Introduction to Data Science 1MS041

Group Assignment 3
By

Group 8
Feruz Redi
Moutushi Sen
Samiha Akhter
Shaheryar
Tebogo Mitane

UPPSALA
UNIVERSITET

Problem Statement 1

Consider a three state (1, 2, 3) Markov chain with transition matrix

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 \end{pmatrix}.$$

(a) Draw the transition diagram.

(b) Find the stationary distribution π.

(c) Given that the chain is in state 1 at time 1, what is the probability that the chain is in state 2 at time 4?

(d) Given that the chain is in state 1 at time 1, what is the expected time until the chain is in state 3 the first time?

(e) What is the period of each state?

Solution:

Part (a): Draw the Transition Diagram

To determine the period of each state in the given Markov chain, we'll analyze the possible return times to each state and compute the greatest common divisor (gcd) of these times. A state's period is defined as the gcd of all possible numbers of steps in which the chain can return to that state with a positive probability.

The transition matrix P is:

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 \end{pmatrix}$$

Understanding Periods:

The period $d(s)$ of a state s is:

$$d(s) = \gcd\{n \geq 1 : P_n(s, s) > 0\}$$

This means we need to find all n such that the probability of returning to state s in n steps is positive.

Computing Possible Return Times:

For each state, we'll compute $P_n(s, s)$ for $n = 1$ to a chosen maximum N. We'll collect all n where $P_n(s, s) > 0$.

Calculating Periods:

For each state's list of n, we'll compute the gcd of these numbers. If the gcd is 1, the state is aperiodic (period 1).

From this matrix, we can derive the transitions:

State 1:

- Possible return times (n): 1, 2, 3, ..., up to max_n.
- Since we can return to state 1 in any number of steps starting from 1, the gcd is 1.
- Stays in State 1 with probability 0.5.
- Moves to State 2 with probability 0.5.

State 2:

- Cannot return in 1 step because $P(2, 2) = 0$.
- Possible return times start from $n = 2$. Since we can return in steps $n = 2, 3, 4, ...$, the gcd is still 1.
- Moves to State 1 with probability 0.5.
- Moves to State 3 with probability 0.5.

State 3:

- Similarly to state 2, the possible return times start from $n = 2$.
- The gcd of these times is 1.

- Stays in state 3 with probability 0.5.
- Moves to state 1 with probability 0.5.

Conclusion:
The period of each state in the given Markov chain is 1. This means the chain is aperiodic, and every state can be revisited in any number of steps greater than or equal to some minimum, without being restricted to multiples of a higher number.

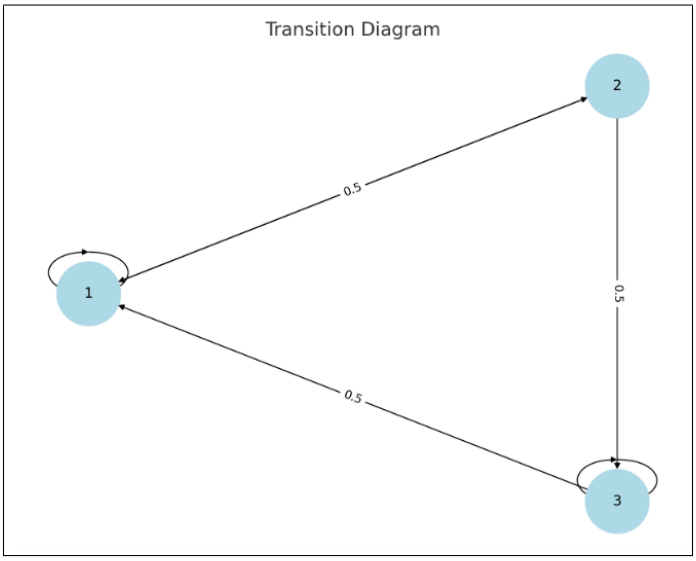

Figure 1: Transition Diagram for the Markov Chain

Part (b): Find the Stationary Distribution π

Let $\pi = (\pi_1, \pi_2, \pi_3)$ be the stationary distribution, satisfying $\pi = \pi P$ and $\pi_1 + \pi_2 + \pi_3 = 1$.

From $\pi = \pi P$:

$\pi_1 = 0.5\pi_1 + 0.5\pi_2 + 0.5\pi_3$

$\pi_2 = 0.5\pi_1$

$\pi_3 = 0.5\pi_2 + 0.5\pi_3$

Simplifying:

From equation 1: $0.5\pi_1 = 0.5\pi_2 + 0.5\pi_3$ $\pi_1 = \pi_2 + \pi_3$

From equation 2: $\pi_2 = 0.5\pi_1$

From equation 3: $0.5\pi_3 = 0.5\pi_2$ $\pi_3 = \pi_2$

Substituting $\pi_1 = 2\pi_2$ and $\pi_3 = \pi_2$ into $\pi_1 + \pi_2 + \pi_3 = 1$: $2\pi_2 + \pi_2 + \pi_2 = 1$ $4\pi_2 = 1$ $\pi_2 = \frac{1}{4}$

Therefore:

$\pi_1 = 2 \times \frac{1}{4} = \frac{1}{2}$, $\pi_3 = \frac{1}{4}$

Stationary Distribution: $\pi = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$

Part (c): Probability that the Chain is in State 2 at Time 4 Given it Started at State 1

We need to compute $P[X_4 = 2|X_1 = 1] = (P^3)_{1,2}$.

Firstly, compute P^2:

$$P^2 = P \times P = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$$

Then compute $P^3 = P^2 \times P$:

$$P^3 = P^2 \times P = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$$

Thus, $P[X_4 = 2|X_1 = 1] = (P^3)_{1,2} = 0.25$.

Part (d): Expected Time Until the Chain Reaches State 3 Starting from State 1

Let m_i be the expected hitting time from state i to state 3.

Set up the equations:
$m_3 = 0$ (since we're already at state 3) $m_1 = 1 + 0.5m_1 + 0.5m_2$ $m_2 = 1 + 0.5m_1$

Solving:

From equation 2: $0.5m_1 = 1 + 0.5m_2$

Substitute equation 3 into the above: $0.5m_1 = 1 + 0.5(1 + 0.5m_1)$

Simplify: $0.5m_1 = 1 + 0.5 + 0.25m_1$ $0.25m_1 = 1.5$ $m_1 = 6$

Answer: The expected time is 6 steps.

Part (e): Period of Each State

The period of a state is the greatest common divisor (gcd) of the lengths of possible cycles returning to that state.

For all states:
They can return to themselves in various numbers of steps without any missing intervals.Since the gcd of all possible return times is 1, each state is aperiodic.

Answer: Each state has a period of 1 (they are aperiodic).

Problem Statement 2

Assume that we are trying to classify a binary outcome Y, i.e., our data is of the form $(X, Y) \sim F_{X,Y}$, where $Y \in \{0, 1\}$ and $X \in \mathbb{R}^d$. We have used data to train a classifier $g(X)$. We can evaluate the performance of the classifier using i.i.d. testing data, $(X_1, Y_1), \ldots, (X_n, Y_n)$. We are interested in estimating the following quantities:

$$\text{Precision: } \mathbb{P}(Y = 1 \mid g(X) = 1),$$

$$\text{Recall: } \mathbb{P}(g(X) = 1 \mid Y = 1).$$

(a) Write down the empirical version of the precision and recall.

(b) Let us now think that the variable Y denotes if a battery's health has deteriorated or not, and let X denote a bunch of constructed health indicators about the battery. If the model $g(X)$ predicts that the battery has deteriorated, you need to run a test to confirm this. The cost of running the test is c when the battery is not deteriorated. On the other hand, if the battery is in fact deteriorated and the test is not run, the battery will die during use, and the cost of this is d. Define a random variable representing the cost of the decision $g(X)$ and write down the formula for the expected cost in terms of the precision and recall.

(c) **Advanced Question:** Can you produce a confidence interval for the expected cost? What about the precision and recall?

Solution:

Part (a): Empirical Versions of Precision and Recall

The empirical versions of Precision and Recall as follows:

Precision:

$$\text{Precision}_{\text{empirical}} = \frac{\sum_{i=1}^{n} \mathbf{1}\{g(X_i) = 1 \text{ and } Y_i = 1\}}{\sum_{i=1}^{n} \mathbf{1}\{g(X_i) = 1\}}, \tag{1}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function, which outputs 1 if the condition is true and 0 otherwise.

Recall:

$$\text{Recall}_{\text{empirical}} = \frac{\sum_{i=1}^{n} \mathbf{1}\{g(X_i) = 1 \text{ and } Y_i = 1\}}{\sum_{i=1}^{n} \mathbf{1}\{Y_i = 1\}}. \tag{2}$$

Part (b): Expected Cost Formula

Cost of Decision: The cost of a decision depends on the outcomes:

$$C(g(X), Y) = \begin{cases} c & \text{if } g(X) = 1 \text{ and } Y = 0 \text{ (false positive)}, \\ d & \text{if } g(X) = 0 \text{ and } Y = 1 \text{ (false negative)}, \\ 0 & \text{otherwise (true positive/negative)}. \end{cases}$$

Expected Cost: The expected cost of the classifier is given as:

$$\mathbb{E}[C(g(X), Y)] = c \cdot P(g(X) = 1 \text{ and } Y = 0) + d \cdot P(g(X) = 0 \text{ and } Y = 1). \tag{3}$$

In Terms of Precision and Recall: Using the definitions of Precision and Recall:

$$P(g(X) = 1 \text{ and } Y = 0) = P(g(X) = 1)(1 - \text{Precision}),$$

$$P(g(X) = 0 \text{ and } Y = 1) = P(Y = 1)(1 - \text{Recall}).$$

Let $\pi = P(Y = 1)$ be the proportion of deteriorated batteries, and note that $P(g(X) = 1) = \pi \cdot \text{Recall/Precision}$. Substituting these:

$$\mathbb{E}[C(g(X), Y)] = c \cdot P(g(X) = 1)(1 - \text{Precision}) + d \cdot \pi(1 - \text{Recall}). \tag{4}$$

Part (c): Confidence Intervals for Metrics and Cost

Confidence Intervals for Precision and Recall: Precision and Recall are proportions and can be approximated using the Central Limit Theorem. The confidence intervals are:

Precision:

$$\text{CI for Precision} = \hat{\text{Precision}} \pm z\sqrt{\frac{\hat{\text{Precision}}(1 - \hat{\text{Precision}})}{n_1}}, \tag{5}$$

where $n_1 = \sum_{i=1}^{n} \mathbf{1}\{g(X_i) = 1\}$ and z is the z-score for the desired confidence level.

Recall:

$$\text{CI for Recall} = \hat{\text{Recall}} \pm z\sqrt{\frac{\hat{\text{Recall}}(1 - \hat{\text{Recall}})}{n_2}}, \tag{6}$$

where $n_2 = \sum_{i=1}^{n} \mathbf{1}\{Y_i = 1\}$.

Confidence Interval for Expected Cost: Using the delta method to approximate the variance of the expected cost:

$$\hat{\mathbb{E}}[C(g(X), Y)] = c \cdot \hat{P}(g(X) = 1)(1 - \hat{\text{Precision}}) + d \cdot \hat{\pi}(1 - \hat{\text{Recall}}),$$

$$\text{Var}[\hat{\mathbb{E}}[C(g(X), Y)]] \approx c^2 \cdot \text{Var}[\hat{P}(g(X) = 1)](1 - \hat{\text{Precision}})^2 + d^2 \cdot \text{Var}[\hat{\pi}](1 - \hat{\text{Recall}})^2.$$

The confidence interval for the expected cost is:

$$\text{CI for Expected Cost} = \hat{\mathbb{E}}[C(g(X), Y)] \pm z \cdot \sqrt{\text{Var}[\hat{\mathbb{E}}[C(g(X), Y)]]}. \tag{7}$$

Problem Statement 3

Let $\mathbf{X}$ and $\mathbf{Y}$ be two d-dimensional zero-mean, unit-variance Gaussian random vectors. Show that $\mathbf{X}$ and $\mathbf{Y}$ are nearly orthogonal by calculating their dot product. Can you, for instance, also bound the probability that the dot product is larger than ε?

Solution:

The Dot Product of X and Y

Given that $\mathbf{X}$ and $\mathbf{Y}$ are zero-mean, unit-variance Gaussian random d-dimensional vectors, we aim to show that they are nearly orthogonal by calculating the dot product $\mathbf{X} \cdot \mathbf{Y}$.

Dot Product Definition

$$\mathbf{X} \cdot \mathbf{Y} = \sum_{i=1}^{d} X_i Y_i$$

where X_i and Y_i are the components of $\mathbf{X}$ and $\mathbf{Y}$, respectively.

Expectation of the Dot Product

Since $\mathbf{X}$ and $\mathbf{Y}$ are zero-mean (i.e., $E[X_i] = 0$ and $E[Y_i] = 0$), we can calculate the expectation of their dot product. Assuming $\mathbf{X}$ and $\mathbf{Y}$ are independent, each term $X_i Y_i$ will have zero expectation:

$$E[\mathbf{X} \cdot \mathbf{Y}] = E\left[\sum_{i=1}^{d} X_i Y_i\right] = \sum_{i=1}^{d} E[X_i Y_i] = 0$$

This shows that the dot product has a mean of zero, suggesting that $\mathbf{X}$ and $\mathbf{Y}$ are likely to be nearly orthogonal in expectation.

Variance of the Dot Product

To understand the typical size of $\mathbf{X} \cdot \mathbf{Y}$, we calculate the variance of the dot product. The variance of $\mathbf{X} \cdot \mathbf{Y}$ will indicate how large the dot product can be.

Variance Calculation

$$\text{Var}(\mathbf{X} \cdot \mathbf{Y}) = \text{Var}\left(\sum_{i=1}^{d} X_i Y_i\right)$$

Since $\mathbf{X}$ and $\mathbf{Y}$ are independent and each X_i and Y_i have unit variance, the variance of each term $X_i Y_i$ is:

$$\text{Var}(X_i Y_i) = E[X_i^2] \cdot E[Y_i^2] = 1 \cdot 1 = 1$$

Summing the Variances

Assuming the terms are independent for different i, we can sum the variances:

$$\text{Var}\left(\sum_{i=1}^{d} X_i Y_i\right) = \sum_{i=1}^{d} \text{Var}(X_i Y_i) = d$$

Therefore, $\text{Var}(\mathbf{X} \cdot \mathbf{Y}) = d$, which quantifies the typical fluctuation around zero.

Bounding the Probability of Large Dot Products

Now we want to bound the probability that the dot product $\mathbf{X} \cdot \mathbf{Y}$ is larger than some threshold ε.

Using Chebyshev's Inequality:

Chebyshev's inequality states that for a random variable Z with mean μ and variance δ^2,

$$\Pr(|Z - \mu| \geq k\delta) \leq \frac{1}{k^2}$$

In our case, $Z = \mathbf{X} \cdot \mathbf{Y}$, $\mu = 0$, and $\delta^2 = d$. So,

$$\Pr(|\mathbf{X} \cdot \mathbf{Y}| \geq \varepsilon) \leq \frac{\text{Var}(\mathbf{X} \cdot \mathbf{Y})}{\varepsilon^2} = \frac{d}{\varepsilon^2}$$

This inequality bounds the probability that the dot product $\mathbf{X} \cdot \mathbf{Y}$ deviates significantly from zero. Using Chebyshev's inequality, we bound the probability that the dot product $\mathbf{X} \cdot \mathbf{Y}$ is larger than ε by $\frac{d}{\varepsilon^2}$.

Problem Statement 4

Let $u_1, \ldots, u_r$ be $n \times 1$ unit length vectors that are linearly independent, i.e., $\sum_{i=1}^{r} \alpha_i u_i = 0$ implies that $\alpha_i = 0$ for all i.

(a) Verify that the matrix $u_i u_i^T$ is a rank-one matrix for all i. What is the null-space and range of $u_i u_i^T$?

(b) Verify that the matrix $U = \sum_{i=1}^{r} u_i u_i^T$ is a rank-r matrix.

(c) (i) If we perform SVD on U, are the vectors $u_1, \ldots, u_r$ the same as the right singular vectors? If not, can you give an example?

(ii) What if the vectors $u_1, \ldots, u_r$ are all orthogonal? In this case, what are the singular values of U?

Solution:

Part (a): Verify the Rank of Matrix to be 1

The matrix $u_i u_i^T$ is an $n \times n$ matrix formed by the outer product of u_i with its transpose:

$$u_i u_i^T = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{in} \end{bmatrix} \begin{bmatrix} u_{i1} & u_{i2} & \cdots & u_{in} \end{bmatrix}$$

Each element (j, k) of $u_i u_i^T$ is given by:

$$(u_i u_i^T)_{jk} = u_{ij} u_{ik}.$$

Since $u_i u_i^T$ is the outer product of u_i, every column of $u_i u_i^T$ is a scalar multiple of u_i. This means all columns are linearly dependent, showing that $u_i u_i^T$ is rank one.

Range of $u_i u_i^T$ The range of $u_i u_i^T$ is the space spanned by u_i, which is a 1-dimensional.

$$\text{Range}(u_i u_i^T) = \text{span}\{u_i\}.$$

Null-Space: The null-space consists of all vectors $v \in \mathbb{R}^n$ such that $u_i u_i^T v = 0$. Since $u_i u_i^T v = (u_i^T v) u_i$, this is zero if and only if $u_i^T v = 0$ (i.e., v is orthogonal to u_i).

$$\text{Null}(u_i u_i^T) = \{v \in \mathbb{R}^n : u_i^T v = 0\}.$$

The null space is the orthogonal complement of $\text{span}\{u_i\}$, which is an $(n-1)$-dimensional subspace of $\mathbb{R}^n$.

Part (b): Verify the Rank of Matrix to be r

$$U = u_1 u_1^T + u_2 u_2^T + \cdots + u_r u_r^T.$$

where $u_1, u_2, \ldots, u_r$ are $n \times 1$ unit-length vectors that are linearly independent. Since $u_i u_i^T$ is a rank-one matrix , U is a sum of r rank-one matrices. For Range of U: it is given by the sum of the ranges of $u_i u_i^T$. Since:

$$\text{Range}(u_i u_i^T) = \text{Span}(\{u_i\}),$$

The range of U is:

$$\text{Range}(U) = \text{Span}(\{u_1, u_2, \ldots, u_r\}).$$

This is an r-dimensional subspace of $\mathbb{R}^n$.

Null-space of U: is the set of all vectors $v \in \mathbb{R}^n$ such that:

$$Uv = \sum_{i=1}^{r} u_i u_i^T v = 0.$$

For this to hold, each term $u_i u_i^T v = 0$, this $u_i^T v = 0$ for all i. Thus, v must be orthogonal to the span of $u_1, u_2, \ldots, u_r$.

The null-space U is, therefore, the orthogonal complement of the span of $u_1, u_2, \ldots, u_r$, which has dimension $n - r$.

So rank of U can be verified by rank-nullity theorem

$$\text{Rank}(U) + \text{Nullity}(U) = n.$$

Since the null-space of U has dimension $n - r$, the rank of U must be:

$$\text{Rank}(U) = r.$$

.

Part (c1): Does SVD on U Changes the Right Singular Vectors?

In SVD, the right singular vectors are the eigenvectors of $U^T U = U^2$. Since U is symmetric ($U = U^T$), the eigenvectors of U^2 are the same as those of U. Therefore, the eigenvectors of U are used as both the left and right singular vectors in SVD.

But the vectors $u_1, \ldots, u_r$ are not necessarily the eigenvectors of U. The eigenvectors of U are orthogonal, while the u_i's are not always orthogonal. So, $u_1, \ldots, u_r$ are not necessarily the right singular vectors of U.

Example

Consider an example with $n = 2$ and $r = 2$. Let:

$$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Compute $U = u_1 u_1^T + u_2 u_2^T$:

$$U = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$U = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

The Eigenvalues are:

$$\lambda_1 = \frac{-\sqrt{5} + 3}{2}$$

$$\lambda_2 = \frac{\sqrt{5} + 3}{2}$$

The Eigenvectors:

$$v_1 = \begin{pmatrix} \frac{-\sqrt{5}+1}{2} \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} \frac{\sqrt{5}+1}{2} \\ 1 \end{pmatrix}$$

- The u_i's (u_1, u_2) are not orthogonal, but the eigenvectors (v_1, v_2) are orthogonal. - The right singular vectors are the eigenvectors (v_1, v_2), not the original u_1, u_2.

Part (c2): Singular VAlues of U for Orthogonal Vectors

Since the vectors $u_i u_i^T$ are orthogonal and unit-length the eigenvectors of U are exactly $u_1, u_2, \ldots, u_r$, corresponding to the directions of the projections. Each $u_i u_i^T$ contributes an eigenvalue of 1 to the eigenvalue corresponding to the vector u_i. The remaining eigenvectors, which are orthogonal to $u_1, \ldots, u_r$, correspond to eigenvalues of 0, as these vectors lie in the null space of U.

We know that:

$$u_i^T u_j = 0 \quad \text{for} \quad i \neq j, \quad \text{and} \quad u_i^T u_i = 1 \quad \text{for all} \quad i$$

These properties ensure that the matrix U is diagonalizable, and its eigenvalues are:

- 1, with multiplicity r,
- 0, with multiplicity $n - r$, where n is the dimension of the vector space.

Since the singular values of a matrix are the square roots of its non-zero eigenvalues, the singular values of U are:

- 1, with multiplicity r,
- 0, with multiplicity $n - r$, where n is the dimension of the vector space.

Thus, the singular values of U are all 1 for the r non-zero eigenvalues.

Problem Statement 5

Let $X \sim \text{Uniform}(B_1)$ and define $Y = \|X\|_2$, the Euclidean norm of X.

1. Find the distribution function of Y.
2. What is the distribution of $\ln\left(\frac{1}{Y}\right)$?
3. Calculate $\mathbb{E}\left[\ln\left(\frac{1}{Y}\right)\right]$, first by using the distribution function of Y and then by using the distribution function of $\ln\left(\frac{1}{Y}\right)$.

Solution:

Part (a): Distribution Function of Y

Breaking the Problem Down

The unit ball B_1 in $\mathbb{R}^n$ is defined as $B_1 = \{X \in \mathbb{R}^n : \|X\|_2 \leq 1\}$. Since X is uniformly distributed over B_1, the probability density function (PDF) is constant within B_1 and zero outside. We aim to find the cumulative distribution function (CDF) of $Y = \|X\|_2$.

Finding the CDF of Y

The CDF $F_Y(y)$ is given by:

$$F_Y(y) = P(Y \leq y) = P(\|X\|_2 \leq y), \quad \text{for } y \in [0, 1].$$

Calculating the Probability

Since X is uniformly distributed over B_1, the probability $P(\|X\|_2 \leq y)$ is the ratio of the volume of a ball with radius y to the volume of the unit ball.

$$F_Y(y) = \frac{\text{Volume of ball with radius } y}{\text{Volume of unit ball}} = \frac{V_n(y)}{V_n(1)},$$

where $V_n(r)$ is the volume of an n-dimensional ball of radius r.

Volume of an n-Dimensional Ball

The volume of an n-dimensional ball of radius r is given by:

$$V_n(r) = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} r^n,$$

where Γ is the Gamma function.

Simplifying the CDF

Since $V_n(r) \propto r^n$, the constants cancel out in the ratio:

$$F_Y(y) = \frac{V_n(y)}{V_n(1)} = \frac{\left(\frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} y^n\right)}{\left(\frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} \cdot 1^n\right)} = y^n.$$

Finally

The distribution function of Y is:

$$F_Y(y) = y^n, \quad \text{for } y \in [0,1].$$

Part (b): Distribution of $\ln\left(\frac{1}{Y}\right)$

Defining the New Variable

Let $W = \ln\left(\frac{1}{Y}\right)$. Our goal is to find the distribution of W, specifically its cumulative distribution function $F_W(w)$.

Finding the CDF of W

$$\begin{aligned} F_W(w) &= P(W \leq w) \\ &= P\left(\ln\left(\frac{1}{Y}\right) \leq w\right) \\ &= P\left(\frac{1}{Y} \leq e^w\right) \\ &= P\left(Y \geq e^{-w}\right). \end{aligned}$$

Relating to the CDF of Y

Since $F_Y(y) = y^n$, the probability $P(Y \leq y) = y^n$. Therefore:

$$P(Y \geq e^{-w}) = 1 - P(Y < e^{-w}) = 1 - F_Y(e^{-w}) = 1 - \left(e^{-w}\right)^n.$$

Simplifying the CDF of W

Thus, the CDF of W is:

$$F_W(w) = 1 - e^{-nw}, \quad \text{for } w \geq 0.$$

Identifying the Distribution

The function $F_W(w) = 1 - e^{-nw}$ is the cumulative distribution function of an exponential distribution with rate parameter n. Therefore, W follows an exponential distribution.

Lastly

$\ln\left(\frac{1}{Y}\right)$ is exponentially distributed with rate parameter n:

$$\ln\left(\frac{1}{Y}\right) \sim \text{Exponential}(n).$$

Part (c): Calculating $\mathbb{E}\left[\ln\left(\frac{1}{Y}\right)\right]$

Method 1: Using the Distribution of Y

Finding the PDF of Y

The probability density function (pdf) $f_Y(y)$ is the derivative of $F_Y(y)$:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}y^n = ny^{n-1}, \quad y \in (0,1).$$

Setting Up the Expected Value

The expected value $\mathbb{E}\left[\ln\left(\frac{1}{Y}\right)\right]$ is:

$$\mathbb{E}\left[\ln\left(\frac{1}{Y}\right)\right] = \int_0^1 \ln\left(\frac{1}{y}\right) f_Y(y)\, dy = \int_0^1 \ln\left(\frac{1}{y}\right) ny^{n-1}\, dy.$$

Simplifying the Integral

Note that $\ln\left(\frac{1}{y}\right) = -\ln y$. So, the integral becomes:

$$\mathbb{E}\left[\ln\left(\frac{1}{Y}\right)\right] = -n\int_0^1 \ln y \cdot y^{n-1}\, dy.$$

Evaluating the Integral

We need to compute:

$$I = \int_0^1 \ln y \cdot y^{n-1}\, dy.$$

Using a Standard Integral:
Recall that for $a > -1$:

$$\int_0^1 x^a \ln x\, dx = -\frac{1}{(a+1)^2}.$$

Here, $a = n - 1$. Therefore:

$$I = -\frac{1}{n^2}.$$

Calculating the Expected Value

Substitute back into the expected value:

$$\mathbb{E}\left[\ln\left(\frac{1}{Y}\right)\right] = -n\left(-\frac{1}{n^2}\right) = \frac{1}{n}.$$

Method 2: Using the Distribution of $\ln\left(\frac{1}{Y}\right)$

Recognizing the Distribution

From Part (b), we know that $W = \ln\left(\frac{1}{Y}\right) \sim \text{Exponential}(n)$.

Using the Expected Value of an Exponential Distribution

The expected value of an exponential random variable with rate parameter λ is:

$$\mathbb{E}[W] = \frac{1}{\lambda}.$$

Given $\lambda = n$:

$$\mathbb{E}\left[\ln\left(\frac{1}{Y}\right)\right] = \frac{1}{n}.$$

Ultimately

Using both methods, the expected value is:

$$\mathbb{E}\left[\ln\left(\frac{1}{Y}\right)\right] = \frac{1}{n}.$$
