Probability and Statistics: MA6.101

Endsem Solutions

5 Mark Questions

Q1: Let $Z = X_1 + X_2 + \cdots + X_N$, where X_i are i.i.d. random variables and N is a positive discrete random variable. Prove that:

$$M_Z(t) = M_N(\log M_X(t)).$$

A: The moment generating function of Z is given by:

$$M_Z(t) = E[e^{tZ}]$$

Conditioning on N, the number of terms, we have:

$$M_Z(t) = E_N\left[E\left[e^{tz} \mid N\right]\right].$$

Expanding Z we have

$$e^{tz} = e^{t(X_1+X_2+\cdots+X_N)} = \prod_{i=1}^{N} e^{tX_i}$$

So the expression is

$$M_Z(t) = E_N\left[E\left[\prod_{i=1}^{N} e^{tX_i} \mid N\right]\right].$$

Since the X_i are independent, we can use $E[XY] = E[X] \times E[Y]$:

$$M_Z(t) = E_N\left[\prod_{i=1}^{N} E\left[e^{tX_i} \mid N\right]\right].$$

Now X_i are i.i.d so $X_i = X_1$ and it is independant of N

$$M_Z(t) = E_N\left[(E\left[e^{tX_1}\right])^N\right].$$

Now $E[e^{tX_1}] = M_X(t)$ by definition of moment generating function. So:

$$M_Z(t) = E_N\left[(M_X(t))^N\right].$$

Now let $u = M_X(t)$. Then:

$$M_Z(t) = E_N[u^N].$$

Now let's think of MGF of N:

$$M_N(t) = E[e^{tN}]$$

if $t = log\ u$

$$M_N(log(u)) = E[e^{log(u)\times N}] = E[e^{log(u^N)}] = E[u^N]$$

Subsituting this in our answer

$$M_Z(t) = M_N(log(u)) = M_N(log(M_X(t))).$$

Hence proved

Q2: Let $Y = e^X$, where $X \sim \mathcal{N}(\mu, \sigma^2)$. Obtain the pdf of Y.

A: The cumulative distribution function (CDF) of Y is given by:

$$F_Y(y) = P(Y \le y) = P(e^X \le y).$$

Taking the natural logarithm, this is equivalent to:

$$F_Y(y) = P(X \le \log y).$$

Note: $y > 0$ for log to be defined

$$F_Y(y) = F_X(\log y).$$

Differentiating $F_Y(y)$ with respect to y, we obtain the PDF of Y:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}\left[F_X(\log y)\right].$$

From the chain rule:

$$f_Y(y) = f_X(\log y) \cdot \frac{d}{dy}(\log y).$$

The derivative of $\log y$ is $\frac{1}{y}$. Substituting:

$$f_Y(y) = f_X(\log y) \cdot \frac{1}{y}.$$

The PDF of X is:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Substituting $x = \log y$, we get:

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log y-\mu)^2}{2\sigma^2}\right) \cdot \frac{1}{y}.$$

Simplifying:

$$f_Y(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log y-\mu)^2}{2\sigma^2}\right), \quad y > 0.$$

8 Mark Questions

Q1: Let $D = \{x_1, x_2, \ldots, x_n\}$ denote i.i.d. samples from a Poisson random variable with unknown parameter γ.

(a) Find the Maximum Likelihood Estimate (MLE) for the unknown parameter γ. (5 marks)

(b) Determine the Mean Squared Error (MSE) of the estimate. (3 marks)

A:

(a) MLE for γ

The probability mass function (PMF) of a Poisson random variable is given by:

$$P(X = x) = \frac{\gamma^x e^{-\gamma}}{x!}, \quad x = 0, 1, 2, \ldots$$

For the i.i.d. samples $D = \{x_1, x_2, \ldots, x_n\}$, the likelihood function is:

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

Taking the natural logarithm of the likelihood function, the log-likelihood is:

$$\ell(D; \gamma) = \sum_{i=1}^{n} \left[x_i \ln(\gamma) - \gamma - \ln(x_i!)\right].$$

To find the MLE, we differentiate $\ell(\gamma; D)$ with respect to γ and set it equal to zero:

$$\frac{\partial \ell(D; \gamma)}{\partial \gamma} = \sum_{i=1}^{n} \frac{x_i}{\gamma} - n = 0.$$

Simplifying:

$$\sum_{i=1}^{n} x_i = n\gamma.$$

Solving for γ, the MLE and corresponding estimator $\hat{\gamma}$ is:

$$\hat{\gamma} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x},$$

where $\bar{x}$ is the sample mean.

Thus MLE estimate can be written as:-

$$\hat{\gamma} = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X},$$

(b) Mean Squared Error (MSE) of $\hat{\gamma}$

Let the sample mean $\hat{\gamma}$ be written as:

$$\hat{\gamma} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The expectation of $\hat{\gamma}$ is:

$$\mathbb{E}[\hat{\gamma}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i].$$

Since $\mathbb{E}[X_i] = \gamma$ for a Poisson random variable:

$$\mathbb{E}[\hat{\gamma}] = \frac{1}{n} \cdot n \cdot \gamma = \gamma.$$

The variance of $\hat{\gamma}$ is:

$$\mathrm{Var}(\hat{\gamma}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Since Var$(X_i) = \gamma$ for a Poisson random variable:

$$\mathrm{Var}(\hat{\gamma}) = \frac{1}{n^2} \cdot n \cdot \gamma = \frac{\gamma}{n}.$$

The Mean Squared Error (MSE) of $\hat{\gamma}$ is:

$$\mathrm{MSE}(\hat{\gamma}) = \mathrm{Var}(\hat{\gamma}) + \left(\mathbb{E}[\hat{\gamma}] - \gamma\right)^2.$$

Since $\mathbb{E}[\hat{\gamma}] = \gamma$, the bias is zero:

$$\mathrm{MSE}(\hat{\gamma}) = \mathrm{Var}(\hat{\gamma}) = \frac{\gamma}{n}.$$

Q2: Consider a Gaussian random variable X with a known mean μ but an unknown variance σ^2. Suppose you observe k iid samples from this random variable, denoted by $D = \{x_1, x_2, \ldots, x_k\}$.

(a) Find the MLE for the unknown variance σ^2. (5 marks)

(b) Is the MLE estimate biased? Justify your answer. (3 marks)

A:

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ be a Gaussian random variable with known mean μ and unknown variance σ^2. The probability density function (PDF) of X is given by:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Given k independent and identically distributed (iid) samples $D = \{x_1, x_2, \ldots, x_k\}$, the likelihood function is expressed as:

$$L(\sigma^2 \mid D) = \prod_{i=1}^{k} f(x_i \mid \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^k \exp\left(-\frac{\sum_{i=1}^{k}(x_i-\mu)^2}{2\sigma^2}\right).$$

To determine the maximum likelihood estimate (MLE) of σ^2, denoted as $\hat{\sigma}^2$, we maximize the likelihood function. Since the logarithm is a monotonically increasing function, we equivalently maximize the log-likelihood function:

$$\ell(\sigma^2 \mid D) = \ln L(\sigma^2 \mid D) = -\frac{k}{2}\ln(2\pi) - \frac{k}{2}\ln(\sigma^2) - \frac{\sum_{i=1}^{k}(x_i-\mu)^2}{2\sigma^2}.$$

To find the MLE, we differentiate $\ell(\sigma^2 \mid D)$ with respect to σ^2 and set the derivative to zero:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{k}{2\sigma^2} + \frac{\sum_{i=1}^{k}(x_i-\mu)^2}{2\sigma^4} = 0.$$

Simplifying, we find:

$$\frac{k}{2\sigma^2} = \frac{\sum_{i=1}^{k}(x_i-\mu)^2}{2\sigma^4}.$$

Multiply through by $2\sigma^4$:

$$k\sigma^2 = \sum_{i=1}^{k}(x_i-\mu)^2.$$

Solve for σ^2:

$$\hat{\sigma}^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i-\mu)^2.$$

Thus, the MLE estimate for the variance is:

$$\boxed{\widehat{\sigma^2} = \frac{1}{k}\sum_{i=1}^{k}(x_i-\mu)^2}$$

To determine whether the MLE estimate is biased, we need to compare $E[\widehat{\sigma^2}]$ with the true variance σ^2.

$$\hat{\sigma}^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i-\mu)^2$$

Now, we know that:

$$E[X] = \mu$$

So,

$$E[\hat{\sigma^2}] = \frac{1}{k}\sum_{i=1}^{k}\left(E[x_i^2] - 2\mu E[x_i] + \mu^2\right)$$

$$= \frac{1}{k}\sum_{i=1}^{k}\left(\sigma^2 + \mu^2 - 2\mu^2 + \mu^2\right)$$

$$= \frac{1}{k}\left(k\sigma^2\right) = \sigma^2$$

$$Bias[\hat{\sigma}^2] = E[\hat{\sigma}^2] - \sigma$$
$$Bias[\hat{\sigma}] = 0$$

Therefore, the estimator $\hat{\sigma}$ is an unbiased estimator.

Q3: Consider a sequence $\{X_n, n = 1, 2, 3, \dots\}$ such that

$$X_n = \begin{cases} n, & \text{with probability } \frac{1}{n^2} \\ 0, & \text{with probability } 1 - \frac{1}{n^2} \end{cases}$$

(a) Show that $X_n \xrightarrow{p} 0$ (Convergence in probability to 0). (4 marks)

(b) Show that $X_n \xrightarrow{\text{a.s.}} 0$ (Almost sure convergence to 0). (4 marks)

A:

(a) Convergence in probability: To show $X_n \xrightarrow{p} 0$, we need to prove:

$$\forall\epsilon > 0, \quad \lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$

$$\forall\epsilon > 0, \quad \lim_{n\to\infty} P(|X_n| > \epsilon) = 0.$$

We have:

$$P(|X_n| > \epsilon) = P(X_n = n) = \frac{1}{n^2}.$$

Hence:

$$\lim_{n\to\infty} P(|X_n| > \epsilon) = \lim_{n\to\infty} \frac{1}{n^2} = 0.$$

Thus, $X_n \xrightarrow{p} 0$.

(b) Almost sure convergence:

By the Borel-Cantelli lemma, if $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$, then $X_n \xrightarrow{\text{a.s.}} X$

$$P\left(|X_n - X| > \epsilon\right)$$

$$= P\left(|X_n| > \epsilon\right)$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

$$= finite$$

This series converges because it is a p-series(type of infinite series that is written as $\sum(1/n^p)$) with $p = 2 > 1$. Therefore:

Hence, $X_n = 0$ for all sufficiently large n with probability 1. Thus, $X_n \xrightarrow{\text{a.s.}} 0$.

Q4: Given a Markov coin with the following transition probability matrix P and initial distribution $\mu = [0.1, 0.9]$, use the following 4 independent $Uniform[0, 1]$ samples $\{0.3, 0.7, 0.23, 0.97\}$ to obtain/generate 4 successive toss outcomes of the Markov coin. (Hint: The first toss is to be sampled from the initial distribution.)

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}$$

A:

We will simulate the successive toss outcomes of the Markov coin using the given transition probability matrix P, initial distribution $\mu = [0.1, 0.9]$, and uniform random samples {0.3, 0.7, 0.23, 0.97}.

Step 1: The first toss The first toss is sampled from the initial distribution μ. Since the first sample is 0.3:

$$\text{Cumulative probabilities for } \mu : \quad [0.1, 1.0]$$

Since $0.3 \leq 1.0$, the first state is 2 (corresponding to μ_2).

Step 2: Successive tosses For each successive toss, the state is determined using the transition probability matrix P and the next sample: so when i get tail as state for first toss, then distribution changes to [0.4 0.6]

- **Second toss:** Current state = 2, transition probabilities = $[0.4, 0.6]$.

  $$\text{Cumulative probabilities: } [0.4, 1.0]$$

  Sample = 0.7. Since $0.7 \leq 1.0$, the next state is 2. so when i get tail as state for second toss, then distribution changes to [0.4 0.6]

- **Third toss:** Current state = 2, transition probabilities = $[0.4, 0.6]$.

  $$\text{Cumulative probabilities: } [0.4, 1.0]$$

  Sample = 0.23. Since $0.23 \leq 0.4$, the next state is 1. so when i get head as state for third toss, then distribution changes to [0.9 0.1]

- **Fourth toss:** Current state = 1, transition probabilities = $[0.9, 0.1]$.

  $$\text{Cumulative probabilities: } [0.9, 1.0]$$

  Sample = 0.97. Since $0.97 \leq 1.0$, the next state is 2.

Final Results

The sequence of states for the 4 tosses is:

$$\{2, 2, 1, 2\}$$

Q5: Let X and Y be independent random variables with common distribution function F.

1. PDF of $Z_1 = \max(X, Y)$

CDF of Z_1:

$$\begin{aligned}
F_{Z_1}(z) &= P(Z_1 \le z) \\
&= P(\max(X, Y) \le z) \\
&= P(X \le z \text{ and } Y \le z) \\
&= P(X \le z) \cdot P(Y \le z) \quad \text{(using independence)} \\
&= F(z) \cdot F(z) \\
&= F(z)^2
\end{aligned}$$

PDF of Z_1:

$$f_{Z_1}(z) = \frac{d}{dz} F(z)^2 = 2F(z)f(z)$$

2. PDF of $Z_2 = \min(X, Y)$

CDF of Z_2:

$$\begin{aligned}
F_{Z_2}(z) &= P(Z_2 \le z) \\
&= P(\min(X, Y) \le z) \\
&= 1 - P(\min(X, Y) > z) \\
&= 1 - P(X > z \text{ and } Y > z) \\
&= 1 - P(X > z) \cdot P(Y > z) \quad \text{(using independence)} \\
&= 1 - (1 - F(z)) \cdot (1 - F(z)) \\
&= 1 - (1 - F(z))^2
\end{aligned}$$

PDF of Z_2:

$$f_{Z_2}(z) = \frac{d}{dz} \left[1 - (1 - F(z))^2\right] = 2(1 - F(z))f(z)$$

Final PDFs

$$\begin{aligned}
f_{Z_1}(z) &= 2F(z)f(z) \\
f_{Z_2}(z) &= 2(1 - F(z))f(z)
\end{aligned}$$

10 Mark Questions

Q1: Let $\mathcal{D} = \{x_1, \dots x_n\}$ 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

A: 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 $\mathcal{D}$ is defined as

$$\begin{aligned} L(x_1, x_2, \dots x_n; a, b) &= f_{U_1, \dots U_n}(x_1, \dots x_n; a, b) \\ &= \prod f_{U_i}(x_i; a, b) \quad \text{as the samples are i.i.d.} \end{aligned}$$

From the pdf, it is clear that $L \neq 0$ if $a \leq x_i \leq b \forall i = 1 \dots n$. So $L \neq 0$ iff $a < min_i(x_i)$ and $b > max_i(x_i)$

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

The log likelihood is

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

The MLE estimate for a is given by $\hat{a}_{ML} = \arg\max_a \log L(x_1, \dots x_n; a, b)$
To find the maxima, we take the derivative w.r.t a

$$\frac{\partial \log L}{\partial a} = \frac{n}{b-a} \quad a \leq min_i(x_i),\, b \geq max_i(x_i)$$

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 $\hat{b}_{ML} = \arg\max_b \log L(x_1, \dots x_n; a, b)$
To find the maxima, we take the derivative w.r.t b

$$\frac{\partial \log L}{\partial b} = \frac{-n}{b-a} \quad a \leq min_i(x_i),\, b \geq max_i(x_i)$$

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

$$\hat{b}_{ML} = \max_i x_i$$

Q2: Bayesian Inference/ Conjugate prior problem: Suppose $D = \{x_1, ..., x_n\}$ is a data set consisting of independent samples of a Poisson random variable with unknown parameter λ^*. Now assume a prior model $\Lambda \sim Gamma(\alpha, \beta)$ on the unknown parameter λ^* (see hint below for gamma distribution). Obtain an expression for the posterior distribution on λ^*. (7mks). What is the MAP estimate for λ^*? (3mks)

Hint: Use Prior belief: $\Lambda \sim Gamma(\alpha, \beta)$,

$$f_\Lambda(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}; \quad \lambda > 0.$$

and Likelihood of observing x given $\Lambda = \lambda$:

$$f_{X|\Lambda}(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

A:

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

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

The posterior distribution can be found using Bayes' rule

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

Let $(\beta+n)\lambda = t$

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

$$= \frac{(\lambda^*)^{\alpha-1+\Sigma_{i=1}^n x_i} e^{-(\beta+n)\lambda^*}}{\frac{1}{(\beta+n)^{\alpha+\Sigma_{i=1}^n x_i}}\int_0^\infty t^{\alpha-1+\Sigma_{i=1}^n x_i} e^{-t}\,dt}$$

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

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

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

(b) To find the MAP estimate of λ^*

$$\lambda_{MAP} = \underset{\lambda^*}{\operatorname{argmax}}\, f_{\Lambda|X_1,\cdots,X_n}(\lambda^*|x_1,\cdots,x_n)$$

Ignoring the variables independent of λ^*

$$\lambda_{MAP} = \underset{\lambda^*}{\operatorname{argmax}} (\lambda^*)^{\alpha-1+\Sigma_{i=1}^n x_i} e^{-(\beta+n)\lambda^*}$$

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

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

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

Thus our MAP estimate for λ^* is

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

Q3: **(i)** For a Markov chain, let F_{ii} denote the probability of the chain ever returning to state i having started in state i, and let f_{ii}^n denote the probability of visiting state i for the first time in exactly n steps, having started in state i. Show that:

$$F_{ii} = \sum_{n=1}^\infty f_{ii}^n.$$

(ii) For a Markov chain with state space $\mathcal{S} = \{1,2,3\}$ and transition matrix:

$$P = \begin{bmatrix} p & 1-p & 0 \\ p & 1-2p & p \\ 0 & 0 & 1 \end{bmatrix},$$

use the above equality to find F_{ii} for $i = 1,2,3$. From the values of F_{ii}, deduce which states are transient and recurrent.

A:

(a) Let

$$S_{ii} := \{\{X_j\}_{j=0}^{n} \text{ s.t. } X_0 = i \text{ and } \exists n \;\; s.t. \;\; X_n = i\}$$

$$s_{ii}^{n} := \{\{X_j\}_{j=0}^{n} \text{ s.t. } X_0 = i \;,\; X_n = i \;,\; X_k \neq i \;\; \forall k \in \{1, \ldots, n-1\}\}$$

$$S'_{ii} := \bigcup_{n=1}^{\infty} s_{ii}^{n}$$

We make the following claims:

i. For $n_1 \neq n_2$, $s_{ii}^{n_1} \cap s_{ii}^{n_2} = \phi$

ii. $\mathbb{P}(S_{ii}) = \mathbb{P}(S'_{ii})$

For the proof of first claim, assume $\exists y \in s_{ii}^{n_2}$
Without loss of generality, assume $n_1 > n_2$.

$$\implies y = \{X_j\}_{j=0}^{n_2-1}, X_{n_2} = i$$

If $y \in s_{ii}^{n_1}$

$$\implies y = \{X_j\}_{j=0}^{n_1-1}, X_{n_1} = i$$

$$\implies y = \{X_j\}_{j=0}^{n_2-1}, X_{n_2} = i, \{X_j\}_{j=n_2+1}^{n_1-1}, X_{n_1} = i \quad \text{[Taking an intermediate stop at } n_2\text{]}$$

But we are given by definition of s_{ii}^{n} that

$$\forall k \in \{1, \ldots, n-1\} \quad X_k \neq i$$

$$\implies y \notin s_{ii}^{n_1}$$

$$\therefore s_{ii}^{n_1} \cap s_{ii}^{n_2} = \phi$$

- To understand how to prove second claim, suppose there exists a sequence $y \in S_{ii}$ of length n such that $y_k = i$ for some $k \in \{1, \ldots, n-1\}$. The sequence can be decomposed as follows:
  - A segment from $X_0 = i$ to $X_k = i$,
  - A subsequent segment from $X_k = i$ to $X_n = i$.
- We can consider X_k to be the starting point of another sequence, which will also satisfy the given conditions. Subsequences following the first time return will not contribute to the previous subsequence.
- Note that we are able to break these sequences, and consider a new starting point, forgetting the info in past as a consequence of Markov Property.
- We can see that $y \in S_{ii}$ is just a concatenation of first time return sequences, we can also claim that the contributions made by the subsequences following the first time return subsequence will be none, and we can reduce our subsequence.
- Since we are dealing with probability of ever returning back to i, it is sufficient to stop at first time visit, and still have the same probability measure.

$$\therefore \mathbb{P}(S_{ii}) = \mathbb{P}(S'_{ii})$$

When we apply probability axioms on these two sets, we observe the following:

- $F_{ii} = \mathbb{P}(S_{ii}) = \mathbb{P}(S'_{ii}) = \mathbb{P}(\bigcup_{n=1}^{\infty} s_{ii}^n)$
- $f_{ii}^n = \mathbb{P}(s_{ii}^n)$

Now using the third axiom of disjoint countable union additivity, we can write:

$$F_{ii} = \mathbb{P}(\bigcup_{n=1}^{\infty} s_{ii}^n) = \sum_{n=1}^{\infty} \mathbb{P}(s_{ii}^n) = \sum_{n=1}^{\infty} f_{ii}^n$$

Note: The proof above is just for reference, any other proofs which mention with proper reasoning why we can add the probabilities, will also be rewarded marks.

A simpler version of this proof:

$$F_{ii} = P(\text{coming back to state i, having started in state i})$$

$$= P(\text{union of all paths of arbitrary length n, with } X_0 = i, X_n = i)$$

- We can simplify the union by seeing that returning back in n steps exact and m steps with $n \neq m$ has no overlapping paths.
- So the event of ever returning to i is equal to a **disjoint countable union of returning back in exactly n steps where the union is over n**.
- Then applying the third axiom of probability we get the summation.

$$F_{ii} = \sum_{n=1}^{\infty} P(\text{returning back to i, starting in i, and not before in n steps})$$

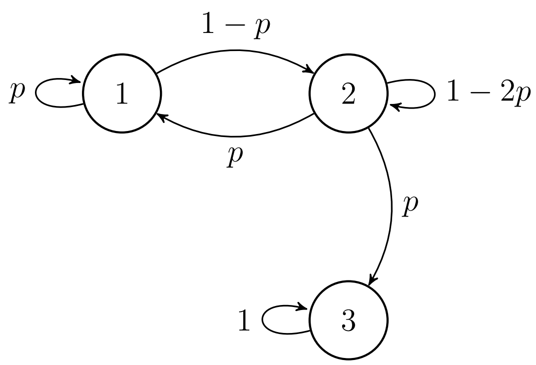
$$\text{Also} \quad f_{ii}^n = P(\text{returning back to i after n steps and not before})$$

$$\implies F_{ii} = \sum_{n=1}^{\infty} P(\text{returning back to i after n exact steps}) = \sum_{n=1}^{\infty} f_{ii}^n$$

(b) For a state i:

- i is recurrent if $F_{ii} = 1$. This is true because you will always come back to i with probability 1.
- i is transient if $F_{ii} < 1$. This is true because there exists some finite non-zero probability $p = 1 - F_{ii}$ which means that with probability p, you will never return to state i.

Markov chain:



To calculate F_{33}, note that $f_{33}^n = 0 \quad \forall \quad n > 1$. This is true because there are no other transitions that will make us leave state 3. So, you cannot visit state 3 starting in 3 after exactly $n > 1$ steps.

$$f_{33}^1 = P(X_1 = 3 \mid X_0 = 3) = 1$$

$$\implies F_{33} = f_{33}^1 + \sum_{n=2}^{\infty} f_{33}^n = 1 + 0 = 1$$

$$\therefore \text{State 3 is } \boxed{\text{recurrent}}$$

—

To calculate F_{22}, observe the transitions. We can write f_{22}^n as follows:

$$f_{22}^1 = P(X_1 = 2 \mid X_0 = 2) = 1 - 2p$$

$$f_{22}^n = P(\text{leave 2, loop } n-2 \text{ times outside 2, then come back to 2})$$

$$\implies f_{22}^n = \sum_{s \in S \setminus \{2\}} P(X_1 = s \mid X_0 = 2) \cdot \prod_{j=2}^{n-1} P(X_j = s \mid X_{j-1} = s) \cdot P(X_n = 2 \mid X_{n-1} = s)$$

$$\implies f_{22}^n = P(X_1 = 1 \mid X_0 = 2) \cdot \prod_{j=2}^{n-1} P(X_j = 1 \mid X_{j-1} = 1) \cdot P(X_n = 2 \mid X_{n-1} = 1)$$

$$[\text{since } P(X_n = 2 \mid X_{n-1} = 3) = 0]$$

$$\implies f_{22}^n = p \cdot \prod_{j=2}^{n-1} p \cdot (1-p) = (1-p) \cdot p^{n-1}$$

Substituting this in the expression, we get:

$$F_{22} = f_{22}^1 + \sum_{n=2}^{\infty} f_{22}^n$$

$$= 1 - 2p + \sum_{n=2}^{\infty} (1-p) \cdot p^{n-1}$$

$$= 1 - 2p + \frac{1-p}{p} \sum_{n=2}^{\infty} p^n$$

$$= 1 - 2p + \frac{1-p}{p} \cdot \frac{p^2}{1-p}$$

$$= 1 - 2p + p = 1 - p < 1$$

$$\therefore \text{State 2 is } \boxed{\text{transient}}$$

—

To calculate F_{11}, we can write f_{11}^n as follows:

$$f_{11}^1 = P(X_1 = 1 \mid X_0 = 1) = p$$

$$f_{11}^n = P(\text{leave 1, loop } n-2 \text{ times outside 1, then come back to 1})$$

$$\implies f_{11}^n = \sum_{s \in S \setminus \{1\}} P(X_1 = s \mid X_0 = 1) \cdot \prod_{j=2}^{n-1} P(X_j = s \mid X_{j-1} = s) \cdot P(X_n = 1 \mid X_{n-1} = s)$$

$$\implies f_{11}^n = P(X_1 = 2 \mid X_0 = 1) \cdot \prod_{j=2}^{n-1} P(X_j = 2 \mid X_{j-1} = 2) \cdot P(X_n = 1 \mid X_{n-1} = 2)$$

$$[\text{since } P(X_n = 1 \mid X_{n-1} = 3) = 0]$$

$$\implies f_{11}^n = (1-p) \cdot \prod_{j=2}^{n-1} (1-2p) \cdot p = p(1-p) \cdot (1-2p)^{n-2}$$

Substituting this in the expression, we get:

$$F_{11} = f_{11}^1 + \sum_{n=2}^{\infty} f_{11}^n$$

$$= p + \sum_{n=2}^{\infty} p(1-p) \cdot (1-2p)^{n-2}$$

$$= p + \frac{p(1-p)}{(1-2p)^2}$$

$$= p + \frac{(1-p)}{2}$$

$$= \frac{p+1}{2} < 1$$

Note that $p < \frac{1}{2}$ since $P(X_i = 2 \mid X_{i-1} = 2) = 1 - 2p > 0$.

$$\therefore \text{State 1 is } \boxed{\text{transient}}$$

Q4: Gaussian: Suppose $\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}$, where A is an $n \times n$ matrix and $\mathbf{Z}$ is a standard normal vector of length n. Derive the expression for the mean $\mathbb{E}[\mathbf{X}]$ and covariance matrix $\mathbf{C_X}$. (5 mks) Also derive the expression for the pdf of $\mathbf{X}$. (5 mks)

A:

(a) **Mean $\mathbb{E}[X]$:**

Applying the expectation operator to both sides of $X = AZ + \mu$, we get:

$$\mathbb{E}[X] = \mathbb{E}[AZ + \mu].$$

$$\mathbb{E}[X] = A\mathbb{E}[Z] + \mathbb{E}[\mu].$$

Since Z is a standard normal vector, $\mathbb{E}[Z] = 0$, and μ is a constant vector, so $\mathbb{E}[\mu] = \mu$.

$$\mathbb{E}[X] = A \cdot 0 + \mu = \mu.$$

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

(b) **Covariance Matrix C_X:**

The covariance matrix of X is given by:

$$C_X = \text{Cov}(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T].$$

Since, $\mathbb{E}[X] = \mu$.

$$C_X = \mathbb{E}[(X - \mu)(X - \mu)^T]$$

Substituting $X = AZ + \mu$ we have:

$$C_X = \mathbb{E}[(AZ)(AZ)^T].$$

$$C_X = A\mathbb{E}[ZZ^T]A^T.$$

Since Z is a standard normal vector, its covariance matrix is the identity matrix I, i.e., $\mathbb{E}[ZZ^T] = I$. Thus:

$$C_X = AIA^T = AA^T.$$

Therefore, the covariance matrix of X is:

$$\boxed{C_X = AA^T.}$$

(c) **PDF of X:**

Let's start with the PDF of the standard normal vector Z. For a standard normal vector where Z_i's are i.i.d. and $Z_i \sim N(0,1)$, the PDF is given by:

$$f_Z(z) = \prod_{i=1}^{n} f_{Z_i}(z_i) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} z^T z\right)$$

When $\mathbf{X} = G(\mathbf{Z})$, where $G : \mathbb{R}^n \to \mathbb{R}^n$ is continuous, invertible, and has continuous partial derivatives, let H denote its inverse. Then

$$f_X(\mathbf{x}) = f_Z(H(\mathbf{x}))|J|$$

where J is the determinant of the Jacobian matrix of H.
Here $\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}$, since A is invertible, we can write:

$$\mathbf{Z} = A^{-1}(\mathbf{X} - \boldsymbol{\mu}),$$

the inverse function H is given by

$$H(\mathbf{x}) = A^{-1}(\mathbf{x} - \boldsymbol{\mu}).$$

Now, let the entries of A^{-1} be:

$$A^{-1} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

The i-th entry of $H_i(\mathbf{x})$ is:

$$H_i(\mathbf{x}) = \sum_{j=1}^{n} a_{ij}(x_j - \mu_j).$$

Differentiating $H_i(\mathbf{x})$ with respect to x_k gives:

$$\frac{\partial H_i}{\partial x_k} = a_{ik}.$$

The Jacobian matrix J is:

$$J = \begin{bmatrix} \frac{\partial H_1}{\partial x_1} & \frac{\partial H_1}{\partial x_2} & \cdots & \frac{\partial H_1}{\partial x_n} \\ \frac{\partial H_2}{\partial x_1} & \frac{\partial H_2}{\partial x_2} & \cdots & \frac{\partial H_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial H_n}{\partial x_1} & \frac{\partial H_n}{\partial x_2} & \cdots & \frac{\partial H_n}{\partial x_n} \end{bmatrix} = A^{-1}.$$

Finally, the determinant of the Jacobian is:

$$|J| = \det(A^{-1}) = \frac{1}{\det(A)}.$$

Thus, we conclude that

$$f_X(x) = \frac{1}{|\det(A)|} f_Z(A^{-1}(x-\mu))$$

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\det(A)|} \exp\left\{-\frac{1}{2}(A^{-1}(x-\mu))^T(A^{-1}(x-\mu))\right\}$$

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\det(A)|} \exp\left\{-\frac{1}{2}(x-\mu)^T A^{-T}A^{-1}(x-\mu)\right\}$$

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\det(A)|} \exp\left\{-\frac{1}{2}(x-\mu)^T (AA^T)^{-1}(x-\mu)\right\}$$

Q5: **Shifted Exponential:** Let X be a random variable following a shifted exponential distribution with rate parameter $\lambda > 0$ and shift parameter μ. The probability density function (PDF) of X is given by:

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

(a) Derive the MGF of X, state its region of convergence (5 mks)

(b) Using the MGF, obtain the first and second moments of X. What is the variance of X? (5 mks)

A:

(a)

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \int_{\mu}^{\infty} e^{tx} \cdot \lambda e^{-\lambda(x-\mu)}\, \mathrm{d}x \\ &= \lambda e^{\lambda\mu} \int_{\mu}^{\infty} e^{-(\lambda-t)x}\, \mathrm{d}x \\ &= \lambda e^{\lambda\mu} \left[-\frac{e^{-(\lambda-t)x}}{\lambda - t}\right]_{\mu}^{\infty} \\ &= \lambda e^{\lambda\mu} \left(\frac{e^{-(\lambda-t)\mu}}{\lambda-t}\right), \text{ for } t < \lambda \\ &= \frac{\lambda e^{\mu t}}{\lambda - t}, \text{ for } t < \lambda \end{aligned}$$

$M_X(t)$ exists for all $t < \lambda$ since $e^{-(\lambda-t)x}$ becomes 0 at infinity only if $(\lambda - t)$ is positive.

(b) To find $\mathbb{E}[X]$, we need to find the derivative of $M_X(t)$, since $\mathbb{E}[X] = M'_X(0)$.

$$M'_X(t) = \frac{\mathrm{d}}{\mathrm{d}t} \lambda e^{\mu t}(\lambda - t)^{-1} = \frac{\mu\lambda e^{\mu t}}{\lambda - t} + \frac{\lambda e^{\mu t}}{(\lambda - t)^2}$$

$$\mathbb{E}[X] = M_X'(0) = \frac{\lambda\mu}{\lambda} + \frac{\lambda}{\lambda^2} = \mu + \frac{1}{\lambda}$$

Similarly, $\mathbb{E}[X^2] = M_X''(0)$.

$$\begin{aligned} M_X''(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\, \mu\lambda e^{\mu t}(\lambda - t)^{-1} + \lambda e^{\mu t}(\lambda - t)^{-2} \\ &= \frac{\mu^2\lambda e^{\mu t}}{\lambda - 1} + 2\frac{\mu\lambda e^{\mu t}}{(\lambda - t)^2} + \frac{2\lambda e^{\mu t}}{(\lambda - t)^3} \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X^2] &= M_X''(0) \\ &= \frac{\mu^2\lambda}{\lambda} + 2\frac{\mu\lambda}{\lambda^2} + \frac{2\lambda}{\lambda^2} \\ &= \left(\mu + \frac{1}{\lambda}\right)^2 + \frac{1}{\lambda^2} \end{aligned}$$

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \left(\mu + \frac{1}{\lambda}\right)^2 + \frac{1}{\lambda^2} - \left(\mu + \frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$