Probability and Statistics

Endsem Solutions

Q1. Moment Generating Functions

(a) Geometric with parameter p

Let $X \sim \text{Geometric}(p)$ denote the number of trials until the first success, where each trial is independent with success probability p, $0 < p < 1$. Then the probability mass function (pmf) of X is

$$\mathbb{P}(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, 3, \ldots$$

The moment generating function (mgf) of X is defined as

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=1}^{\infty} e^{tk} \mathbb{P}(X = k).$$

Substituting the pmf, we get

$$\begin{aligned} M_X(t) &= \sum_{k=1}^{\infty} e^{tk} (1-p)^{k-1} p \\ &= p \sum_{k=1}^{\infty} \left(e^t\right)^k (1-p)^{k-1} \\ &= pe^t \sum_{k=1}^{\infty} \left(e^t(1-p)\right)^{k-1} \\ &= pe^t \sum_{n=0}^{\infty} \left(e^t(1-p)\right)^n, \end{aligned}$$

where we have re-indexed with $n = k-1$. This is a geometric series with ratio $r = e^t(1-p)$, and it converges for

$$|e^t(1-p)| < 1.$$

Using the sum of a geometric series,

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \quad |r| < 1,$$

we obtain

$$M_X(t) = pe^t \cdot \frac{1}{1-(1-p)e^t} = \frac{pe^t}{1-(1-p)e^t},$$

valid for $e^t(1-p) < 1$, i.e., $t < -\ln(1-p)$. Thus,

$$\boxed{M_X(t) = \frac{pe^t}{1-(1-p)e^t}, \quad \text{for } t < -\ln(1-p).}$$

(b) Exponential with parameter λ

Let $X \sim \text{Exponential}(\lambda)$ with parameter $\lambda > 0$. Then the probability density function (pdf) of X is

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

The moment generating function of X is

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_0^\infty e^{tx} f_X(x)\, dx = \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx.$$

Combine the exponents:

$$M_X(t) = \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx.$$

For the integral to converge, we require $\lambda - t > 0$, i.e., $t < \lambda$. Using the standard result

$$\int_0^\infty e^{-ax}\, dx = \frac{1}{a}, \quad a > 0,$$

we get

$$M_X(t) = \lambda \cdot \frac{1}{\lambda - t} = \frac{\lambda}{\lambda - t},$$

valid for $t < \lambda$. Thus,

$$\boxed{M_X(t) = \frac{\lambda}{\lambda - t}, \quad \text{for } t < \lambda.}$$

Q2. Monte Carlo Estimation of π

Estimation Method:

Use two independent uniform samples to produce a point in the unit square $[0, 1]^2$. Count the number of points that fall inside the quarter of the unit circle of radius 1 centered at the origin. Repeat n times and rescale the fraction of "hits" to obtain an estimator for π.

More formally, let (X_i, Y_i) for $i = 1, 2, \ldots, n$ be i.i.d. samples from the uniform distribution on $[0, 1]^2$. Define the Bernoulli random variable:

$$\mathbb{I}_{\{X_i^2+Y_i^2\leq 1\}}$$

which is 1 if the point (X_i, Y_i) lies inside the quarter circle and 0 otherwise. The estimator for π is given by:

$$\hat{\pi} = 4 \cdot \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i^2+Y_i^2\leq 1\}}$$

Justification

Intuitively, the ratio of the area of the quarter circle to the area of the unit square gives the probability that a randomly chosen point in the unit square falls inside the quarter circle.

However, a more formal justification is as follows:

$$\int_0^1 \sqrt{1-x^2}dx = \int_0^{\pi/2} \cos^2(\theta)d\theta = \frac{\pi}{4}$$

Now, if we evaluate the integral using Monte Carlo integration, we get a way to estimate π.

$$\int_0^1 \sqrt{1-x^2}dx = \int_0^1 \int_0^{\sqrt{1-y^2}} 1dxdy = \int_0^1 \int_0^1 \mathbb{I}_{\{x^2+y^2\leq 1\}}dxdy = \mathbb{E}_{(X,Y)\sim \text{Uniform}(0,1)^2}[\mathbb{I}_{\{X^2+Y^2\leq 1\}}]$$

Where $\mathbb{I}_{\{x^2+y^2\leq 1\}}$ is a Bernoulli random variable. Now, using Monte Carlo integration, we can estimate the above integral as follows:

$$\mathbb{E}_{(X,Y)\sim \text{Uniform}(0,1)^2}[\mathbb{I}_{\{X^2+Y^2\leq 1\}}] \approx \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i^2+Y_i^2\leq 1\}}$$

Where $X_i, Y_i \sim \text{Uniform}(0, 1)$ i.i.d. Thus, we have an unbiased estimator for π as:

$$\hat{\pi} = 4 \cdot \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i^2+Y_i^2\leq 1\}}$$

Q3. Sum of Independent Exponential Random Variables

Part 1: Derivation using Convolution Formula

Let $Z = X_1 + X_2$. Since X_1 and X_2 are independent continuous random variables, the PDF of Z is given by the convolution of their marginal PDFs:

$$f_Z(z) = (f_{X_1} * f_{X_2})(z) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z-x)\, dx$$

Given that $X_i \sim \text{Exp}(\lambda)$, the marginal PDFs are:

$$f_{X_i}(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

Determining the Limits of Integration:

For the integrand $f_{X_1}(x) f_{X_2}(z-x)$ to be non-zero, both individual terms must be non-zero simultaneously. This imposes two constraints on the integration variable x:

1. From $f_{X_1}(x)$, we require $x \geq 0$.
2. From $f_{X_2}(z-x)$, we require argument $z - x \geq 0$, which implies $x \leq z$.

Combining these constraints ($0 \leq x \leq z$):

- **Case 1 ($z < 0$):** The interval $[0, z]$ is empty (since x cannot be greater than 0 and less than a negative number). Thus, the integral is 0.
- **Case 2 ($z \geq 0$):** The valid range for integration is $x \in [0, z]$.

Evaluating the Integral for $z \geq 0$:

Substituting the explicit functional forms into the integral:

$$\begin{aligned} f_Z(z) &= \int_0^z \left(\lambda e^{-\lambda x}\right) \left(\lambda e^{-\lambda(z-x)}\right)\, dx \\ &= \int_0^z \lambda^2 e^{-\lambda x} e^{-\lambda z} e^{\lambda x}\, dx \\ &= \lambda^2 e^{-\lambda z} \int_0^z 1\, dx \quad \text{(Since terms with } z \text{ are constant w.r.t } x) \\ &= \lambda^2 e^{-\lambda z} [x]_0^z \\ &= \lambda^2 e^{-\lambda z}(z - 0) \\ &= \lambda^2 z e^{-\lambda z} \end{aligned}$$

Conclusion:

Combining both cases, the PDF of Z is:

$$f_Z(z) = \begin{cases} \lambda^2 z e^{-\lambda z} & z \geq 0 \\ 0 & z < 0 \end{cases}$$

- *Note 1: This is the PDF of the Erlang(2, λ) distribution or Gamma(2, λ).*
- **Note 2: Several valid approaches begin from first principles and inherently derive the convolution formula as the problem is solved. Marks have been awarded for all correct solutions.**

Part 2: Mean of Z

The mean is defined as $E[Z] = \int_{-\infty}^{\infty} z f_Z(z)\, dz$. Using our derived PDF (where $f_Z(z) = 0$ for $z < 0$):

$$E[Z] = \int_0^\infty z(\lambda^2 z e^{-\lambda z})\, dz = \lambda^2 \int_0^\infty z^2 e^{-\lambda z}\, dz$$

Method 1: Integration by Parts

Let $u = z^2$ and $dv = e^{-\lambda z} dz$. Then $du = 2z dz$ and $v = -\frac{1}{\lambda} e^{-\lambda z}$.

$$\begin{aligned} \int_0^\infty z^2 e^{-\lambda z}\, dz &= \left[-\frac{z^2}{\lambda} e^{-\lambda z} \right]_0^\infty - \int_0^\infty \left(-\frac{1}{\lambda} e^{-\lambda z} \right) 2z\, dz \\ &= 0 + \frac{2}{\lambda} \int_0^\infty z e^{-\lambda z}\, dz \end{aligned}$$

We recognize $\int_0^\infty z e^{-\lambda z} dz = \frac{1}{\lambda^2}$ (mean of standard exponential is $1/\lambda$, so integral without λ pre-factor is $1/\lambda^2$).

$$E[Z] = \lambda^2 \left(\frac{2}{\lambda} \cdot \frac{1}{\lambda^2} \right) = \frac{2}{\lambda}$$

Method 2: Gamma Function Substitution

Alternatively, in the integral $I = \lambda^2 \int_0^\infty z^2 e^{-\lambda z}\, dz$, let $u = \lambda z \implies dz = du/\lambda$.

$$E[Z] = \lambda^2 \int_0^\infty \left(\frac{u}{\lambda} \right)^2 e^{-u} \frac{du}{\lambda} = \frac{1}{\lambda} \int_0^\infty u^{3-1} e^{-u} du = \frac{1}{\lambda} \Gamma(3)$$

Using $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$:

$$E[Z] = \frac{1}{\lambda}(2!) = \frac{2}{\lambda}$$

Or you could have recognized that the integral is equivalent to $\lambda . E[X^2]$ where $X \sim Exp(\lambda)$

Final Answer:

The PDF is $f_Z(z) = \lambda^2 z e^{-\lambda z}$ for $z \geq 0$ (and 0 otherwise). The Mean is $E[Z] = \frac{2}{\lambda}$.

Q4. Mixture Distribution

Let Z be a discrete random variable taking values in $\{1, \dots, n\}$ with

$$\mathbb{P}(Z = i) = p_i, \qquad i = 1, \dots, n.$$

Conditioned on $Z = i$, let X be distributed as Y_i. Thus we define

$$X = Y_Z.$$

(a) Find $\mathbb{E}[X]$.

(b) Also prove that

$$f_X(x) = \sum_{i=1}^{n} p_i \, f_{Y_i}(x).$$

Part (a)

Using the law of total expectation,

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X \mid Z]\big].$$

Given $Z = i$, the random variable X has the same distribution as Y_i, so

$$\mathbb{E}[X \mid Z = i] = \mathbb{E}[Y_i].$$

Taking expectation over Z gives

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{P}(Z = i) \, \mathbb{E}[X \mid Z = i] = \sum_{i=1}^{n} \mathbb{P}(Z = i) \, \mathbb{E}[Y_i] = \sum_{i=1}^{n} p_i \, \mathbb{E}[Y_i].$$

Thus,

$$\boxed{\mathbb{E}[X] = \sum_{i=1}^{n} p_i \, \mathbb{E}[Y_i]}.$$

Part (b)

Apply the law of total probability at the point x:

$$f_X(x) = \sum_{i=1}^{n} \mathbb{P}(Z = i) \, f_{X|Z=i}(x).$$

Since $X \mid \{Z = i\}$ has the same distribution as Y_i, we have $f_{X|Z=i}(x) = f_{Y_i}(x)$, and therefore

$$\boxed{f_X(x) = \sum_{i=1}^{n} p_i \, f_{Y_i}(x)}.$$

Q5. MLE for Uniform Distribution

Question: Let $D = \{x_1, ...x_n\}$ denote i.i.d. samples from a uniform random variable $U[a, b]$ where a and b are unknown. Find an MLE estimate for the unknown parameters a and b.

Solution

The pdf of a $U[a, b]$ random variable is given by

$$f_U(u) = \begin{cases} \frac{1}{b-a} & a \leq u \leq b \\ 0 & \text{o.w} \end{cases}$$

The likelihood of D is defined as

$$L(x_1, x_2, ...x_n; a, b) = f_{U_1,...,U_n}(x_1, ...x_n; a, b) = \prod f_{U_i}(x_i; a, b)$$

as the samples are i.i.d.

From the pdf, it is clear that $L \neq 0$ if $a \leq x_i \leq b \; \forall i = 1...n$. So $L \neq 0$ iff

$$a \leq \min_i(x_i) \quad \text{and} \quad b \geq \max_i(x_i)$$

$$L(x_1, ...x_n; a, b) = \begin{cases} \frac{1}{(b-a)^n}, & a \leq \min_i(x_i), \; b \geq \max_i(x_i) \\ 0 & \text{o.w} \end{cases}$$

The log likelihood is

$$\log L(x_1, ...x_n; a, b) = \begin{cases} -n \log(b - a), & a \leq \min_i(x_i), \; b \geq \max_i(x_i) \\ -\infty & \text{o.w} \end{cases}$$

The MLE estimate for a is given by

$$\hat{a}_{ML} = \arg\max_a \log L(x_1, ...x_n; a, b)$$

To find the maxima, we take the derivative w.r.t a

$$\frac{\partial \log L}{\partial a} = \frac{n}{b - a}$$

for $a \leq \min_i(x_i), \; b \geq \max_i(x_i)$.

The derivative w.r.t a is monotonically increasing in the region $a \leq \min_i(x_i)$, so to maximize the likelihood we take the maximum value a can take in the region which is given by

$$\hat{a}_{ML} = \min_i x_i$$

Similarly, the MLE estimate for b is given by

$$\hat{b}_{ML} = \arg\max_{b} \log L(x_1, ...x_n; a, b)$$

To find the maxima, we take the derivative w.r.t b

$$\frac{\partial \log L}{\partial b} = -\frac{n}{b-a}$$

for $a \leq \min_i(x_i),\ b \geq \max_i(x_i)$.

The derivative w.r.t b is monotonically decreasing in the region $b \geq \max_i(x_i)$, so to maximize the likelihood we take the minimum value b can take in the region which is given by

$$\hat{b}_{ML} = \max_{i} x_i$$

Q6. MLE for Poisson Distribution

Part (a): Finding the MLE

Let $X \sim \text{Poisson}(\gamma)$ and let $\mathcal{D} = \{x_1, \ldots, x_n\}$ be i.i.d. samples.

The pmf is

$$P(X = x_i) = \frac{\gamma^{x_i} e^{-\gamma}}{x_i!}.$$

Thus the likelihood is

$$L(\gamma) = \prod_{i=1}^{n} \frac{\gamma^{x_i} e^{-\gamma}}{x_i!} = \frac{\gamma^{\sum_{i=1}^{n} x_i} e^{-n\gamma}}{\prod_{i=1}^{n} x_i!}.$$

Taking the log-likelihood:

$$\ell(\gamma) = \log L(\gamma) = \sum_{i=1}^{n} x_i \log \gamma - n\gamma - \sum_{i=1}^{n} \log(x_i!).$$

Differentiate w.r.t. γ and set to zero:

$$\frac{d\ell}{d\gamma} = \frac{\sum_{i=1}^{n} x_i}{\gamma} - n = 0.$$

Solving gives the MLE:

$$\hat{\gamma}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{X}.$$

Finding the Second Derivative of ℓ wrt γ,

$$\frac{d^2\ell}{d\gamma^2} = -\frac{\sum_{i=1}^{n} x_i}{\gamma^2}$$

Since the samples are from a Poisson distribution, each sample is positive, making the numerator positive. Hence the overall term is still negative, satisfying the maximality of the estimator.

Part (b): Bias and MSE

Since $X \sim \text{Poisson}(\gamma)$, we have

$$\mathbb{E}[X] = \gamma, \qquad \text{Var}(X) = \gamma.$$

Thus for the estimator $\hat{\gamma} = \bar{X}$,

$$\mathbb{E}[\hat{\gamma}] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(x_i) = \frac{n\gamma}{n} = \gamma,$$

Hence, the Bias $= 0$.

$$\mathrm{Var}(\hat{\gamma}) = \mathrm{Var}(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(x_i) = \frac{n\gamma}{n^2} = \frac{\gamma}{n}.$$

Hence the mean squared error (MSE) is

$$\mathrm{MSE}(\hat{\gamma}) = \mathrm{Var}(\hat{\gamma}) + \mathrm{Bias}^2 = \frac{\gamma}{n}.$$

Q7. Convergence in Distribution and Probability

Given: Let $X_n \sim \text{Uniform}(5 - \frac{1}{n}, 5 + \frac{1}{n})$.

(a) Show that $X_n \xrightarrow{d} 5$.

To show convergence in distribution ($X_n \xrightarrow{d} X$), we must show that the Cumulative Distribution Function (CDF) of X_n, denoted $F_n(x)$, converges to the CDF of the limiting random variable X at all points where $F_X(x)$ is continuous.

Step 1: Identify the limiting CDF

The limiting variable is the constant $X = 5$. The CDF of a constant random variable is a step function:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 5 \\ 1 & \text{if } x \geq 5 \end{cases}$$

The function $F_X(x)$ is continuous everywhere except at $x = 5$. Therefore, we need to show that $\lim_{n\to\infty} F_n(x) = F_X(x)$ for all $x \neq 5$.

Step 2: Define the CDF of X_n

The random variable X_n follows a uniform distribution on $[a_n, b_n]$ where $a_n = 5 - \frac{1}{n}$ and $b_n = 5 + \frac{1}{n}$. The length of the interval is $b_n - a_n = \frac{2}{n}$. The CDF is given by:

$$F_n(x) = \begin{cases} 0 & \text{if } x < 5 - \frac{1}{n} \\ \frac{x-(5-\frac{1}{n})}{2/n} & \text{if } 5 - \frac{1}{n} \leq x \leq 5 + \frac{1}{n} \\ 1 & \text{if } x > 5 + \frac{1}{n} \end{cases}$$

Step 3: Evaluate the limit as $n \to \infty$

- **Case 1:** $x < 5$

  Since $x < 5$, let $\delta = 5 - x > 0$. We can choose an integer N such that $\frac{1}{N} < \delta$. For all $n > N$, we have $\frac{1}{n} < 5 - x$, which implies $x < 5 - \frac{1}{n}$. In this region, $F_n(x) = 0$.

  $$\lim_{n\to\infty} F_n(x) = 0 = F_X(x).$$

- **Case 2:** $x > 5$

  Since $x > 5$, let $\delta = x - 5 > 0$. We can choose an integer N such that $\frac{1}{N} < \delta$. For all $n > N$, we have $\frac{1}{n} < x - 5$, which implies $x > 5 + \frac{1}{n}$. In this region, $F_n(x) = 1$.

  $$\lim_{n\to\infty} F_n(x) = 1 = F_X(x).$$

Conclusion:

Since $\lim_{n\to\infty} F_n(x) = F_X(x)$ for all continuity points of F_X (i.e., for all $x \neq 5$), we conclude that:

$$X_n \xrightarrow{d} 5.$$

(b) Compute $P(|X_n - 5| > \varepsilon)$ explicitly and show that it converges to 0 as $n \to \infty$.

We fix an arbitrary $\varepsilon > 0$. We want to compute the probability that X_n falls outside the interval $(5 - \varepsilon, 5 + \varepsilon)$.

The probability density function (PDF) of X_n is:

$$f_n(x) = \frac{1}{2/n} = \frac{n}{2}, \quad \text{for } 5 - \frac{1}{n} \leq x \leq 5 + \frac{1}{n},$$

and 0 otherwise. The maximum distance of X_n from 5 is $\frac{1}{n}$, that is:

$$|X_n - 5| \leq \frac{1}{n}$$

We compare $\frac{1}{n}$ with ε:

- **Case 1: Large n (specifically $n \geq \frac{1}{\varepsilon}$)**
  If $n \geq \frac{1}{\varepsilon}$, then $\frac{1}{n} \leq \varepsilon$. The entire support interval $[5 - \frac{1}{n}, 5 + \frac{1}{n}]$ is contained within $[5 - \varepsilon, 5 + \varepsilon]$. Therefore, the probability of X_n falling outside this range is 0.

  $$P(|X_n - 5| > \varepsilon) = 0.$$

- **Case 2: Small n (specifically $n < \frac{1}{\varepsilon}$)**
  If $n < \frac{1}{\varepsilon}$, then $\frac{1}{n} > \varepsilon$. Hence $5 - \frac{1}{n} < 5 - \varepsilon$ and $5 + \frac{1}{n} > 5 + \varepsilon$.
  The regions where $|X_n - 5| > \varepsilon$ are:

  $$\left[5 - \frac{1}{n}, 5 - \varepsilon\right) \cup \left(5 + \varepsilon, 5 + \frac{1}{n}\right]$$

  The total length of these two regions is:

  $$2 \times \left((5 + \frac{1}{n}) - (5 + \varepsilon)\right) = 2\left(\frac{1}{n} - \varepsilon\right).$$

  Hence

  $$P(|X_n - 5| > \varepsilon) = \frac{n}{2} \times 2\left(\frac{1}{n} - \varepsilon\right) = n\left(\frac{1}{n} - \varepsilon\right) = 1 - n\varepsilon.$$

Explicit Formula:

$$P(|X_n - 5| > \varepsilon) = \begin{cases} 1 - n\varepsilon & \text{if } n < \frac{1}{\varepsilon} \\ 0 & \text{if } n \geq \frac{1}{\varepsilon} \end{cases}$$

Convergence:

To find the limit as $n \to \infty$, we observe that for every $\varepsilon > 0$ there exists an integer $N > \frac{1}{\varepsilon}$. For all $n > N$, we are in "Case 1" above, where the probability is exactly 0.

$$\lim_{n \to \infty} P(|X_n - 5| > \varepsilon) = 0.$$

This confirms that X_n converges to 5 in probability ($X_n \xrightarrow{p} 5$).

Q8. Markov Chain - Hitting Probabilities

Problem: Consider a Markov chain with state space $\mathcal{S} = \{1, 2, 3\}$ and transition matrix P:

$$P = \begin{bmatrix} 0.1 & 0.9 & 0 \\ 0.2 & 0.6 & 0.2 \\ 0 & 0 & 1 \end{bmatrix}$$

Let F_{ij} denote the probability of the Markov chain ever returning to (or hitting) state i having started in state j. Calculate F_{ii} for $i = 1, 2, 3$ and classify the states as transient or recurrent.

State Transition Diagram

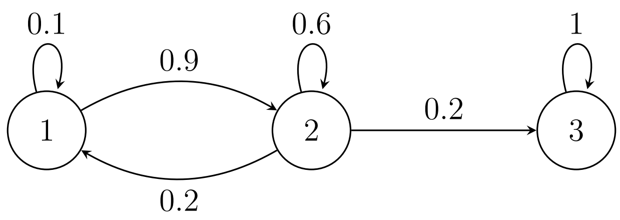

Figure 1: State Transition Diagram

0. General Formula Used

The probability of ever reaching a general state p, having started in the state q could be expressed in the following form:

$$F_{pq} = P_{qp} + \sum_{k \neq q} P_{pk} F_{qk}$$

1. Analysis of State 3 (F_{33})

State 3 is an absorbing state because $P_{33} = 1$. Once the system enters state 3, it cannot leave.

$$F_{33} = P_{33} + \sum_{k \neq 3} P_{3k} F_{3k} = 1 + 0 = 1$$

Conclusion: Since $F_{33} = 1$, State 3 is **Recurrent**.

2. Analysis of State 1 (F_{11})

We calculate the probability of returning to state 1 starting from state 1. We use First Step Analysis.

$$F_{11} = P_{11}.1 + P_{12}.F_{12} + P_{13}.F_{13}$$

Substituting values from matrix P (note that $P_{13} = 0$):

$$F_{11} = 0.1 + 0.9F_{12} \tag{1}$$

Now we must calculate F_{12} (probability of hitting 1 starting from 2):

$$F_{12} = P_{21}.1 + P_{22}.F_{12} + P_{23}.F_{13}$$

Note that $F_{13} = 0$ because if we enter state 3, we are absorbed and never reach state 1.

$$\begin{aligned} F_{12} &= 0.2 + 0.6F_{12} + 0.2(0) \\ F_{12} &= 0.2 + 0.6F_{12} \\ 0.4F_{12} &= 0.2 \\ F_{12} &= \frac{0.2}{0.4} = 0.5 \end{aligned}$$

Substitute $F_{12} = 0.5$ back into Equation (1):

$$F_{11} = 0.1 + 0.9(0.5) = 0.1 + 0.45 = 0.55$$

Conclusion: Since $F_{11} = 0.55 < 1$, State 1 is **Transient**.

3. Analysis of State 2 (F_{22})

We calculate the probability of returning to state 2 starting from state 2.

$$F_{22} = P_{22}.1 + P_{21}.F_{21} + P_{23}.F_{23}$$

Since state 3 is absorbing, $F_{23} = 0$ (cannot return to 2 from 3).

$$F_{22} = 0.6 + 0.2F_{21} \tag{2}$$

Now we must calculate F_{21} (probability of hitting 2 starting from 1):

$$F_{21} = P_{12}.1 + P_{11}.F_{21} + P_{13}.0$$

$$\begin{aligned} F_{21} &= 0.9 + 0.1F_{21} \\ 0.9F_{21} &= 0.9 \\ F_{21} &= 1 \end{aligned}$$

Substitute $F_{21} = 1$ back into Equation (2):

$$F_{22} = 0.6 + 0.2(1) = 0.8$$

Conclusion: Since $F_{22} = 0.8 < 1$, State 2 is **Transient**.

Q9. Bayesian Estimation

(a) Likelihood for the data

We first obtain the expression for likelihood. Let X_i be the random variable corresponding to sample x_i:

$$\begin{aligned}
f_{X_1,\cdots,X_n|\Lambda}(x_1,\cdots,x_n|\lambda^*) &= \prod_{i=1}^{n} f_{X_i|\Lambda}(x_i|\lambda^*) \\
&= \prod_{i=1}^{n} \lambda^* e^{-\lambda^* x_i} \\
&= (\lambda^*)^n e^{-\lambda^* \sum_{i=1}^{n} x_i}.
\end{aligned}$$

The posterior distribution can be found using Bayes' rule:

$$\begin{aligned}
f_{\Lambda|X_1,\cdots,X_n}(\lambda^*|x_1,\cdots,x_n) &= \frac{f_{X_1,\cdots,X_n|\Lambda}(x_1,\cdots,x_n|\lambda^*)f_\Lambda(\lambda^*)}{f_{X_1,\cdots,X_n}(x_1,\cdots,x_n)} \\
&= \frac{(\lambda^*)^n e^{-\lambda^* \sum_{i=1}^{n} x_i} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}(\lambda^*)^{\alpha-1}e^{-\beta\lambda^*}}{\int_0^\infty f_{X_1,\cdots,X_n|\Lambda}(x_1,\cdots,x_n|\lambda)f_\Lambda(\lambda)\,d\lambda} \\
&= \frac{\frac{\beta^\alpha}{\Gamma(\alpha)}(\lambda^*)^{\alpha-1+n}e^{-(\beta+\sum_{i=1}^{n} x_i)\lambda^*}}{\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1+n}e^{-(\beta+\sum_{i=1}^{n} x_i)\lambda}\,d\lambda} \\
&= \frac{(\lambda^*)^{\alpha-1+n}e^{-(\beta+\sum_{i=1}^{n} x_i)\lambda^*}}{\int_0^\infty \lambda^{\alpha-1+n}e^{-(\beta+\sum_{i=1}^{n} x_i)\lambda}\,d\lambda}.
\end{aligned}$$

Let $(\beta + \sum_{i=1}^{n} x_i)\lambda = t$:

$$\begin{aligned}
f_{\Lambda|X_1,\cdots,X_n}(\lambda^*|x_1,\cdots,x_n) &= \frac{(\lambda^*)^{\alpha-1+n}e^{-(\beta+\sum_{i=1}^{n} x_i)\lambda^*}}{\int_0^\infty \left(\frac{t}{\beta+\sum_{i=1}^{n} x_i}\right)^{\alpha-1+n} e^{-t}\frac{1}{\beta+\sum_{i=1}^{n} x_i}\,dt} \\
&= \frac{(\lambda^*)^{\alpha-1+n}e^{-(\beta+\sum_{i=1}^{n} x_i)\lambda^*}}{\frac{1}{(\beta+\sum_{i=1}^{n} x_i)^{\alpha+n}}\int_0^\infty t^{\alpha-1+n}e^{-t}\,dt}.
\end{aligned}$$

We know that $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt$:

$$f_{\Lambda|X_1,\cdots,X_n}(\lambda^*|x_1,\cdots,x_n) = \frac{(\beta+\sum_{i=1}^{n} x_i)^{\alpha+n}}{\Gamma(\alpha+n)}(\lambda^*)^{\alpha-1+n}e^{-(\beta+\sum_{i=1}^{n} x_i)\lambda^*}.$$

Which gives us $\Lambda|X_1,\cdots,X_n \sim \text{Gamma}(\alpha+n, \beta+\sum_{i=1}^{n} x_i)$.

(b) MAP Estimate

To find the MAP estimate of λ^*:

$$\lambda_{MAP} = \arg\max_{\lambda^*} f_{\Lambda|X_1,\cdots,X_n}(\lambda^*|x_1,\cdots,x_n).$$

Ignoring the variables independent of λ^*:

$$\lambda_{MAP} = \arg\max_{\lambda^*} (\lambda^*)^{\alpha-1+n} e^{-(\beta+\sum_{i=1}^{n} x_i)\lambda^*}.$$

Differentiating the expression with respect to λ^* and setting it to zero to obtain the maximum point:

$$(\alpha - 1 + n)(\lambda^*)^{\alpha-2+n} e^{-(\beta+\sum_{i=1}^{n} x_i)\lambda^*} - (\beta + \sum_{i=1}^{n} x_i)(\lambda^*)^{\alpha-1+n} e^{-(\beta+\sum_{i=1}^{n} x_i)\lambda^*} = 0$$

$$\lambda^* = \frac{\alpha - 1 + n}{\beta + \sum_{i=1}^{n} x_i}.$$

Thus our MAP estimate for λ^* is

$$\boxed{\lambda_{MAP} = \frac{\alpha - 1 + n}{\beta + \sum_{i=1}^{n} x_i}}.$$

Q10(a). Stationary vs Limiting Distribution

Question: Give an example of a 3-state Markov chain that has a stationary distribution but does not have a limiting distribution. Obtain its stationary distribution. (Hint: You saw a two-state example in class.)

Any correct example of a 3-state Markov chain that satisfies the conditions is acceptable, as long as it:

- clearly shows the transition structure,
- correctly finds a stationary distribution,
- and argues why there is no limiting distribution.

Example solution:

Consider the 3-state Markov chain with transition matrix

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

This chain cycles through the states in order: $1 \to 2 \to 3 \to 1$, repeatedly.

Stationary distribution:

Let $\pi = (\pi_1, \pi_2, \pi_3)$. We need

$$\pi P = \pi, \qquad \pi_1 + \pi_2 + \pi_3 = 1.$$

Computing,

$$\pi P = (\pi_3, \pi_1, \pi_2),$$

and setting $\pi P = \pi$ gives

$$\pi_1 = \pi_2 = \pi_3.$$

Since they must sum to 1,

$$\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right).$$

Why there is no limiting distribution:

Starting from $(1, 0, 0)$,

$$(1,0,0)P^n = \begin{cases} (1,0,0), & n \equiv 0 \pmod 3, \\ (0,1,0), & n \equiv 1 \pmod 3, \\ (0,0,1), & n \equiv 2 \pmod 3. \end{cases}$$

The distribution keeps rotating among the states and never converges to a single fixed distribution.

(b) Define the Mean square error of an estimator. Explain the bias-variance tradeoff.

Solution

1. Definition of Mean Square Error (MSE)
Let $\hat{\Theta}_n$ be a point estimator for an unknown parameter θ^*. The Mean Square Error (MSE) of the estimator is defined as the expected value of the squared difference between the estimator and the true parameter:

$$MSE(\hat{\Theta}_n) = \mathbb{E}[(\hat{\Theta}_n - \theta^*)^2]$$

2. Bias-Variance Tradeoff
The MSE can be decomposed into two components: the variance of the estimator and the square of its bias. This decomposition represents the bias-variance tradeoff.

Derivation:
We expand the MSE term by adding and subtracting the expected value of the estimator, $\mathbb{E}[\hat{\Theta}_n]$ (analogous to the expansion used in the total variance identity in L25, Source 31):

$$\begin{aligned} MSE(\hat{\Theta}_n) &= \mathbb{E}[(\hat{\Theta}_n - \theta^*)^2] \\ &= \mathbb{E}[(\hat{\Theta}_n - \mathbb{E}[\hat{\Theta}_n] + \mathbb{E}[\hat{\Theta}_n] - \theta^*)^2] \end{aligned}$$

Let $A = \hat{\Theta}_n - \mathbb{E}[\hat{\Theta}_n]$ and $B = \mathbb{E}[\hat{\Theta}_n] - \theta^*$. Note that B is a constant (the Bias). Expanding the square:

$$= \mathbb{E}[(\hat{\Theta}_n - \mathbb{E}[\hat{\Theta}_n])^2] + (\mathbb{E}[\hat{\Theta}_n] - \theta^*)^2 + 2(\mathbb{E}[\hat{\Theta}_n] - \theta^*)\mathbb{E}[\hat{\Theta}_n - \mathbb{E}[\hat{\Theta}_n]]$$

The last term vanishes because $\mathbb{E}[\hat{\Theta}_n - \mathbb{E}[\hat{\Theta}_n]] = \mathbb{E}[\hat{\Theta}_n] - \mathbb{E}[\hat{\Theta}_n] = 0$.

Thus, we are left with:

$$MSE(\hat{\Theta}_n) = \underbrace{\mathbb{E}[(\hat{\Theta}_n - \mathbb{E}[\hat{\Theta}_n])^2]}_{\text{Variance}} + \underbrace{(\mathbb{E}[\hat{\Theta}_n] - \theta^*)^2}_{\text{Bias}^2}$$

$$MSE(\hat{\Theta}_n) = Var(\hat{\Theta}_n) + [Bias(\hat{\Theta}_n)]^2$$

Explanation:
This equation shows that the error in estimation (MSE) comes from two sources:

- **Variance:** How spread out the estimator is around its expected value.
- **Bias:** The difference between the estimator's expected value and the true parameter θ^* ($Bias = \mathbb{E}[\hat{\Theta}] - \theta^*$).

The "tradeoff" implies that minimizing MSE often requires balancing these two; reducing bias might increase variance, and vice versa.