Problem 1
– Akshit
Stochastic Simulation: Suppose you want to generate samples from a discrete random variable X having pmf $\{p_j, j \geq 0\}$. Now assume that you have a access to samples from another discrete random variable Y with pmf $\{q_j, j \geq 0\}$ with the property that $\frac{p_j}{q_j} \leq c$ for some constant c and for all j such that $p_j > 0$. The rejection method generates samples of X as follows:

1. Simulate/Generate the value of Y with mass function $\{q_j, j \geq 0\}$
2. Generate random number U which is uniform in the interval [0,1]
3. If $U < \frac{p_Y}{cq_Y}$, set X=Y and stop. Otherwise return to step 1.

Prove that samples of X generated using the above algorithm indeed have pmf $\{p_j, j \geq 0\}$.

Solution:

$$\begin{aligned}
P(X = x) &= \sum_{n=1}^{\infty} \text{P(reject } n-1 \text{ times, draw Y=}x \text{ and accept it)} \\
&= \sum_{n=1}^{\infty} \text{P(reject)}^{n-1} \cdot \text{P(draw Y=}x\text{ and accept)} \\
\text{Now, P(draw Y=}x\text{ and accept it)} &= \text{P(draw Y=}x) \cdot \text{P(accept Y|Y=}x) \\
&= q(x) \cdot P(U < \frac{p(Y)}{c \cdot q(Y)} | Y = x) \\
&= q(x) \cdot \frac{p(Y)}{c \cdot q(Y)} \\
&= \frac{p(x)}{c} \\
\text{P(reject Y)} &= \sum_{x \in \Omega} \text{P(draw Y=x and reject it)} \\
&= \sum_{x \in \Omega} q(x) \cdot P(U \geq \frac{p(Y)}{c \cdot q(Y)} | Y = x) \\
&= \sum_{x \in \Omega} q(x)(1 - \frac{p(x)}{c \cdot q(x)}) \\
&= (1 - \frac{1}{c}) \\
\text{So, P(X=x)} &= \sum_{n=1}^{\infty} \text{P(reject)}^{n-1} \cdot \text{P(draw Y=}x\text{ and accept)} \\
&= \sum_{n=1}^{\infty} (1 - \frac{1}{c})^{n-1} \cdot \frac{p(x)}{c} \\
&= \frac{1}{c} \sum_{n=0}^{\infty} (1 - \frac{1}{c})^{n} \cdot p(x) \\
&= \frac{1}{c} \cdot \frac{p(x)}{1 - (1 - \frac{1}{c})} \\
&= p(x)
\end{aligned}$$

So, the samples of X generated using given algorithm indeed have the pmf $\{p_j, j \geq 0\}$
Marking Scheme :-

- Full marks only for analytically showing that pdf is actually p_j
- Partial marks for deriving intermediate probabilities used in final calculation
- 4 marks for solving for pdf of accepted samples
- 3+3 marks for both calculations, of the other probability terms used in final calculation

---

Problem 2
– Sriya
MGF: Derive the expression for the MGF of a Gaussian $\mathcal{N}(\mu, \sigma^2)$ random variable and use the

MGF to identify the first and the second moment. Furthermore, using MGF, show that sum of n independent Gaussian $\mathcal{N}(\mu, \sigma^2)$ random variables is also a Gaussian random variable. What are the resulting mean and variance parameters?

Solution: Let $X \sim \mathcal{N}(\mu, \sigma^2)$
The cdf is defined as $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]$
MGF is defined as $M_X(t) = E[e^{tX}]$
$M_X(t) = \int e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$
Let $z = \frac{x-\mu}{\sigma}$,
$\Rightarrow x = z\sigma + \mu$
$\Rightarrow \frac{dx}{dz} = \sigma$
Therefore, the MGF becomes
$M_X(t) = \int e^{\mu t + z\sigma t} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2} \frac{dx}{dz} dz$
Since $\frac{dx}{dz} = \sigma$ we get
$M_X(t) = e^{\mu t} \int e^{z\sigma t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$
Now, consider the term $e^{z\sigma t} e^{-\frac{1}{2}z^2}$

$$\begin{aligned}
e^{z\sigma t} e^{-\frac{1}{2}z^2} &= e^{-\frac{1}{2}z^2 + z\sigma t} \\
&= e^{-\frac{1}{2}z^2 + z\sigma t - \frac{1}{2}(\sigma t)^2 + \frac{1}{2}(\sigma t)^2} \\
&= e^{-\frac{1}{2}(z-\sigma t)^2 + \frac{1}{2}(\sigma t)^2} \\
&= e^{-\frac{1}{2}(z-\sigma t)^2} e^{\frac{1}{2}(\sigma t)^2}
\end{aligned}$$

Substituting in the MGF equation,

$$\begin{aligned}
M_X(t) &= e^{\mu t} e^{\frac{1}{2}(\sigma t)^2} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-\sigma t)^2} dz \\
&= e^{\mu t + \frac{1}{2}\sigma^2 t^2}
\end{aligned}$$

The final equation follows from the fact that the expression under the integral is the $\mathcal{N}(\sigma t, 1)$ probability density function which integrates to 1.
Therefore, the MGF of a Gaussian $\mathcal{N}(\mu, \sigma^2)$ random variable X is $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

The nth moment of a random variable X is given by $M_X^{(r)}(0)$
For the first moment,
$M_X^{(1)}(t) = (\mu + \sigma^2 t) e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
$M_X^{(1)}(0) = \mu$
For the second moment,
$M_X^{(2)}(t) = (\sigma^2 + (\mu + \sigma^2 t)^2) e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
$M_X^{(2)}(0) = \sigma^2 + \mu^2$

Let $X_1, X_2, ..., X_n$ be n independent Gaussian $\mathcal{N}(\mu, \sigma^2)$ random variables.
We know that the MGF of the sum of independent random variables is equal to the product of the MGFs of each individual random variable.

Let $Y = X_1 + X_2 + ... + X_n$

$$\begin{aligned} M_Y(t) &= M_{X_1}(t)M_{X_2}(t)...M_{X_n}(t) \\ &= e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2} e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2} ... e^{\mu_n t + \frac{1}{2}\sigma_n^2 t^2} \\ &= e^{(\mu_1 + \mu_2 + ... + \mu_n)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2)t^2} \end{aligned}$$

We can see that the MGF of Y is the same as the MGF of a Gaussian $\mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.
By uniqueness property of MGFs, Y must be a Gaussian random variable with mean $\sum_{i=1}^n \mu_i$ and variance $\sum_{i=1}^n \sigma_i^2$

Grading Scheme:
5 marks - deriving MGF
2 marks - finding the moments
3 marks - proving that the sum is a gaussian random variable with the correct mean and variance

---

Problem 3
– Rudransh
MLE: Get the MLE of mean and standard deviation for k i.i.d samples from a Gaussian Random variable
Is the MLE for standard deviation biased? Give reason.

Solution:

$$L(\mu, \sigma; x_1, x_2, \dots, x_k) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

$$\ln L(\mu, \sigma; x_1, x_2, \dots, x_k) = -\frac{k}{2}\ln(2\pi) - k\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{k}(x_i - \mu)^2$$

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

$$\frac{1}{\sigma^2}\sum_{i=1}^{k}(x_i - \mu) = 0$$

$$\hat{\mu} = \frac{1}{k}\sum_{i=1}^{k} x_i$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{k}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{k}(x_i - \mu)^2$$

$$-\frac{k}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{k}(x_i - \mu)^2 = 0$$

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

$$\hat{\mu} = \frac{1}{k}\sum_{i=1}^{k} x_i$$

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

$$E[\hat{\mu}^2] = (E[\hat{\mu}])^2 + Var(\hat{\mu})$$
$$= \mu^2 + \frac{\sigma^2}{n}$$

$$E[\hat{\sigma}^2] = \frac{1}{k}(\sum_{k=1}^{n} Ex_k^2 - kE[\hat{\mu}]^2)$$

$$= \frac{1}{k}(k(\mu^2 + \sigma^2) - k(\mu^2 + \frac{\sigma^2}{k})) = \frac{k-1}{k}\sigma^2$$
$$E[\hat{\sigma}^2] < \sigma^2$$

$$E[\hat{\sigma}^2] = E[\hat{\sigma}]^2 + Var(\hat{\sigma})$$

Since $Var(\hat{\sigma}) > 0$

$$E[\hat{\sigma}^2] > E[\hat{\sigma}]^2$$
$$E[\hat{\sigma}]^2 < E[\hat{\sigma}^2] < \sigma^2$$
$$E[\hat{\sigma}] < \sigma$$

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

$$Bias[\hat{\sigma}] \neq 0$$

Therefore, biased estimator

Marking Scheme:

- 2.5 Marks for MLE of mean
- 2.5 Marks for MLE of s.d
- 1 Marks mentioning biased.
- 2 Marks for Calculating till $E[\sigma^2]$
- 2 for finally showing biased.

Problem 4
– Aman
Let $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ where X_1 and X_2 are independent exponential random variables with parameter 1. Find the probability density function of $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ where $U_1 = X_1 + X_2$ and $U_2 = \frac{X_1}{X_1+X_2}$.

Solution: We have the probability density of X,

$$f_X(\begin{bmatrix} x_1 & x_2 \end{bmatrix}) = f_{X_1}(x_1) f_{X_2}(x_2) \text{ (since } X_1 \text{ and } X_2 \text{ are independent)}$$
$$= e^{-x_1} e^{-x_2} \Rightarrow f_X(\begin{bmatrix} x_1 & x_2 \end{bmatrix}) = e^{-x_1-x_2} ...(1)$$

Now,
Let G denotes a function that maps X to U i.e.

$$G(\begin{bmatrix} X_1 & X_2 \end{bmatrix}) = \begin{bmatrix} X_1 + X_2 & \frac{X_1}{X_1+X_2} \end{bmatrix}$$

Now, notice that G is both one-one and onto. Thus, it is an invertible function. Let $H = G^{-1}$. Now, we have

$$\begin{bmatrix} U_1 & U_2 \end{bmatrix} = G(\begin{bmatrix} X_1 & X_2 \end{bmatrix}) = \begin{bmatrix} X_1 + X_2 & \frac{X_1}{X_1+X_2} \end{bmatrix}$$
$$\Rightarrow U_1 = X_1 + X_2, U_2 = \frac{X_1}{X_1+X_2}$$
$$\Rightarrow X_1 = U_1 U_2, X_2 = U_1(1 - U_2)$$

Thus, the inverse function of G is

$$H(\begin{bmatrix} U_1 & U_2 \end{bmatrix}) = \begin{bmatrix} U_1 U_2 & U_1(1 - U_2) \end{bmatrix}$$

Thus, we need to find $f_U(U)$ where $U = G(X) \Rightarrow X = G^{-1}(U) = H(U)$.
By the method of transforms applied to vectors, we have

$$f_U(U) = f_X(X)|J|$$

where $|J|$ denotes the positive determinant of Jacobian matrix at X.
Now, we have

$$|J| = \begin{vmatrix} \frac{\partial X_1}{\partial U_1} & \frac{\partial X_1}{\partial U_2} \\ \frac{\partial X_2}{\partial U_1} & \frac{\partial X_2}{\partial U_2} \end{vmatrix}$$
$$= \begin{vmatrix} U_2 & U_1 \\ 1 - U_2 & -U_1 \end{vmatrix} = |U_2(-U_1) - U_1(1 - U_2)| = U_1$$

Also, we have

$$f_X(X) = e^{-X_1-X_2}$$
$$= e^{-(U_1U_2)-U_1(1-U_2)} = e^{-U_1}$$

So, we get by putting all these values,

$$\Rightarrow f_U(U) = e^{-U_1} U_1$$

So, the probability distribution is given by

$$f_U(U) = U_1 e^{-U_1}$$

Marking Scheme:

- 10 Marks for getting correct probability distribution.
- Partial marking will be done wherever applicable.

Problem 5
–Vansh
Bayesian Inference problem: Suppose $D = x_1, \cdots, x_n$ is a data set consisting of independent samples of a Bernoulli random variable with unknown parameter θ, i.e. $f(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$ for $x_i \in 0, 1$. Obtain an expression for the posterior distribution on θ. Using this, obtain θ_{MAP} and the conditional expectation estimator θ_{CE}. (Hint: you may use the fact that $\int_0^1 \theta^m(1-\theta)^r d\theta = \frac{m!r!}{(m+r+1)!}$

Solution: We know that the posterior of θ given $x_1, \cdots, x_n$ is given by

$$\begin{aligned} f(\theta|x_1, ..., x_n) &= \frac{f(x_1, ..., x_n, \theta)}{f(x_1, ..., x_n)} \\ &= \frac{f(x_1, ..., x_n|\theta)p(\theta)}{\int_0^1 f(x_1, ..., x_n|\theta)\rho(\theta)d\theta} \\ &= \frac{\theta^{\Sigma_{i=1}^n x_i}(1-\theta)^{n-\Sigma_{i=1}^n x_i}}{\int_0^1 \theta^{\Sigma_{i=1}^n x_i}(1-\theta)^{n-\Sigma_{i=1}^n x_i} d\theta} \end{aligned}$$

Now using the fact that for integral values m and r

$$\int_0^1 \theta^m(1-\theta)^r d\theta = \frac{m!r!}{(m+r+1)!}$$

and letting $s = \sum_{i=1}^n x_i$, the expression for the posterior becomes:

$$f(\theta|x_1, ..., x_n) = \frac{(n+1)!\theta^s(1-\theta)^{n-s}}{s!(n-s)!}$$

Now, solving for θ_{MAP}:

$$\begin{aligned} \theta_{MAP} &= \arg\max_\theta P(x_1, \cdots, x_n|\theta) \\ &= \arg\max_\theta \theta^s(1-\theta)^{n-s} \end{aligned}$$

Taking the derivative and setting to 0, we get:

$$\begin{aligned} s\theta^{s-1}(1-\theta)^{n-s} - (n-s)\theta^s(1-\theta)^{n-s-1} &= 0 \\ \theta^{s-1}(1-\theta)^{n-s-1}\left[s(1-\theta) - (n-s)\theta\right] &= 0 \end{aligned}$$

Since, θ cannot take the value of 0 or 1 for maximizing the posterior, the second part of the above expression must go to 0.

$$\begin{aligned} s(1-\theta) - (n-s)\theta &= 0 \\ s - s\theta - n\theta + s\theta &= 0 \\ \theta &= \frac{s}{n} \end{aligned}$$

We got that the MAP estimate for θ is $\frac{s}{n}$ which is just the sample mean $\overline{X}$ of the data.

For θ_{CE}, we have:

$$
\begin{aligned}
\theta_{CE} &= E[\theta|x_1, \cdots, x_n] \\
&= \int_0^1 \theta f(\theta|x_1, \cdots, x_n) d\theta \\
&= \frac{(n+1)!}{s!(n-s)!} \int_0^1 \theta^{1+s}(1-\theta)^{n-s} d\theta \\
&= \frac{(n+1)!}{s!(n-s)!} \frac{(1+s)!(n-s)!}{(n+2)!} \\
&= \frac{s+1}{n+2}
\end{aligned}
$$

Marking Scheme

- 5 marks for calculating the posterior
- 5 marks for finding the MAP estimate
- 5 marks for finding the CE estimate

Problem 6a
– Ishwar
Let $X_1, X_2, X_3, ..$ be a sequence of random variables with density

$$f_{X_n}(x) = \frac{n}{2} e^{-n|x|}.$$

Show that X_n converges to 0 in probability and in distribution

Solution: For any ε greater than 0,

$$
\begin{aligned}
\lim_{n\to\infty} P\left(|X_n - 0| \geq \varepsilon\right) &= \lim_{n\to\infty} P\left(|X_n| \geq \varepsilon\right) \\
&= \lim_{n\to\infty} 2P\left(X_n \geq \varepsilon\right) \quad \text{(by symmetry)} \\
&= \lim_{n\to\infty} \int_{\varepsilon}^{\infty} n\, e^{-nx} dx \\
&= \lim_{n\to\infty} \frac{1}{e^{n\varepsilon}} \\
&= 0.
\end{aligned}
$$

Therefore X_n converges to 0 in probability

$$
\begin{aligned}
\lim_{n\to\infty} F_{X_n}(x) &= \lim_{n\to\infty} P(X_n \le x) \\
&= \lim_{n\to\infty} \int_{-\infty}^{x} \frac{n}{2}\, e^{-n|x|} dx \\
&= \begin{cases} \lim_{n\to\infty} \int_{-\infty}^{0} \frac{n}{2}\, e^{nx} dx + \int_{0}^{x} \frac{n}{2}\, e^{-nx} dx & \text{if } x \ge 0 \\ \lim_{n\to\infty} \int_{-\infty}^{x} \frac{n}{2}\, e^{nx} dx & \text{if } x < 0 \end{cases} \\
&= \begin{cases} \lim_{n\to\infty} 1 - \frac{1}{2}\, e^{-nx} & \text{if } x \ge 0 \\ \lim_{n\to\infty} \frac{1}{2}\, e^{nx} & \text{if } x < 0 \end{cases} \\
&= \begin{cases} 1 & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases} \\
&= F_X(x)
\end{aligned}
$$

where X is the constant random variable 0.
Therefore X_n converges to 0 in distribution.

Marking Scheme:

- 1 + 1 Marks for correctly applying the definitions
- 3 + 3 Marks for the Proof.
- Partial Marks will be awarded
- No partial marks will be awarded for using the result mentioned in the paper that was instructed not to be used.

Problem 6b
– Gowlapalli Rohit
Let X_n be a Poisson($n\lambda$) random variable for $n = 1, 2, 3, \dots$. Consider the sequence of random variables $Y_n = \frac{1}{n} X_n$, for $n = 1, 2, 3, \dots$.Show that Y_n converges in mean square sense to λ

Solution:

$$
\begin{aligned}
E[|Y_n - \lambda|^2] &= E[(Y_n - \lambda)^2] \\
&= E\left[\left(\frac{1}{n} X_n - \lambda\right)^2\right] \\
&= E\left[\frac{1}{n^2}(X_n - n\lambda)^2\right] \\
&= \frac{1}{n^2} E[(X_n - n\lambda)^2] \\
&= \frac{1}{n^2} E[(X_n - E[X_n])^2] \quad \textit{(Expected value of a Poisson random variable} = n\lambda\textit{)} \\
&= \frac{1}{n^2} \mathrm{Var}[X_n] \\
&= \frac{1}{n^2}(n\lambda) \quad \textit{(Variance of a Poisson random variable} = n\lambda\textit{)} \\
&= \frac{\lambda}{n}.
\end{aligned}
$$

Thus, $\lim_{n\to\infty} E[|Y_n - \lambda|^2] = 0$, hence $Y_n \xrightarrow{\text{m.s.}} \lambda$.

Grading Scheme:

* 7 Marks - Proving Mean Square Convergence of Y_n
* Partial Marks will be awarded incase of Alternative Approaches

---

Problem 7a
– Gnana Prakash

Let $\mathcal{D} = \{x_1, \ldots, x_n\}$ denote i.i.d samples from a uniform random variable $U[0, a]$, where a is unknown. Find an MLE estimate for the unknown parameter a.

Solution: Since we know that the samples are drawn from the uniform distribution $U[0, a]$, we have that the PDF of a sample x_i is given by

$$f(x_i) = \frac{1}{a-0} = \frac{1}{a} \quad \forall\, 0 \le x_{min} \le x_{max} \le a$$

Now, the likelihood function $\mathcal{L}$ is given by

$$\mathcal{L}(x_i, \ldots, x_n;\, a) = P(x_1, \ldots, x_n;\, a)$$

Since we have that the samples are i.i.d,

$$P(x_1, \ldots, x_n;\, a) = \prod_{i=1}^{n} P(x_i;\, a)$$

$$\implies \mathcal{L}(x_1, \ldots, x_n;\, a) = \prod_{i=1}^{n} P(x_i;\, a) = \frac{1}{a^n}$$

$$\implies \mathcal{L}(\mathbf{x};\, a) = \frac{1}{a^n} \quad \forall\, 0 \le \mathbf{x} \le a$$

On calculating the derivative of $\mathcal{L}(\mathbf{x};\, a)$ with respect to a and setting it 0 to maximize $\mathcal{L}(a)$, we get the following expression:

$$\frac{d\mathcal{L}}{da} = \frac{-n}{a^{n+1}} \neq 0$$

So, we analyse the the expression on the RHS of the likelihood function and we can infer that $1/a^n$ is a decreasing function. Hence, the maximum value of the likelihood function will be at the minimum possible value of a, i.e., $max(\mathcal{L}) = min(a)$
We also have to note the constraint on a that $0 \le x_{min} \le x_{max} \le a$ and hence, the maximum likelihood estimate for the unknown parameter a is given by

$$a_{MLE} = max(x_1, \ldots, x_n)$$

Grading Scheme:

* Deriving the expression of MLE - 4 Marks
* Showing that direct differentiation does not give MLE - 2 Marks
* Analytical derivation of MLE based on constraints - 2 Marks

---

Problem 7b
- Revanth

Let $\mathcal{D} = \{x_1, \ldots, x_n\}$ denote i.i.d samples from a poison random variable with unknown parameter γ. Find an MLE estimate for the unknown parameter γ.

Solution:
Poisson distribution with parameter γ is given by -

$$f(x;\gamma) = \frac{\gamma^x e^{-\gamma}}{x!}$$

Now the MLE estimate γ^* is given by $argmax \prod_i f(x_i;\gamma)$. Also we know $\gamma = \text{argmax}_\gamma f(x;\gamma) = \text{argmax}_\gamma log(f(x;\gamma))$ as log is monotonically increasing. Now let us proceed with the calculations -

$$\begin{aligned}
\gamma^* &= \text{argmax}_\gamma \log(\prod_i f(x_i,\gamma)) \\
&= \text{argmax}_\gamma \log\left(\frac{\gamma^{\sum x_i} e^{-n\gamma}}{x_1! x_2! \cdots x_n!}\right) \\
&= \text{argmax}_\gamma \log\left(\gamma^{\sum x_i} e^{-n\gamma}\right) \\
&= \text{argmax}_\gamma \left(\left[\sum x_i\right] \log(\gamma) - n\gamma\right)
\end{aligned}$$

In the third step, we note that x_i's are irrelevant for argmax and we drop them out of calculations. Now to find maximum and argmax we use the standard practice of differentiation and setting it to 0 to find argmax γ. Differentiating the last equation gives -

$$\begin{aligned}
\frac{\sum x_i}{\gamma} - n &= 0 \\
\gamma &= \frac{\sum x_i}{n}
\end{aligned}$$

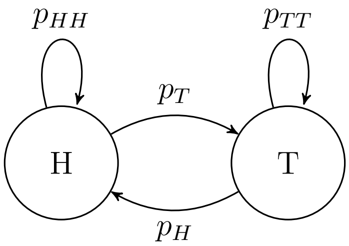
Hence our MLE estimate for γ is $\sum x_i / n$
Scheme:

* Deriving the expression of MLE - 2 Marks
* for explaing why log (if used) for finding MLE - 1 Mark
* Derivation of MLE- 4 Marks

noindent

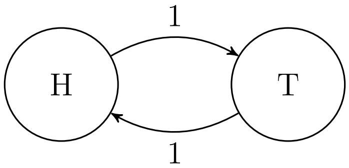
Problem 8a
- Anush
Consider a discrete-time Markov chain with the following transition probabilities $p_{ij} = 0$ when $i = j$. Let the states be denoted as 'H' for head and 'T' for tail. The initial distribution is given by $\mu = [\mu_1, \mu_2]$. Find the probability of head and tail in the n-th step, expressed in terms of μ_1 and μ_2.

Solution:
Lets begin by making a discrete-time Markov chain with states H and T each representing Heads and Tails



Given in the question that $p_{HH} = 0$ and $p_{TT} = 0$ and we know that $p_{TH} + p_{HH} = 1$ and $p_{HT} + p_{TT} = 1$

So Now we get the following markov chain



For $n = 1$, the probabilities are:

$$p_1(H) = p_H \mu_2 \quad \text{and} \quad p_1(T) = p_T \mu_1$$

$$p_1(H) = \mu_2 \quad \text{and} \quad p_1(T) = \mu_1$$

For $n = 2$, the probabilities are:

$$p_2(H) = p_1(T) p_H + p_1(H) p_{HH} \quad \text{and} \quad p_2(T) = p_T p_1(H) + p_1(T) p_{TT}$$

On substituting,

$$p_2(H) = \mu_1 \quad \text{and} \quad p_2(T) = \mu_2$$

So in general we can write, For even steps ($n = 2k$), where $k \geq 1$, the probabilities are given by:

$$p_{2k}(H) = (p_T p_H)^k \mu_1 \quad \text{and} \quad p_{2k}(T) = (p_H p_T)^k \mu_2$$
$$p_{2k}(H) = \mu_1 \quad \text{and} \quad p_{2k}(T) = \mu_2$$

For odd steps ($n = 2k + 1$), where $k \geq 0$, the probabilities are given by:

$$P_{2k+1}(H) = (p_{2k}(T) p_H) \quad \text{and} \quad P_{2k+1}(T) = (p_{2k}(H) p_T)$$
$$P_{2k+1}(H) = \mu_2 \quad \text{and} \quad P_{2k+1}(T) = \mu_1$$

Therefore we get the general formulas for even and odd steps are:

$$p_{2k}(H) = \mu_1 \quad \text{and} \quad p_{2k}(T) = \mu_2$$
$$p_{2k+1}(H) = \mu_2 \quad \text{and} \quad p_{2k+1}(T) = \mu_1$$

where k ≥ 0

Grading Scheme:

* Even Case - 3.5 marks
* Odd Case - 3.5 marks

Problem 8b
- Pranav
Find the limiting distribution and stationary distribution π for Markov chains with the following transition probability matrix

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Now obtain the values of F_{ii}(probability of ever returning to state i having started in state i) for each of the 4 states and based on the values identify if each state is transient or recurrent. (8 marks)

Solution:
let $\pi = [\pi_1, \pi_2, \pi_3, \pi_4]$. Thus,

$$\pi = \pi P$$

Solving the above, we get the relations

$$\pi_1 = \pi_2$$

$$\sum_{k=1}^{4} \pi_k = 1 \text{ (As sum of probabilities should be equal to 1)}$$

Thus, π can be written as $[a, a, b, 1 - 2a - b]$ where $0 \leq a, b \leq 1$ and $2a + b \leq 1$ (2 marks)
The limiting distribution does not exists. This can be shown as as P does not change on multiplying by itself, i.e. $P^2 = P$. Thus the rows of P do not converge to a common distribution. (1 mark)
Now for states 3 and 4, clearly we come back to them after 1 step if we start in them. So, F_{ii} is 1 for them and they are recurrent. (2 marks)
Now $F_{ii} = \sum_{n=1}^{\infty} f_{ii}^n$ where f_{ii}^n represents the probability of coming back to state i for the first time in n steps. For states 1 and 2 :-

$$f_{ii}^n = (0.5)(0.5)^{n-1}$$

Thus,

$$\sum_{n=1}^{\infty} f_{ii}^n = 0.5 \sum_{n=0}^{\infty} 0.5^n = (0.5)2 = 1$$

Thus, states 1 and 2 are recurrent with $F_{ii} = 1$ (3 marks)
0.5 marks for giving the correct F_{ii} for each state.
0.5 marks for stating correct transience or recurrence.
0.5 marks for deriving the F_{ii} for state 1 and 2