}$$

$$\boxed{E[X] = \mu}$$

Variance can be calculated using $Var(X) = E[X^2] - E[X]^2$

$$\begin{aligned} E[X^2] &= \sum_{k=0}^{\inf} k^2 e^{-\mu}\frac{\mu^k}{k!} = \mu e^{-\mu}\sum_{k=1}^{\inf} k\frac{\mu^{k-1}}{(k-1)!} \\ &= \mu e^{-\mu}\sum_{k=0}^{\inf}(k+1)\frac{\mu^k}{k!} \\ &= \mu e^{-\mu}(\mu\sum_{k=1}^{\inf}\frac{\mu^{k-1}}{(k-1)!} + \sum_{k=0}^{\inf}\frac{\mu^k}{k!}) \\ &= \mu e^{-\mu}(\mu e^{\mu} + e^{\mu}) \\ Var(X) &= E[X^2] - E[X]^2 \end{aligned}$$

$$\boxed{Var(X) = \mu}$$

Thus, mean and variance is the same for a Poisson variable.

14.3 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.

$$P(k,\tau) = \text{Probability of } k \text{ arrivals in interval of duration } \tau$$

$$\sum_k P(k,\tau) = 1 \qquad \text{for a given } \tau$$

Assumptions

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

14.4 A Special Counting Process

A counting process $N_t : t \in [0, \inf)$ is a Poisson process with rate λ if

1. $N_0 = 0$
2. N_t is composed of independent and stationary increments
3. The number of arrivals in any time interval $\tau > 0$ has $Possion(\lambda\tau)$ distribution

Hence, for a Poisson process, the number of arrivals in any interval is dependent only on the length of that interval and not the location. Further, the number of arrivals in the interval will follow a Poisson distribution.

14.5 Derivation from Bernoulli Process

For a very small interval δ,

$$P(k,\delta) = \begin{cases} 1-\lambda\delta & k=0 \\ \lambda\delta & k=1+O(\delta^2) \\ 0 & k>2 \end{cases}$$

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

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

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

14.6 Time till kth 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,

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

14.7 Time of 1st 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.

14.8 Renewal Process

Poisson process can be seen as a special case of a renewal process, when the interarrival times are all exponentially distributed.

$$\begin{aligned} \text{Interarrival time } \xi_i \quad &= \lambda e^{-\lambda t} \\ \text{Number of arrivals } P(N_t = n) &= \frac{(\lambda t)^n}{n!}e^{-\lambda t} \\ \text{Time till } n\text{th arrival } P(S_n = t) \quad &= \lambda\frac{(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda x} \text{ for } t > 0 \\ \text{Cumulative distribution } P(S_n \le t) \quad &= \begin{cases} 1 - e^{-\lambda t}\sum_{k=1}^{n-1}\frac{(\lambda t)^k}{k!} & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

14.9 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

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

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

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

14.10 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,

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

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

14.11 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.

14.12 Non Homogenous Poisson Process

Sometimes, it may not be accurate to use a simple Poisson process to model arrival. For example, a restaurant will not have the same rate of influx throughout the day. This rate itself is a function of time. In such cases, we model the arrivals as Non Homogenous Poisson Process.

For such a process, we have $\lambda(t) : [0, \inf) \to [0, \inf)$ and the counting process N_t is non homogenous if the following hold

1. $N_0 = 0$
2. The increments to N_t are **independent but not stationary**
3. For any small time interval δ, the probability of more than 1 arrival in the interval is zero

The distribution of arrivals in a time interval is still Poisson, but the Poisson parameter is now dependent on the location of the interval itself (since the process does not have stationary increments)

$$N_{t+s} - N_t \sim Poisson(\int_t^{t+s} \lambda(\alpha)d\alpha)$$

15 Gamma Distribution

A random variable is said to have a Gamma distribution if for parameters $\lambda > 0, \alpha > 0$ it has the following probability distribution

$$p_X(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The denominator in the above fomula acts as nothing but a normalization constant and is defined as

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^{\inf} \lambda e^{-\lambda x}(\lambda x)^{\alpha-1} \\
&= \int_0^{\inf} e^{-y} y^{\alpha-1} dy && \text{by letting } \lambda x = y \\
&= (\alpha - 1) \int_0^{\inf} e^{-y} y^{\alpha-2} dy \text{ using integration by parts} \\
&= (\alpha - 1)\Gamma(\alpha - 1)
\end{aligned}$$

Note that at $\alpha = 1$, $\Gamma(1) = \int_0^{\inf} \lambda e^{-\lambda x} = 1$. Hence, if α is an integer, $\Gamma(\alpha) = \alpha!$ using the recursion relation derived above.

For a fixed λ, as the value of α becomes large, the distribution takes the form of a normal distribution.

15.1 Mean and Variance

Mean and variance are easily obtainable for this using the moment generating function. Recall

$$\begin{aligned}
\phi(t) &= E[e^{tX}] \\
\phi^n(t) &= E[X^n]
\end{aligned}$$

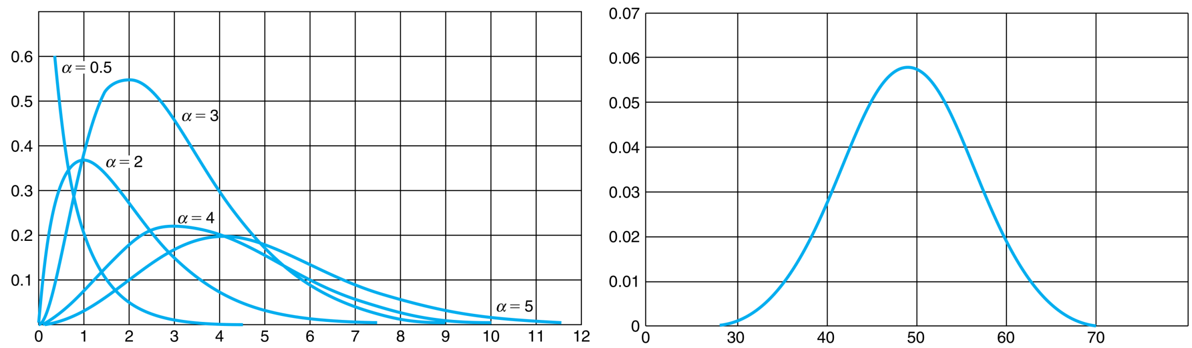


Figure 1: Gamma distribution for $\lambda = 1$ and different values of α. The bottom figure shows the distribution for $\alpha = 50$.

For the current distribution,

$$\begin{aligned}\phi(t) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\inf} e^{tx} e^{-\lambda x} x^{\alpha-1} \\ &= (\frac{\lambda}{\lambda - t})^\alpha\end{aligned}$$

Differentiating,

$$\begin{aligned}\phi'(t) &= \frac{\alpha\lambda^\alpha}{(\lambda - t)^{\alpha+1}} \\ \phi''(t) &= \frac{\alpha(\alpha+1)\lambda^\alpha}{(\lambda - t)^{\alpha+2}} E[X] && = \phi'(0) \\ &\boxed{E[X] = \frac{\alpha}{\lambda}} \\ Var(X) &= \phi''(0) \\ &\boxed{Var(X) = \frac{\alpha}{\lambda^2}}\end{aligned}$$

15.2 Sum of Gamma Distributions

Let $X_1, X_2, \dots, X_n$ be n random variables that are gamma distributed with parameters $(\alpha_1, \lambda), (\alpha_2, \lambda), \dots, (\alpha_n, \lambda)$. Then the distribution of the sum of these random variables is itself a gamma distribution with the parameters $\alpha' = \sum_{i=1}^{n} \alpha_i$ and $\lambda' = \lambda$

16 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(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,

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

16.1 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.

16.2 Steady State Probabilities

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

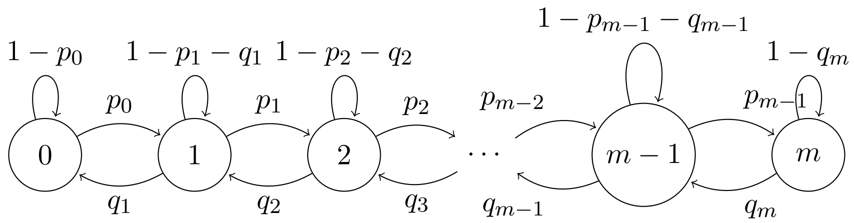
- recurrent states are all in a single class
- single recurrent class is not periodic (otherwise oscillations are possible)

Assuming yes,

$$
\begin{aligned}
r_{ij}(n) &= \sum_k r_{ik}(n-1) p_{kj} \\
\lim_{n \to \inf} r_{ij}(n) &= \sum_k r_{ik}(n-1) p_{kj} \\
\pi_{ij} &= \sum_k \pi_{ik} p_{kj} && \text{balance equations} \\
\sum_i \pi_i &= 1 \\
\text{frequency of transitions } k \to j &= \pi_k p_{kj} && \text{in one step} \\
\text{frequency of transitions into } j &= \sum_k \pi_k p_{kj} && \text{influx from all connected states}
\end{aligned}
$$

16.3 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

$$\begin{tikzcd} i \arrow[r, bend left, "p_i"] & i+1 \arrow[l, bend left, "q_{i+1}"] \end{tikzcd}$$

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 $q_i = q \; \forall \; i$,

$$\begin{aligned}
\rho &= \frac{p}{q} && \text{load factor} \\
\pi_{i+1} &= \pi_i \frac{p}{q} = \pi_i \rho \\
\pi_i &= \pi_0 \rho^i && i = 0, \dots, m \\
\text{Using } \sum_{i=0}^{m} \pi_0 \rho^i &= 1, \\
\pi_0 &= \frac{1}{\sum_{i=0}^{m} \rho^i} \\
\text{if } p < q \text{ and } m \to \inf, \\
\pi_0 &= 1 - \rho \\
\pi_i &= (1-\rho)\rho^i \\
E[X_n] &= \frac{\rho}{1-\rho} && \text{Exponential Distribution}
\end{aligned}$$

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

16.4 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,

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

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

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

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

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

17 Central Limit Theorem

17.1 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,

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$\lim_{n \to \inf} E[M_n] = E[X] = \text{The true mean}$$

17.2 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.

17.3 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,

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

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.

18 Linear Regression

We are given pairs of data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ (all independent) where we assume that x and y are governed by the linear relation

$$y \approx \theta_0 + \theta_1 x$$

The aim is to determine the model which is parametric consisting of two parameters θ_0 and θ_1. We find it using the least squares estimate, i.e., minimizing

$$\underset{\theta_0,\theta_1}{\text{minimize}} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x)^2$$

The true model also includes noise and is given by

$$Y_i = \theta_0 + \theta_1 X_i + W_i$$

where we assume the noise $W_i \sim \mathcal{N}(0, \sigma^2)$ and is independently and identically distributed. Observing some X and Y is same as observing the noise.

$$P(X = x, Y = y) = P(W = y - \theta_0 - \theta_1 x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{(y - \theta_0 - \theta_1 x)^2}{2})$$

$$\begin{aligned}
P(X_1 = x_1, Y_1 = y_1, \ldots, X_n = x_n, Y_n = y_n) &= \prod_{i=1}^{n} P(X_1 = x_i, Y_i = y_i) \\
&= \prod_{i=1}^{n} W_i = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp(-\frac{(y_i - \theta_0 - \theta_1 x_i)^2}{2})
\end{aligned}$$

Maximizing the above product is maximizing the likelihood of the occurrence of the data under the model parameters θ_0 and θ_1. Since taking log will not change the maxima, we usually maximize the log likelihood

$$\underset{\theta_0,\theta_1}{\text{maximize}} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp(-\frac{(y_i - \theta_0 - \theta_1 x_i)^2}{2}) = \underset{\theta_0,\theta_1}{\text{minimize}} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2$$

We can take derivatives with respect to the parameters of the above function to get the estimate for the parameters as

$$\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i,\ \bar{y} && = \frac{1}{n}\sum_{i=1}^{n} y_i \\
\hat{\theta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{E[(X-\bar{X})(Y-\bar{Y})]}{E[(X-\bar{X})^2]} = \frac{Cov(X,Y)}{Var(X)} \\
\hat{\theta}_0 &= \bar{y} - \hat{\theta}_1 \bar{x}
\end{aligned}$$

The above formulae can also be derived if the additives are a function of X. Since the linear relationship will still be respected and the loglikelihood can be maximized to get the estimates of the parameters.

19 Exercises

19.1 Problems

1. **Independence in Complements**
   Given $A \perp B$, show $A \perp B^c$ and $A^c \perp B^c$. Solution

2. **Conditional Independence**
   A, B, and C are independent with $P(C) > 0$. Show that $A \perp B|C$. Solution

3. **Geometry of Meeting**
   R and J have to meet at a given place and each will arrive at the given place independent of each other with a delay of 0 to 1hr uniformly distributed. The pairs of delays are all equally likely. The first to arrive waits for 15 minutes and leaves. What is the probability of meeting ? Solution

4. **Expectation of Function**
   Let X and Y be random variables with $Y = g(X)$. Show $E[Y] = \sum_x g(x)p_X(x)$. Solution

5. **Cumulative Distribution Function**
   A random variable X is a combination of a continuous and discrete distribution as follows

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

   Find the Cumulative Distribution of X. Solution

6. **Number of tosses till first head**
   When tossing a fair coin, what is the $E[\#$ tosses till the first H$]$. Solution

7. **Iterated Expectation Proof**
   For discrete variables, show $E[X] = E[E[X|Y]]$. Solution

8. **Iterated Expectation for three variables**
For three random variables X, Y and Z, show $E[Z|X] = E[E[Z|X,Y]|X]$. Solution

9. **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, $E[X|Y=1] = 90, E[X|Y=2] = 60, Var(X|Y=1) = 10, Var(X|Y=2) = 20$, find $E[X]$ and $Var(X)$. Solution

10. **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 ? Solution

11. **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. Solution

12. **Convolution of Exponentials**
Suppose $X \sim exp(\lambda)$ and $Y \sim exp(\mu)$, find the probability distribution $p_{X+Y}(x)$.

13. **Triangles from a Stick**
We have a stick of length 1. We randomly choose two points on the stick and break the stick at those points. Calculate the probability that the three pieces form a triangle. Solution

14. **PMF of g(X)**
Let X be uniform in $[0,2]$, then find the PMF of $Y = X^3$. Solution

15. **Waiting for Taxi**
A taxi stand and bus stop near Al's home are at the same location. Al goes there and if a taxi is waiting $P = \frac{2}{3}$, he boards it. Otherwise, he waits for a taxi or bus to come, whichever is first. Taxi takes anywhere between 0 to 10 mins (uniform) while a bus arrives in exactly 5 mins. He boards whichever is first. Find CDF and $E[\text{wait time}]$. Solution

16. **Bayes Theorem**
Let Q be a continuous random variable with PDF

$$f_Q(q) = \begin{cases} 6q(1-q) & 0 \leq q \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

where Q represents $P(success)$ for a Bernoulli X, i.e., $P(X=1|Q=q) = q$. Find $f_{Q|X}(q|x) \forall x \in [0,1] and q$. Solution

17. **A Normal Transformation**
Let $X \sim \mathcal{N}(0,1)$ and $Y = g(X)$. Find $p_Y(y)$.

$$g(t) = \begin{cases} -t & t \leq 0 \\ \sqrt{t} & t > 0 \end{cases}$$

Solution

18. **Binomial Shooter**
A shooter takes 10 hits in a shooting range and each shot has $p = 0.2$ of hitting target independent of each other. Let X = number of hits. Find

(a) PMF of X

(b) $P(no\ hits)$

(c) $P(scoring\ more\ than\ misses)$

(d) $E[X]$ and $Var(X)$

(e) Suppose the entry is \$3 and each shot fetches \$2. Let $Y =$ profit. Find $E[Y] and Var(Y)$.

(f) Suppose entry is free and total reward is square of number of hits. Let Z be profit. Find $E[Z]$.

Solution

19. **Mosquito and Tick**
Every second, a mosquito lands with $P = 0.5$. Once it lands, it bites with $P = 0.2$. Let X be the time between successive mosquito bites. Find $E[X]$ and $Var(X)$.
Now suppose a tick comes into play independent of mosquito. It lands with $P = 0.1$ and once landed, bites with $) = 0.7$. Let Y be the time between successive bug bites. Find $E[Y]$ and $Var(Y)$. Solution

20. **HH or TT**
Given a coin with $P(H) = p$, find the E[number of tosses till HH or TT]. Solution

21. **A Three Coin Game**
Let 3 fair coins be tossed at every turn. Given all coins and turns are independent, calculate the following (assuming success is defined as all three coins landing the same side up))

    (a) PMF of K, no of trials upto but not including the 2^{nd} success

    (b) E and Var of M, the E[number of tails] before first success.

    Solution

22. **Linear Expectations**
Bob conducts trials in a similar manner to Problem 21, but with four coins. He repeatedly removes a coin at success until just a single coin remains. Calculate the Expected number of tosses till the finish of experiment. Solution

23. **Papers Drawn with Replacement**
Suppose there are n papers in a drawer. We take one paper, sign it, and then put it back into the drawer. We take one more paper out and if it is not signed, we sign it and put it back in the drawer. If the paper is already signed, we simply put it back in the drawer. We repeat this process until all the papers are signed. Find the E[papers drawn till all papers are signed]. What is the value of this quantity as $n \to large$. Solution

24. **A Three Variable Inequality**
Let X, Y, Z be three exponentially distributed random variables with parameters λ, μ, and ν respectively. Find $P(X < Y < Z)$. Solution

25. **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)

    Solution

26. **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]

Solution

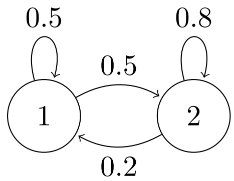
27. **Poisson Lightbulbs**
We have three identical but independent lightbulbs whose lifetimes are modelled by a Poisson process with parameter λ. Given that we start all the three bulbs together, find the E[time until last bulb dies out]. Solution

28. **Two Poisson Lightbulbs**
Beginning at $t = 0$, we begin using bulbs one at a time until failure. Any broken bulb is immediately replaced. Each new bulb is selected independently and equally likely from type A(exponential life with $\lambda = 1$) or type B(exponential life with $\lambda = 3$). Lifetimes of all bulbs are independent.

(a) Find E[time until first failure].
(b) P(no bulb failure before time t).
(c) Given that there are no failures until time t, determine the conditional probability that the first bulb used is of type A.
(d) Find the probability that the total illumination by two type B bulbs > one type A.
(e) Suppose the process terminates after 12 bulbs fail. Determine the expected value and variance of the total illumination provided by type B bulbs while the process is in operation.
(f) Given there are no failures until time t, find the expected value of time until first failure.

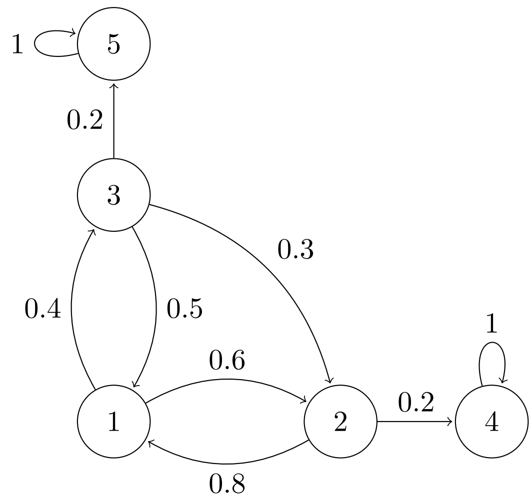
Solution

29. **Steady State Markov Process**
Find the steady state probabilites of the following Markov Process



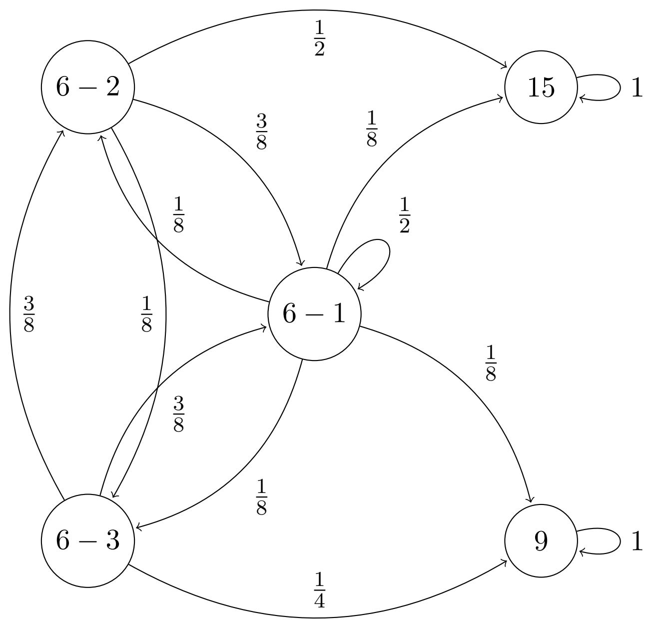
Solution

30. **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 $p_{32} = 0.5$) Solution

31. **Selecting Courses with Markov Process**



Consider the above markov process for changing courses. The probability being in some course tomorrow given a course today is mentioned along the edges. Suppose we start with course 6-1 (Note that course 6 is the combination of courses 6-1, 6-2 and 6-3). Calculate the following

(a) P(eventually leaving course 6).

(b) P(eventually landing in course 15).

(c) E[number of days till leaving course 6].

(d) At every switch for 6-2 to 6-1 or 6-3 to 6-1, we buy an ice cream (but a maximum of two). Calculate the E[number of ice creams before leaving course 6].

(e) Suppose we end up in 15. What is the E[number of steps to reach 15].

(f) Suppose we don't want to take course 15. Accordingly, when in 6-1, we stay there with probability 1/2 while other three options have equal probabilities. If we are in 6-2, probability of going to 6-1 and 6-3 are in the same ratio as before. Calculate the E[number of days until we enter course 9].

(g) Assuming $P(X_{n+1} = 15|X_n = 9) = P(X_{n+1} = 9|X_n = 15) = P(X_{n+1} = 15|X_n = 15) = P(X_{n+1} = 9|X_n = 9) = 1/2$, what is $P(X_n = 15)$ and $P(X_n = 9)$ far into the future.

(h) Suppose $P(X_{n+1} = 6{-}1|X_n = 9) = 1/8, P(X_{n+1} = 9|X_n = 9) = P(X_{n+1} = 15|X_n = 15) = 7/8$. What is the E[number of days till return to 6-1].

Solution

32. **Estimating Binomial with CLT, 1/2 correction**
Given a Bernoulli Process with $n = 36$ and $p = 0.5$, find $P(S_n \leq 21)$. Solution

33. **MLE Estimate**
Suppose we observe n independent and identically distributed samples $x_1, x_2, \ldots, x_n$ from an exponential distribution. Estimate the parameter of the exponential.

34. **LMS Estimate**
Given the prior $f_{\Theta|(\theta)}$, uniform in $[4, 10]$, and $f_{X|\Theta}(x|\theta)$ is uniform in $[\theta-1, \theta+1]$, estimate the posterior of θ. Solution

35. **Probability Convergence**
Let X be uniformly distributed between $[-1, 1]$. Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed with the same distribution as X. Find whether the following sequences are convergent in probability and also find the limit.

(a) X_i

(b) $Y_i = X_i/i$

(c) $Z_i = (X_i)^i$

19.2 Solutions

1. Question

(a)

$$\begin{aligned}
P(A \cap B) &= P(A)P(B) \\
P(A) &= P((A \cap B) \cup (A \cap B^c)) \\
&= P(A \cap B) + P(A \cap B^c) && \text{since disjoint} \\
P(A \cap B^c) &= P(A) - P(A)P(B) \\
&= P(A)(1 - P(B)) = P(A)P(B^c)
\end{aligned}$$

(b)

$$\begin{aligned}
(A \cup B)^c &= A^c \cap B^c \\
P(A^c \cap B^c) &= 1 - P(A \cup B) \\
&= 1 - P(A) - P(B) + P(A \cap B) \\
&= (1 - P(A))(1 - P(B)) \\
&= P(A^c)P(B^c)
\end{aligned}$$

2. Question

$$P(A \cap B|C) = \frac{P(A \cap B \cap C)}{P(C)} = P(A)P(B) = P(A|C)P(B|C) \quad \text{Due to independence}$$

3. Question
Suppose R arrives at x hours. J has to arrive between x hrs to x hrs + 15 mins. Similarly if J arrives at y hours, R has to arrive between y hours to y hours + 15 mins. These are regions enclosed by the regions $x \leq 1, y \leq 1, y \leq x + \frac{1}{4}$ *and* $y \geq x - \frac{1}{4}$. The probability is then $1 - P(\textit{not meeting}) = 1 - 2(\frac{1}{2}\frac{3}{4}\frac{3}{4}) = \frac{7}{16}$.

4. Question

$$\begin{aligned} E[Y] = \sum_y y p_Y(y) &= \sum_y \sum_{x:g(x)=y} p_X(x) = \sum_y \sum_{x:g(x)=y} y p_X(x) \\ &= \sum_y \sum_{x:g(x)=y} g(x) p_X(x) = \sum_x g(x) p_X(x) \end{aligned}$$

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

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

6. Question
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.

7. Question

Note that $E[X|Y]$ is a function of y.

$$\begin{aligned}
E[E[X|Y]] &= \sum_y E[X|Y]p_Y(y) \\
&= \sum_y \sum_x x p_{X|Y} p_Y \\
&= \sum_y \sum_x x p_{X,Y}(x,y) \\
&= \sum_x x \sum_y p_{X,Y}(x,y) \\
&= \sum_x x p_X(x) \\
&= E[X]
\end{aligned}$$

8. Question

Note that $E[Z|X,Y]$ will be a function of both X and Y.

$$\begin{aligned}
E[Z|X,Y] &= \sum_z z p_{Z|X,Y}(z|x,y) \\
E[E[Z|X,Y]|X] &= \sum_y E[Z|X,Y]p_{X,Y|X}(x,y|x) \\
&= \sum_y \sum_z z p_{Z|X,Y}(z|x,y) p_{Y|X}(y|x) \\
&= \sum_y \sum_z z \frac{p_{X,Y,Z}(x,y,z)}{p_X(x)} \\
&= \sum_z z \sum_y \frac{p_{X,Y,Z}(x,y,z)}{p_X(x)} \\
&= \sum_z z \frac{p_{X,Z}(x,z)}{p_X(x)} \\
&= \sum_z z p_{Z|X}(z|x) \\
&= E[Z|X]
\end{aligned}$$

9. Question

We use the formulae from iterated expectation to calculate these.

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

10. Question

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,

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

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

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

Putting these values in the original equation for variance

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

11. Question

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

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

12. Question

Note that the required probability distribution is given by the following formula

$$p_{X+Y}(x) = \int_{-\inf}^{\inf} p_X(x-y) * p_Y(y)dy$$

However, note that the exponential distribution is not positive everywhere. For values < 0, the probability density is 0. Hence, we break the integral into three parts as follows

$$p_{X+Y}(x) = \int_{-\inf}^{0} p_X(x-y) * p_Y(y)dy + \int_0^x p_X(x-y) * p_Y(y)dy + \int_x^{\inf} p_X(x-y) * p_Y(y)dy$$

Carefully note that for y in range $(-\inf, 0]$, $p_Y(y) = 0$, and in the range $[x, \inf)$, $x - y < 0$, which implies $p_X(x) = 0$. Hence,

$$p_{X+Y}(x) = \int_0^x p_X(x-y) * p_Y(y)dy$$
$$= \lambda\mu\exp(-\lambda x)\int_0^x \exp((\lambda-\mu)y)dy$$
$$= \frac{\lambda\mu}{\lambda-\mu}\exp(-\lambda x)(\exp((\lambda-\mu)x) - 1)$$
$$= \frac{\lambda\mu}{\lambda-\mu}(\exp(\mu x) - \exp(-\lambda x))$$

13. Question
Assume that we break the stick at points X and Y. Assume $X < Y$. Then for the stick to form a triangle, the three lengths $X, Y - X$ and $1 - Y$ should satisfy the following three inequalities

$$\begin{aligned} X + (Y - X) &> 1 - Y \\ (Y - X) + (1 - Y) &> X \\ X + (1 - Y) &> Y - X \end{aligned}$$

which is nothing but the triangluar region between the points $(0, 0.5), (0.5, 0.5)$ and $(0.5, 1)$ and has the area of $1/8$. We should also consider the case $Y < X$ and by symmetry, the area is same. Now, X and Y comprise of the entire square region $X \leq 1$ and $Y \leq 1$. Hence the required probability is $2 * 1/8 = 1/4$.

14. Question
Always solve such questions using the cumulative distribution approach.

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

$$\begin{aligned} P(Y \leq y) &= P(X^3 \leq y) = P(X \leq y^{\frac{1}{3}}) \\ &= \begin{cases} 0 & y < 0 \\ \frac{1}{2}y^{\frac{1}{3}} & 0 \leq y^{\frac{1}{3}} \leq 2 \\ 1 & 2 < y^{\frac{1}{3}} \end{cases} \\ f_Y(y) &= \frac{dP(Y <= y)}{dy}(y) \\ &= \begin{cases} 0 & y < 0 \\ \frac{1}{6}y^{\frac{-2}{3}} & 0 \leq y \leq 8 \\ 0 & 8 < y \end{cases} \end{aligned}$$

15. Question
Let X be the waiting time and $F_X(x)$ be the CDF. Then,

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{2}{3} & x = 0 \\ \frac{2}{3} + \frac{1}{30}x & 0 < x < 5 \\ 1 & 5 \leq x \end{cases}$$

The PDF is simply the derivate of the CDF. Thus, expectation is

$$E[X] = \frac{2}{3}(0) + \int_0^5 \frac{1}{30}xdx + \frac{1}{6}(5) = \frac{5}{4}mins$$

16. Question
From Bayes' theorem

$$\begin{aligned} f_{Q|X}(q|x) &= \frac{f_{X|Q}(x|q)f_Q(q)}{f_X(x)} \\ &= \frac{f_{X|Q}(x|q)f_Q(q)}{\int_0^1 f_{X|Q}(x|q)f_Q(q)dq} \end{aligned}$$

We will need to solve separately for $x = 0$ and $x = 1$ as x is discrete.

$$f_{Q|X=0}(q|x=0) = \frac{(1-q)*6q(1-q)}{\int_0^1 (1-q)*6q(1-q)dq} = 12q(1-q)^2$$
$$f_{Q|X=1}(q|x=1) = \frac{q*6q(1-q)}{\int_0^1 q*6q(1-q)dq} = 12q^2(1-q)$$

17. Question
Questions of this type must only be approached through CDF. First find the CDF of Y and then it's PDF.

$$\begin{aligned}
F_Y(y) &= P(Y \le y) = P(g(X) <= y) \\
&= P(X \in [-y, 0] \cup X \in [0, y^2]) \\
&= (F_X(0) - F_X(-y)) + (F_X(y^2) - F_X(0)) \\
&= F_X(y^2) - F_X(-y) \\
p_Y(y) &= \frac{F_Y(y)}{dy} \\
&= \frac{dF_X(y^2)}{dx}\frac{d(y^2)}{dy} - \frac{dF_X(-y)}{dx}\frac{d(-y)}{dy} \\
&= 2yp_X(y^2) + p_X(-y) \\
&= 2y\frac{1}{\sqrt{2\pi}}e^{-\frac{y^4}{2}} + \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}
\end{aligned}$$

18. Question
   (a) $P(X = k) = \binom{10}{k}0.2^k 0.8^{10-k}$
   (b) $P(no\ hits) = 0.8^{10}$
   (c) $P(X >= 6)\ \sum_{k=6}^{10}\binom{10}{k}0.2^k 0.8^{10-k}$
   (d) $E[X] = np = 2, Var(X) = np(1-p) = 1.6$ for Bernoulli distribution
   (e) $Y = 2X - 3, E[Y] = 2E[X] - 3 = 1, Var(Y) = 4Var(X) = 6.4$
   (f) $Z = X^2, E[Z] = E[X^2] = Var(X) + E[X]^2 = 5.6$

19. Question
For the mosquito, $P(bite) = P(land)P(bite|land) = 0.1$. X is a geometric random variable. $E[X] = 1/p = 10$ and $Var(X) = \frac{1-p}{p^2} = 90$.
For the mosquito and tick combined, $P(\text{mosquito and tick}) = 0.1 + 0.1*0.7 - 0.1*0.1*0.7 = 0.163$. This is again a geometric random variable with $E[Y] = 1/0.163$ and $Var(Y) = (1-0.163)/(0.163^2)$.s

20. Question
This quantity can be calculated using the law of total expectation

$$E[X] = E[X|A_1]P(A_1) + E[X|A_2]P(A_2) + \cdots + E[X|A_n]P(A_n) \quad \text{where } A_i \text{ are disjoint}$$

Let H_1 denote heads at first toss, H_2 denote heads at the second toss, T_1 denote tails at first toss and T_2 denote tails at the second toss. Then,

$$\begin{aligned}
E[X] &= E[X|H_1]P(H_1) + E[X|T_1]P(T_1) \\
E[X|H_1] &= E[X|H_1H_2]P(H_2|H_1) + E[X|H_1T_2]P(T_2|H_1) \\
&= 2p + (1 + E[X|T_1])(1-p) \\
E[X|T_1] &= E[X|T_1T_2]P(T_2|T_1) + E[X|T_1H_2]P(H_2|T_1) \\
&= 2(1-p) + (1 + E[X|H_1])p
\end{aligned}$$

$E[X|H_1T_2] = 1 + E[X|T_1]$ because the tails after the first heads implies the first heads is now irrelevant and we have wasted one toss on the heads. The remaining process is same as starting from the first coin toss as tails.
Solving for the conditional expectations,

$$\begin{aligned}
E[X|H_1] &= \frac{3-2p+p^2}{1-p+p^2} \\
E[X|T_1] &= \frac{2+p^2}{1-p+p^2} \\
E[X] &= \frac{2+p-p^2}{1-p+p^2}
\end{aligned}$$

21. Question
Define X as the following random variable

$$X = \begin{cases} 1, p = \frac{1}{4} & HHH \text{ or } TTT \\ 0, p = \frac{3}{4} & \text{otherwise} \end{cases}$$

(a) K is simply a binomial distribution, where we want the 2^{nd} success to happen at the $K+1$th trial.

$$p_K(k) = \binom{k}{1} \frac{1}{4}^2 \frac{3}{4}^{k-1} \qquad \text{since the last trial is success}$$

(b) M = number of tails before first success. Let the success be at $N+1$. Defin Y as

$$\begin{aligned}
Y &= \begin{cases} 1 \ \ p = \frac{1}{2} & HHT,\ HTH, \text{ or } THH \\ 2 \ \ p = \frac{1}{2} & HTT,\ THT, \text{ or } TTH \end{cases} \\
E[Y] &= 1 * \frac{1}{2} + 2 * \frac{1}{2} \\
Var(Y) &= (1 - \frac{3}{2})^2 * \frac{1}{2} + (2 - \frac{3}{2})^2 * \frac{1}{2} \\
E[N+1] &= \frac{1}{p} = 4 \\
Var(N+1) = Var(N) &= \frac{1-p}{p^2} = \frac{1-\frac{1}{4}}{\frac{1}{4}^2} \\
M &= Y_1 + Y_2 + \cdots Y_N \\
E[M] &= E[Y_1 + Y_2 + \cdots Y_N] \\
Var(M) &= Var(Y_1 + Y_2 + \cdots Y_N)
\end{aligned}$$

Note that both Y and N are random variables here. Using the formulae for random number of random variables,

$$E[M] = E[E[M|N]] = E[NE[Y]] = E[N]E[Y] = (4-1)*\frac{3}{2} = \frac{9}{2}$$
$$Var(M) = Var(E[M|N]) + E[Var(M|N)] = Var(NE[Y]) + E[NVar(Y)]$$
$$= E[Y]^2Var(N) + E[N]Var(Y) = \frac{9}{4}*12 + 3*\frac{1}{4} = \frac{111}{4}$$

22. Question
Let X be the number of tosses till the first coin is removed. This is a geometric random variable with $P(\text{success}) = \frac{1}{8}$. then $E[X] = 1/p = 8$. Now Y be the number of tosses till the second coin is removed (counting tosses after removal of first coin). Note that geometric random variables are memory less and what happened before the start of the "experiment" will not matter. Thus, $E[Y] = 1/(1/4) = 4$. Similarly, Z is the tosses till the last coin is removed and $E[Z] = 1/(1/2) = 2$. Note that the number of tosses till the end of experiment is simply $X+Y+Z$. $E[X+Y+Z] = E[X]+E[Y]+E[Z] = 14$.

23. Question
Note that the process till the end is a combination of multiple binomial process, such that any process lasts till the first success. Suppose we sign a paper and keep this in the drawer. Now the total signed papers in the drawer is k out of n and the $P(\text{success}) = \frac{n-k}{n}$ and $E[\text{draws till next unsigned paper}] = \frac{1}{p} = \frac{n}{n-k}$. Total draws

$$E = \frac{n}{1} + \frac{n}{2} + \cdots + \frac{n}{n}$$
$$= n(1 + \frac{1}{2} + \cdots + \frac{1}{n})$$
$$\lim_{n \to large} E = n\log(n)$$

24. Question
A very straightforward way is to use a triple integral

$$P(X < Y < Z) = \int_0^{\inf}\int_0^z\int_0^y \lambda e^{-\lambda x}\mu e^{-\mu y}\nu e^{-\nu z}dxdydz = \frac{\lambda\mu}{(\lambda+\mu+\nu)(\mu+\nu)}$$

$P(X < Y < Z)$ can be broken down as $P(X < min(Y,Z))P(Y < Z)$. Consider just P(Y ¡ Z)

$$P(Y < Z) = \int_0^{\inf}\int_0^z \mu e{-}\mu y\nu e{-}\nu z dydz = \frac{\mu}{\mu+\nu}$$

Thus, when two exponential processes are considered, probaility of arrival of 1st before 2nd is simply the percentage ratio of parameters. Thus,

$$P(X < min(Y,Z)) = \frac{\lambda}{\lambda+(\mu+\nu)} \qquad Y \text{ and } Z \text{ can be combined as a single process}$$
$$P(Y < Z) = \frac{\mu}{\mu+\nu}$$
$$P(X < Y < Z) = P(X < min(Y,Z))P(Y < Z)$$
$$= \frac{\lambda\mu}{(\lambda+\mu+\nu)(\mu+\nu)}$$

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

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

26. Question

- 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, $E[T_1] = \frac{1}{\lambda}$

27. Question
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] $= \frac{1}{3\lambda} + \frac{1}{2\lambda} + \frac{1}{\lambda}$

28. Question

(a)

$$\begin{aligned} E[\text{time till failure}] &= E[\text{time till failure}|A]P(A) + E[\text{time till failure}|B]P(B) \\ &= \frac{1}{\lambda_A}\frac{1}{2} + \frac{1}{\lambda_B}\frac{1}{2} \\ &= \frac{1}{2}(1 + \frac{1}{3}) = \frac{2}{3} \end{aligned}$$

(b) Let C denote the event of no failure till time t. $P(C)$ for a given λ will be $\int_t^{\text{inf}} \lambda e^{-\lambda t}$. Then,

$$\begin{aligned} P(C) &= P(C|A)P(A) + P(C|B)P(B) && \text{Using total probability theorem} \\ &= e^{-t}(\frac{1}{2}) + e^{-3t}(\frac{1}{2}) \\ &= \frac{1}{2}(e^{-t} + e^{-3t}) \end{aligned}$$

(c) Let C denote the event of no failure till time t. Then,

$$\begin{aligned}
P(A|C) &= \frac{P(C|A)P(A)}{P(C)} \\
&= \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C|B)P(B)} \\
&= \frac{\frac{1}{2}e^{-t}}{\frac{1}{2}(e^{-t} + e^{-3t})} \\
&= \frac{1}{1 + e^{-2t}}
\end{aligned}$$

(d) Let T_{B1}, T_{B2} and T_A denote the life times of the first B bulb, second B bulb and the A bulb respectively. First consider the solution to $P(T_{B1} + T_{B2} = t)$

$$\begin{aligned}
P(T_{B1} + T_{B2} = t) &= \int_0^t P(T_{B1} = t_1)P(T_{B2} = t - t_1)dt_1 \qquad \text{Using independence} \\
&= \int_0^t 3e^{-3t_1}3e^{-3(t-t_1)}dt_1 \\
&= \int_0^t 9e^{-3t}dt_1 \\
&= 9te^{-3t}
\end{aligned}$$

Now, we can rewrite the requred probability in a slightly different format

$$\begin{aligned}
P(T_{B1} + T_{B2} > T_A) &= P(T_{B1} + T_{B2} = t)P(T_A \le t) \\
&= \int_0^{\text{inf}} 9te^{-3t}(\int_0^t e^{-t_1}dt_1)dt \\
&= \int_0^{\text{inf}} 9te^{-3t}(1 - e^{-t})dt \\
&= \int_0^{\text{inf}} 9te^{-3t} - 9te^{-4t}dt
\end{aligned}$$

Using integration by parts, $\int uv' = uv - \int u'v$ and choosing $u = t, v = e^{-3t}/3$,

$$\begin{aligned}
P(T_{B1} + T_{B2} > T_A) &= \left[9[te^{-3t}]_0^{\text{inf}} - 3\int_0^{\text{inf}} e^{-3t}dt - 9[te^{-4t}]_0^{\text{inf}} + \frac{9}{4}\int_0^{\text{inf}} e^{-4t}dt\right] \\
&= 0 + 1 - 0 - \frac{9}{16} = \frac{7}{16}
\end{aligned}$$

(e) Let there be N bulbs of type B out of the 12 bulbs. Clearly N is a random variable and can be seen as the "successes" of choosing a given bulb as B. and the probability of choosing any ith bulb as B is 1/2.
Let the life time of any bulb of type B be T. Then the total lifetime of all the type B bulbs will be NT, which is nothing but the sum of a random number of random variables.

$$\begin{aligned}
E[NT] &= E[N]E[T] = np * \frac{1}{\lambda} = 12 * \frac{1}{2} * \frac{1}{3} = 2 \\
Var(NT) &= E[Var(NT|N)] + Var(E[NT|N]) = E[N]Var(T) + E[T]^2Var(N) \\
&= np * \frac{1}{\lambda^2} + (\frac{1}{\lambda})^2 np(1-p) = 1
\end{aligned}$$

(f) Let D be the event that the lifetime is greater thatn t or $T > t$. Then,

$$\begin{aligned}
E[T|D] &= E[T|D, A]P(A|D) + E[T|D, B]P(B|D) \\
&= t + (E[T - t|D, A]P(A|D) + E[T - t|D, B]P(B|D)) \\
&= t + (\frac{1}{1}P(A|D) + \frac{1}{3}P(B|D)) && \text{Using memoryless property} \\
&= t + (\frac{1}{1 + e^{-2t}} + \frac{1}{3}(1 - \frac{1}{1 + e^{-2t}})) && \text{Using part 28c} \\
&= t + \frac{1}{3} + \frac{2}{3}\frac{1}{1 + e^{-2t}}
\end{aligned}$$

29. Question

Using balance equations, we have

$$\begin{aligned}
\pi_1 &= \pi_1 p_{11} + \pi_2 p_{21} \\
\pi_2 &= \pi_1 p_{12} + \pi_2 p_{22} \\
\pi_1 + \pi_2 &= 1
\end{aligned}$$

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

30. Question

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 $\mu_3 = \frac{111}{8}$

31. Question

(a) The probability of eventually leaving course 6 is 1 as states 15 and 9 are absorbing states.

(b) Here we have to calculate the probability of absortion into state 15. Let a_i denote the probability of absorption into state 15 from state i. Then, $a_{15} = 1$ and $a_9 = 0$. Using equations from 16.4,

$$\begin{aligned}
a_{6-1} &= \frac{1}{2}a_{6-1} + \frac{1}{8}a_{6-2} + \frac{1}{8}a_{6-3} + \frac{1}{8}a_9 + \frac{1}{8}a_{15} \\
a_{6-2} &= \frac{1}{2}a_{15} + \frac{3}{8}a_{6-1} + \frac{1}{8}a_{6-3} \\
a_{6-3} &= \frac{1}{4}a_9 + \frac{3}{8}a_{6-1} + \frac{3}{8}a_{6-2}
\end{aligned}$$

Solving the 3 equations, 3 variable system, $a_{6-1} = 105/184, a_{6-2} = 143/184$ and $a_{6-3} = 93/184$.

(c) Let μ_i denote the expected number of steps to get absorbed starting from state i. Then, $\mu_{15} = \mu_9 = 0$. Using equations from 16.4,

$$\begin{aligned}
\mu_{6-1} &= 1 + \frac{1}{2}\mu_{6-1} + \frac{1}{8}\mu_{6-2} + \frac{1}{8}\mu_{6-3} + \frac{1}{8}\mu_9 + \frac{1}{8}\mu_{15} \\
\mu_{6-2} &= 1 + \frac{1}{2}\mu_{15} + \frac{3}{8}\mu_{6-1} + \frac{1}{8}\mu_{6-3} \\
\mu_{6-3} &= 1 + \frac{1}{4}\mu_9 + \frac{3}{8}\mu_{6-1} + \frac{3}{8}\mu_{6-2}
\end{aligned}$$

Solving, $\mu_{6-1} = 81/23, \mu_{6-2} = 63/23$ and $\mu_{6-3} = 77/23$.

(d) This question can be done in a manner similar to the equations described above but with a small adjustment. Note that, we can either have 0, 1, or 2 ice creams. Consider $v_i(j)$ as the probability of making j additional ice creams from 6-2 to 6-1 or 6-3 to 6-1 transitions, given the current state is i. Note $v_{15}(0) = v_9(0) = 1$. Then,

$$\begin{aligned}
v_{6-1}(0) &= \frac{1}{2}v_{6-1}(0) + \frac{1}{8}v_{6-2}(0) + \frac{1}{8}v_{6-3}(0) + \frac{1}{8}v_9(0) + \frac{1}{8}v_{15}(0) \\
v_{6-2}(0) &= \frac{1}{2}v_{15}(0) + \frac{3}{8}(0) + \frac{1}{8}v_{6-3}(0) \\
v_{6-3}(0) &= \frac{1}{4}v_9(0) + \frac{3}{8}(0) + \frac{3}{8}v_{6-2}(0)
\end{aligned}$$

Some of the transitions have been directly replaced with 0 as we are considering 0 ice creams and thus those transitions are not possible (6-2 to 6-1 for instance). Solving, $v_{6-1}(0) = 46/61, v_{6-2}(0) = 34/61$ and $v_{6-3}(0) = 28/61$.

The same way, we can construct equations for 1 additional steps where $v_{15}(1) = v_9 = 0$.

$$\begin{aligned}
v_{6-1}(1) &= \frac{1}{2}v_{6-1}(1) + \frac{1}{8}v_{6-2}(1) + \frac{1}{8}v_{6-3}(1) + \frac{1}{8}v_9(1) + \frac{1}{8}v_{15}(1) \\
v_{6-2}(1) &= \frac{1}{2}v_{15}(1) + \frac{3}{8}v_{6-1}(0) + \frac{1}{8}v_{6-3}(1) \\
v_{6-3}(1) &= \frac{1}{4}v_9(1) + \frac{3}{8}v_{6-1}(0) + \frac{3}{8}v_{6-2}(1)
\end{aligned}$$

In the second equation, after going from 6-2 to 6-1, we can only get 0 more ice creams. Hence, some of the values have been replaced with the $v_i(0)$ calculated above. Solving, $v_{6-1}(1) = 690/3721, v_{6-2}(1) = 1242/3721$ and $v_{6-3}(1) = 1518/3721$.

Note that since the total ice creams are 0, 1, or 2, we have $v_{6-1}(0) + v_{6-1}(1) + v_{6-1}(2) = 1$. $E[\text{ice creams}] = 0 * v_{6-1}(0) + 1 * v_{6-1}(1) + 2 * v_{6-1}(2) = 1140/3721$

(e) We need to recalculate the the transition probabilities since we are conditioning on the event A that we land up in state 15.

$$\begin{aligned}
P_{ij|A} &= P(X_{n+1} = j | X_i = i, A) \\
&= \frac{P(X_{n+1} = j, X_n = i, A)}{P(X_n = i, A)} \\
&= \frac{P(A|X_{n+1} = j, X_n = i)P(X_{n+1} = j|X_n = i)P(X_n = i)}{P(A|X_n = i)P(X_n = i)} \\
&= \frac{P(A|X_{n+1} = j)P(X_{n+1} = j|X_n = i)}{P(A|X_n = i)} \\
&= \frac{a_j}{a_i}P_{ij}
\end{aligned}$$

where a_i is the probability of absorption into state 15 starting from state i. Since markov process is only dependent on the last state, absoprtion probabilities are not dependent on n.

We can write equations similar to 16.4 for calculating the expected number of steps with the adjusted transition probabilities

$$\begin{aligned}
\mu_{6-1} &= 1 + \frac{a_{6-1}}{a_{6-1}}\frac{1}{2}\mu_{6-1} + \frac{a_{6-2}}{a_{6-1}}\frac{1}{8}\mu_{6-2} + \frac{a_{6-3}}{a_{6-1}}\frac{1}{8}\mu_{6-3} + \frac{a_{15}}{a_{6-1}}\frac{1}{8}\mu_{15} + \frac{a_9}{a_{6-1}}\frac{1}{8}\mu_9 \\
\mu_{6-2} &= 1 + \frac{a_{6-1}}{a_{6-2}}\frac{3}{8}\mu_{6-1} + \frac{a_{6-3}}{a_{6-2}}\frac{1}{8}\mu_{6-3} + \frac{a_{15}}{a_{6-2}}\frac{1}{2}\mu_{15} \\
\mu_{6-3} &= 1 + \frac{a_{6-1}}{a_{6-3}}\frac{3}{8}\mu_{6-1} + \frac{a_{6-2}}{a_{6-3}}\frac{3}{8}\mu_{6-2} + \frac{a_9}{a_{6-3}}\frac{1}{4}\mu_9
\end{aligned}$$

where $\mu_{15} = \mu_9 = 0, a_{15} = 1$, and $a_9 = 0$. The absorption probabilities can be taken from the part 31b. Solving, $\mu_{6-1} = 1763/483$.

(f) The changed probabilites become $P(X_{n+1} = 15|X_n = 6-1) = P(X_{n+1} = 6-2|X_n = 6-1) = P(X_{n+1} = 6-3|X_n = 6-1) = 1/6, P(X_{n+1} = 6-1|X_n = 6-2) = 3/4$ and $P(X_{n+1} = 6-3|X_n = 6-2) = 1/4$. We then use equations from 16.4 to calculate the expected values

$$\begin{aligned}
\mu_{6-1} &= 1 + \frac{1}{2}\mu_{6-1} + \frac{1}{6}\mu_{6-2} + \frac{1}{6}\mu_{6-3} + \frac{1}{6}\mu_9 \\
\mu_{6-2} &= 1 + \frac{3}{4}\mu_{6-1} + \frac{1}{4}\mu_{6-3} \\
\mu_{6-3} &= 1 + \frac{1}{4}\mu_9 + \frac{3}{8}\mu_{6-1} + \frac{3}{8}\mu_{6-2}
\end{aligned}$$

where $\mu_{15} = 0$. Solving, $\mu_{6-1} = 86/13, \mu_{6-2} = 98/13$ and $\mu_{6-3} = 82/13$.

(g) If we look carefully at the new probabilities, states 15 and 9 become recurrent. Far into the future, we are sure to land up in those states, and will be in either one of those. By symmetry, the two should be same. $\pi_{15} = \pi_9 = 1/2$.

(h) We assume that 6-1 is an absorbing state, and accordingly calculate the probabilities. Note that there will not be an equation for 6-1 since we are then already in the final state.

$$\mu_{6-2} = 1 + \frac{1}{8}\mu_{6-3} + \frac{1}{2}\mu_{15}$$
$$\mu_{6-3} = 1 + \frac{3}{8}\mu_{6-2} + \frac{1}{4}\mu_{9}$$
$$\mu_{9} = 1 + \frac{7}{8}\mu_{9}$$
$$\mu_{15} = 1 + \frac{7}{8}\mu_{15}$$

Solving, $\mu_{6-2} = 344/61, \mu_{6-3} = 312/61$ and $\mu_9 = \mu_{15} = 8$. Plugging these into the following equation (which corresponds to taking one step out of 6-1),

$$\mu_{6-1} = 1 + \frac{1}{2}\mu_{6-1} + \frac{1}{8}\mu_{15} + \frac{1}{8}\mu_{6-2} + \frac{1}{8}\mu_{6-3} + \frac{1}{8}\mu_{15} = \frac{265}{61}$$

32. Question The exact answer will be

$$\sum_{k=0}^{21} \binom{36}{k} (\frac{1}{2})^{36} = 0.8785$$

But the same can be estimated using the CLT as follows

$$\mu = np = 18$$
$$\sigma^2 = np(1-p) = 9$$
$$P(S_n \leq 21) \approx P(\frac{S_n - 18}{3} \leq \frac{21-18}{3}) \approx 0.843$$

Our estimate is in the rough range of the answer but not quite close. We can do better using the $\frac{1}{2}$ correction

$$P(S_n \leq 21) = P(S_n < 22) \text{ since } S_n \text{ is an integer}$$
$$\text{Consider } P(S_n <= 21.5) \text{ as a compromise between the two}$$
$$P(S_n <= 21.5) = P(\frac{S_n - 18}{3} \leq \frac{21.5-18}{3}) \approx 0.879$$

In a similar manner, $P(S_n = 19) = P(18.5 \leq S_n \leq 19.5)$ using $\frac{1}{2}$ correction.

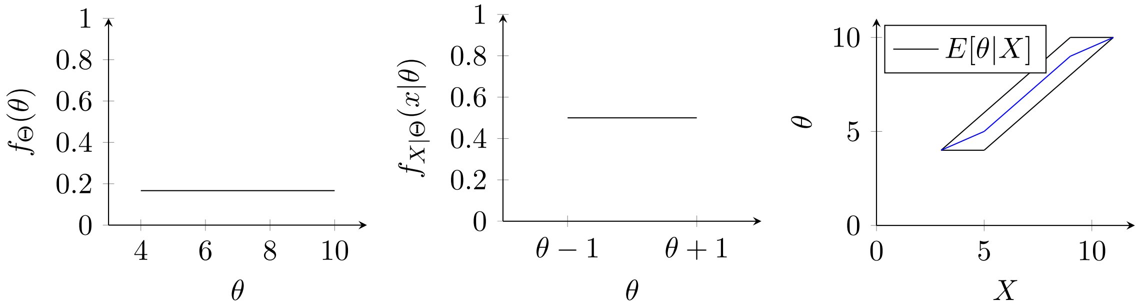
33. Question
Since the observations are independent, the likelihood of all the observations under some θ is given by

$$p_{X|\Theta}(x|\theta) = \prod_{i=1}^{n} \theta \exp(-\theta x_i)$$
$$log(p_{X|\Theta}(x|\theta)) = nlog(\theta) - \theta(\sum_{i=1}^{n} x_i)$$

Taking the derivatie and maximizing with respect to θ, $\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^{n} x_i}$

34. Question
We need to evaluate $f_{\Theta|X}(\theta|x)$ in order to get $E[\Theta|X]$.
$f_{X,\Theta}(x,\theta) = f_X(x) f_{\Theta|X}(\theta|x)$ which is a parallelogram on the $\theta - x$ plane at the points

(3, 4), (5, 4), (9, 10) and (11, 10). Then $E[\Theta|X]$ can be obtained by drawing vertical lines on the planes and calculating the $E[\theta]$ over that line. It is a line which bends at two points.



35. Question

(a) No, since X_i is also uniform in $[-1, 1]$

(b) Yes, $E[Y_i] = 0$ by symmetry. For $\epsilon > 0$,

$$\begin{aligned}\lim_{i\to\inf}(P|Y_i - \mu_i| > \epsilon) &= \lim_{i\to\inf} P(|\frac{X_i}{i} - 0| > \epsilon)\\ &= \lim_{i\to\inf} P(\frac{X_i}{i} > \epsilon \text{ and } \frac{X_i}{i} < -\epsilon)\\ &= \lim_{i\to\inf}[P(X_i > i\epsilon) + P(X_i < -i\epsilon)] = 0\end{aligned}$$

(c) Yes, $E[Y_i] = 0$ by symmetry. For $\epsilon > 0$,

$$\begin{aligned}\lim_{i\to\inf} P(|Z_i - 0| > \epsilon) &= \lim_{i\to\inf} P((X_i)^i > \epsilon \text{ or } (X_i)^i < -\epsilon)\\ &= \lim_{i\to\inf}[\frac{1}{2}(1 - \epsilon^{1/i}) + \frac{1}{2}(1 - \epsilon^{1/i})]\\ &= \lim_{i\to\inf}(1 - \epsilon^{1/i}) = 0\end{aligned}$$