NATIONAL UNIVERSITY OF SINGAPORE

Mathematics PhD Qualifying Exam Paper 4
Stochastic Processes and Machine Learning

January 2025

Time allowed : 3 hours

INSTRUCTIONS TO CANDIDATES

1. Please write your matriculation/student number only. Do not write your name.
2. Including this page, this examination paper comprises **5** printed pages.
3. At the top right corner of every page of your answer script, write the question and page numbers(eg. Q1 P1, Q1 P2, Q2 P1,. . .).
4. This examination contains **EIGHT (8)** questions. Answer **ALL** questions. **Properly justify your answers.**
5. There is a total of **ONE HUNDRED (100)** points. The points for each question are indicated at the beginning of the question.
6. Please start each part of a question (i.e., (a), (b), etc.) on a new page.
7. This is a CLOSED BOOK examination. The use of a double-sided A4-size cheat-sheet is allowed. No electronic device (such as calculator, tablet, laptop or phone) is allowed. You need to have your reference materials in hard copy with you.
8. A list containing information on the probability density / mass function, mean, variance and moment generating functions of some common distributions has been provided at the end of this exam paper for possible consultation.

Q 1 **[10 points]**

Consider the one-dimensional lattice of positive integers $\{1, 2, 3, \ldots\}$. At each site $x \geq 1$, independently place either *one ball* or *no ball* according to a Bernoulli probability distribution with mean $p \in (0, 1/2)$.

A boy starts at site 0 with an initial (random) number of balls in his backpack, denoted by $B(0)$. The random variable $B(0)$ is distributed according to a probability measure μ on $\{0, 1, 2, \ldots\}$. The boy visits the sites $1, 2, 3, \ldots$ in order. At each site x:

(a) If site x contains a ball, the boy picks it up and places it into his backpack.

(b) If site x is empty and his backpack is *not* empty, he removes one ball from his backpack and deposits it on the site.

Define $B(x)$ to be the number of balls in the boy's backpack *after* he visits site x. The process $\{B(x)\}_{x=0,1,2,\ldots}$ is therefore a Markov chain taking values in $\{0, 1, 2, \ldots\}$. Find the *invariant distribution* of the Markov chain $\{B(x)\}_{x=0,1,2,\ldots}$.

Q 2 **[20 points]** Let N be a positive integer, and let $x, y \in \mathbb{R}$. Let μ be a probability distribution on $\mathbb{R}$. A *random excursion* of length N from x to y with increment distribution μ is defined as a discrete-time stochastic process $\{B_n\}_{n=0}^{N}$ that satisfies:

- $B_0 = x$ and $B_N = y$ almost surely, i.e., the process starts at x and ends at y after N steps.
- For each $n = 1, 2, \ldots, N$, the increment $B_n - B_{n-1}$ is distributed according to μ, independently of the other increments, but *conditioned* on the event $\{B_0 = x,\, B_N = y\}$.

In the following we will set $\mu = \mathcal{N}(0, 1)$.

(a) **(7 points)** Write the joint distribution of the N-steps random excursion $\{B_n\}_{n=0,\ldots,N}$ from x to y and normally distributed increments.

(b) **(7 points)** Describe a sampling procedure for a 2-step random excursion with normally distributed increments.

(c) **(6 points)** Set $N = 2^k$ for some positive integer k. Describe a sampling procedure for a N-steps random excursion from x to y with normally distributed increments.

Q 3 **[10 points]** Let σ be a uniform random permutation of $\{1, \ldots, n\}$. Let E_k be the event that the cycle containing 1 has length exactly equal to k.

(a) **(5 points)** For any fixed $1 \leq k \leq n$ compute $\mathbb{P}(E_k)$.

(b) **(5 points)** Let B_n be the event that the permutation σ has no cycles of length larger than $\lfloor n/2 \rfloor$. Compute

$$\lim_{n \to +\infty} \mathbb{P}(B_n).$$

Q 4 **[10 points]** Let $\mathcal{P} = \{0 = p_0 < p_1 < p_2 < \cdots\}$ be a Poisson Point Process on $[0, \infty)$ and fix $\varepsilon \in (0, 1)$.

(a) **(5 points)** For any $T > 1$ define the event $E_{\varepsilon,T}$ that all points of $\mathcal{P}$ in the interval $[0, T)$ are spaced by more than ε. Compute

$$\lim_{T \to +\infty} \frac{1}{T} \log \mathbb{P}\left(E_{\varepsilon,T}\right).$$

(b) **(5 points)** Consider a partition of the interval $[0, T)$ as

$$[0, T) = \bigcup_{i=1,\ldots,T/\varepsilon} I_i, \qquad \text{with } I_i = [(i-1)\varepsilon, i\varepsilon)$$

Similarly to the previous point define the event $F_{\varepsilon,T}$ that no two consecutive intervals I_i, I_{i+1} contain points of $\mathcal{P}$. Compute

$$\lim_{T \to +\infty} \frac{1}{T} \log \mathbb{P}\left(F_{\varepsilon,T}\right).$$

Q 5 [Graph kernel] **[10 points]** Let $G = (\mathcal{V}, \mathcal{E})$ be an undirected graph with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$. $\mathcal{V}$ could represent a set of documents or biosequences and $\mathcal{E}$ the set of connections between them. Let $w[e] \in \mathbb{R}$ denote the weight assigned to edge $e \in \mathcal{E}$. The weight of a path is the product of the weights of its constituent edges. Show that the kernel K over $\mathcal{V} \times \mathcal{V}$ where $K(p, q)$ is the sum of the weights of all paths of length two between p and q is positive definite symmetric.

Q 6 [Double centering in Isomap] **[15 points]** Let $\boldsymbol{X}$ be an uncentered data matrix and let $\bar{\boldsymbol{x}} = \frac{1}{m}\sum_i \boldsymbol{x}_i$ be the sample mean of the columns of $\boldsymbol{X}$. Define $\boldsymbol{X}^*$ as the centered version of $\boldsymbol{X}$, that is, let $\boldsymbol{x}_i^* = \boldsymbol{x}_i - \bar{\boldsymbol{x}}$ be the ith column of $\boldsymbol{X}^*$. Let $\boldsymbol{K} = \boldsymbol{X}^\top \boldsymbol{X}$, and let $\boldsymbol{D}$ denote the Euclidean distance matrix, i.e., $\boldsymbol{D}_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$.

(a) **(5 points)** Show that $\boldsymbol{K}_{ij} = \frac{1}{2}(\boldsymbol{K}_{ii} + \boldsymbol{K}_{jj} + \boldsymbol{D}_{ij}^2)$.

(b) **(5 points)** Show that $\boldsymbol{K}_{ij}^* = -\frac{1}{2}\left[\boldsymbol{D}_{ij}^2 - \frac{1}{m}\sum_{k=1}^m \boldsymbol{D}_{ik}^2 - \frac{1}{m}\sum_{k=1}^m \boldsymbol{D}_{kj}^2 + \bar{\boldsymbol{D}}\right]$, where $\boldsymbol{K}^* = \boldsymbol{X}^{*\top}\boldsymbol{X}^*$ and $\bar{\boldsymbol{D}} = \frac{1}{m^2}\sum_u \sum_v \boldsymbol{D}_{uv}^2$ is the mean of the m^2 entries in $\boldsymbol{D}$.

(c) **(5 points)** Show that $\boldsymbol{K}^* = -\frac{1}{2}\boldsymbol{H}\boldsymbol{D}\boldsymbol{H}$, where $\boldsymbol{H} = \boldsymbol{I}_m - \frac{1}{m}\mathbf{1}\mathbf{1}^\top$ and $\mathbf{1}$ is the column vector whose elements are 1.

Q 7 [Stopping Strategy] **[15 points]** A fair six sided dice is rolled repeatedly and you observe outcomes sequentially. Formally, dice roll outcomes are independently and uniformly sampled from the set $\{1, 2, 3, 4, 5, 6\}$. At every time step before the hth roll you can-choose between two actions:

Stop: stop and receive a reward equal to the number shown on the dice or,

Roll: roll again and receive no immediate reward.

If not having stopped before then, at time step h (which would be reached after $h-1$ rolls) you are forced to take the action Stop, you receive the corresponding reward and the game ends.

We will model the game as a finite horizon MDP with six states and two actions. The state at time step k corresponds to the number shown on the dice at the kth roll. Assume that the discount factor, γ, is 1.

(a) **(2 points)** The value function at time step h, when it is no longer possible to roll the dice again, is $V^h(1) = 1, V^h(2) = 2, \ldots, V^h(6) = 6$. Compute the value function at time step $h-1$.

(b) **(3 points)** Express the value function at time step $k-1$, with $2 < k \leq h$ recursively in terms of the value function at roll k, so in terms of $V^k(1), V^k(2), \ldots, V^k(6)$.

The Q function at time step k for action "Roll" does not depend on the state since the number shown by the dice is irrelevant once you decided to roll. We use the shorthand notation $q(k) = Q^k$ (state, "Roll") since the only dependence is on k.

(c) **(2 points)** Compute $q(h-1)$.

(d) **(3 points)** Express $q(k-1)$ recursively as a function of $q(k)$, with $2 < k \leq h$.

(e) **(5 points)** What is the optimal policy $\pi^k(s)$ at roll k as a decision rule based on the current state s and $q(k)$?

Q 8 [Random Fourier Features] **[10 points]**

Given the Gaussian kernel $k(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{y})^\top(\boldsymbol{x}-\boldsymbol{y})\right)$ where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^D$. Find a mapping $\boldsymbol{z} : \mathbb{R}^D \to \mathbb{R}^R$, where ideally $R \ll N$, such that:

$$k(\boldsymbol{x}, \boldsymbol{y}) \approx \boldsymbol{z}(\boldsymbol{x})^\top \boldsymbol{z}(\boldsymbol{y}).$$

—— **End of Paper** ——

- Bernoulli (p) :
  $\mathbb{P}(X=i) = \begin{cases} p \text{ if } i=1 \\ 1-p \text{ if } i=0. \end{cases}$
  $\mathbb{E}[X]=p, \quad \mathrm{Var}[X]=p(1-p), \quad \mathbb{E}[e^{tX}]=(1-p)+pe^t.$

- Binomial (n,p):
  $\mathbb{P}(X=i)=\binom{n}{i}p^i(1-p)^{n-i}; 0\le i\le n.$
  $\mathbb{E}[X]=np, \quad \mathrm{Var}[X]=np(1-p), \quad \mathbb{E}[e^{tX}]=[(1-p)+pe^t]^n.$

- Geometric (p) :
  $\mathbb{P}(X=i)=(1-p)^{i-1}p; i\ge 1.$
  $\mathbb{E}[X]=\frac{1}{p}, \quad \mathrm{Var}[X]=\frac{1-p}{p^2}, \quad \mathbb{E}[e^{tX}]=\frac{pe^t}{1-(1-p)e^t}$ for $t<-\log(1-p).$

- Poisson (λ):
  $\mathbb{P}(X=i)=e^{-\lambda}\frac{\lambda^i}{i!}; i\ge 1.$
  $\mathbb{E}[X]=\lambda, \quad \mathrm{Var}[X]=\lambda, \quad \mathbb{E}[e^{tX}]=\exp(\lambda(e^t-1)).$

- Uniform (a,b) :
  $f(x) = \begin{cases} \frac{1}{b-a} \text{ if } a\le x\le b \\ 0 \text{ otherwise }. \end{cases}$
  $\mathbb{E}[X]=(a+b)/2, \quad \mathrm{Var}[X]=\frac{(b-a)^2}{12}, \quad \mathbb{E}[e^{tX}]=\frac{e^{tb}-e^{ta}}{t(b-a)}$ if $t\neq 0.$

- Uniform on the square $(a,b)\times(c,d)$:
  $f(x,y) = \begin{cases} \frac{1}{(b-a)(d-c)} \text{ if } a\le x\le b, c\le y\le d \\ 0 \text{ otherwise }. \end{cases}$

- Normal / Gaussian ($N(\mu,\sigma^2)$):
  $f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2}).$
  $\mathbb{E}[X]=\mu, \quad \mathrm{Var}[X]=\sigma^2, \quad \mathbb{E}[e^{tX}]=\exp(\mu t+\frac{1}{2}\sigma^2t^2).$

- Exponential (λ):
  $f(x) = \begin{cases} \lambda\exp(-\lambda x) \text{ if } x>0 \\ 0 \text{ otherwise}. \end{cases}$
  $\mathbb{E}[X]=1/\lambda, \quad \mathrm{Var}[X]=1/\lambda^2, \quad \mathbb{E}[e^{tX}]=\frac{\lambda}{\lambda-t}$ for $t<\lambda.$