NATIONAL UNIVERSITY OF SINGAPORE

Mathematics PhD Qualifying Exam Paper 4
Stochastic Processes and Machine Learning

August 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 two independent simple symmetric random walks $\{X_n\}_{n\geq 0}$ and $\{Y_n\}_{n\geq 0}$ on the integer lattice with initial conditions $X_0 = 0$ and $Y_0 = a > 0$ respectively. Define the stopping time τ by:

$$\tau = \inf\{n \geq 0 : X_n = Y_n\}$$

(a) Show that the difference $Y_n - X_n$ is itself a symmetric random walk starting at a.

(b) Using part (a), compute the probability that the random walks ever meet, i.e. $\mathbb{P}(\tau < \infty)$.

(c) Calculate the expected value $\mathbb{E}[\tau]$ as a function of a.

Q 2 [10 points]

Let $\{X_k\}_{k\geq 1}$ be a sequence of i.i.d. exponential random variables of parameter 1, i.e. $\mathbb{P}[X_1 > x] = e^{-x}$, for all $x > 0$. Let $M_n = \max\{X_1, \ldots, X_n\}$ be the running maximum of the family of random variables $\{X_k\}_{k\geq 1}$.

(a) Show that

$$\limsup_n \frac{X_n}{\log n} \leq 1 \qquad \text{a.s.}$$

and that

$$\lim_{n\to\infty} \frac{M_n}{\log n} \leq 1.$$

(b) Define the random variable $G_n = M_n - \log n$. Show that G_n converges weakly to a random variable G. Moreover find the cumulative distribution function of G.

Q 3 [15 points]

For a natural number $n \geq 1$, consider S_n to be the set of permutations of n elements $\{1, \ldots, n\}$ and equip S_n with the uniform probability measure

(a) A *fixed point* of a permutation $\pi = (\pi_1, \ldots, \pi_n)$ is a value $i \in \{1, \ldots, n\}$ such that $\pi_i = i$ and we define the function $F(\pi)$ as

$$F(\pi) = \text{number of fixed points in } \pi.$$

Compute the expectation $\mathbb{E}[F(\pi)]$.

(b) A *transposition* of a permutation $\pi = (\pi_1, \ldots, \pi_n)$ is a pair of different values $i, j \in \{1, \ldots, n\}$ such that $\pi_i = j$ and $\pi_j = i$. Define the function

$$T(\pi) = \text{number of transpositions in } \pi.$$

Compute the expectation $\mathbb{E}[T(\pi)]$.

Q 4 **[15 points]**

Let β be a positive random variable such that $\mathbb{E}(\beta)$ and $\mathbb{E}(\beta^2)$ are both well defined. Let $B = \{b_i\}_{i\geq 0}$ be the point process constructed setting $b_0 = 0$ and such that the gaps $\{b_{i+i} - b_i\}_{i\geq 0}$ are i.i.d. random variables with $b_1 - b_0 \sim \beta$. For any $T > 0$ define

$$N(T) = \max\{n : b_n < T\}.$$

(a) Prove that the random variable $\frac{N(T)}{T}$ converges almost surely and determine its limit.

(b) Let $p_T \sim \text{Unif}(0, T)$ be a point sampled uniformly at random in the interval $(0, T)$ and define the random variable

$$W_T = \inf\{t > 0 : p_T + t \in B\},$$

that is, the gap between the point p_T and the nearest point in B to its right. Compute the limit

$$\lim_{T\to\infty} \mathbb{E}[W_T].$$

Q 5 [Reducing elastic net to lasso] **[10 points]**

Define

$$J_1(\boldsymbol{w}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|^2 + \lambda_2\|\boldsymbol{w}\|_2^2 + \lambda_1\|\boldsymbol{w}\|_1$$

and

$$J_2(\boldsymbol{w}) = \|\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{w}\|^2 + c\lambda_1\|\boldsymbol{w}\|_1$$

where $\|\boldsymbol{w}\|^2 = \|\boldsymbol{w}\|_2^2 = \sum_i w_i^2$ is the squared 2-norm, $\|\boldsymbol{w}\|_1 = \sum_i |w_i|$ is the 1-norm, $c = (1 + \lambda_2)^{-\frac{1}{2}}$, and

$$\tilde{\boldsymbol{X}} = c\begin{pmatrix} \boldsymbol{X} \\ \sqrt{\lambda_2}\boldsymbol{I}_d \end{pmatrix}, \quad \tilde{\boldsymbol{y}} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{0}_{d\times 1} \end{pmatrix}$$

Show

$$\arg\min J_1(\boldsymbol{w}) = c(\arg\min J_2(\boldsymbol{w}))$$

i.e.

$$J_1(c\boldsymbol{w}) = J_2(\boldsymbol{w})$$

and hence that one can solve an elastic net problem using a lasso solver on modified data.

Q 6 [Reward modification] **[15 points]** A key technique for solving sequential decision problems is the modification of reward functions that leaves the optimal policy unchanged while improving sample efficiency or convergence rates. This problem looks at simple ways of modifying rewards and understanding how these modifications affect the optimal policy.

Consider two Markov decision processes $M \doteq (X, A, p, r)$ and $M' \doteq (X, A, p, r')$ where the reward function r is modified to obtain r', and the rewards are bounded and discounted by the discount factor $\gamma \in [0, 1)$. Let $\pi^\star_M$ be the optimal policy for M.

(a) (5 points) Suppose $r'(x) = \alpha r(x)$, where $\alpha > 0$. Show that the optimal policy $\pi^\star$ of M is also an optimal policy of M'.

(b) (5 points) Given a modification of the form $r'(x) = r(x) + c$, where $c > 0$ is a constant scalar, show that the optimal policy $\pi^\star_M$ can be different from $\pi^\star_{M'}$.

(c) (5 points) Another way of modifying the reward function is through *reward shaping* where one supplies additional rewards to the agent to guide the learning process. When one has no knowledge of the underlying transition dynamics p, a commonly used transformation is

$$r'(x, x') = r(x, x') + f(x, x')$$

where f is a *potential-based shaping function* defined as

$$f(x, x') \doteq \gamma\phi(x') - \phi(x), \quad \phi : X \to \mathbb{R}.$$

Show that the optimal policy remains unchanged under this definition of f.

Q 7 [High-dimensional mapping] **[15 points]** Let $\Phi : \mathcal{X} \to \mathcal{H}$ be a feature mapping such that the dimension N of $\mathcal{H}$ is very large and let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive semi-definite (PDS) kernel defined by

$$K(x, x') = \mathbb{E}_{i\sim\mathcal{D}}\left[[\Phi(x)]_i[\Phi(x')]_i\right], \tag{1}$$

where $[\Phi(x)]_i$ is the ith component of $\Phi(x)$ (and similarly for $\Phi(x')$) and where $\mathcal{D}$ is a distribution over the indices i. We shall assume that $|[\Phi(x)]_i| \leq R$ for all $x \in \mathcal{X}$ and $i \in [N]$. Suppose that the only method available to compute $K(x, x')$ involved direct computation of the inner product in (1), which would require $O(N)$ time. Alternatively, an approximation can be computed based on random selection of a subset I of the N components of $\Phi(x)$ and $\Phi(x')$ according to $\mathcal{D}$, that is:

$$K'(x, x') = \frac{1}{n}\sum_{i\in I}\mathcal{D}(i)[\Phi(x)]_i[\Phi(x')]_i,$$

where $|I| = n$.

(a) (10 points) Fix x and x' in $\mathcal{X}$. Prove that

$$\mathbb{P}_{I\sim\mathcal{D}^n}\left[\left|K(x, x') - K'(x, x')\right| > \epsilon\right] \leq 2e^{-\frac{n\epsilon^2}{2R^2}}. \tag{0.1}$$

(b) (5 points) Let $\mathbf{K}$ and $\mathbf{K}'$ be the kernel matrices associated to K and K'. Show that for any $\epsilon, \delta > 0$, for $n > \frac{R^2}{\epsilon^2} \log \frac{m(m+1)}{\delta}$, with probability at least $1 - \delta$,

$$|\mathbf{K}'_{ij} - \mathbf{K}_{ij}| \leq \epsilon \quad \text{for all } i, j \in [m].$$

Q 8 [Nyström method] **[10 points]** Define the following block representation of a kernel matrix:

$$\mathbf{K} = \begin{bmatrix} \mathbf{W} & \mathbf{K}_{21}^\top \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} \mathbf{W} \\ \mathbf{K}_{21} \end{bmatrix}.$$

The Nyström method uses $\mathbf{W} \in \mathbb{R}^{l \times l}$ and $\mathbf{C} \in \mathbb{R}^{m \times l}$ to generate the approximation

$$\tilde{\mathbf{K}} = \mathbf{C}\mathbf{W}^\dagger\mathbf{C}^\top \approx \mathbf{K}.$$

If $\mathrm{rank}(\mathbf{K}) = \mathrm{rank}(\mathbf{W}) = r \ll m$, show that $\tilde{\mathbf{K}} = \mathbf{K}$.

Note: this statement holds whenever $\mathrm{rank}(\mathbf{K}) = \mathrm{rank}(\mathbf{W})$, but is of interest mainly in the low-rank setting.

— **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):
  $\mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i}; 0 \leq i \leq n.$
  $\mathbb{E}[X] = np, \quad \mathrm{Var}[X] = np(1-p), \quad \mathbb{E}[e^{tX}] = [(1-p) + pe^t]^n.$
- Geometric (p) :
  $\mathbb{P}(X = i) = (1-p)^{i-1} p; i \geq 1.$
  $\mathbb{E}[X] = \frac{1}{p}, \quad \mathrm{Var}[X] = \frac{1-p}{p^2}, \quad \mathbb{E}[e^{tX}] = \frac{pe^t}{1-(1-p)e^t}$ for $t < -\log(1-p)$.
- Poisson (λ):
  $\mathbb{P}(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}; i \geq 1.$
  $\mathbb{E}[X] = \lambda, \quad \mathrm{Var}[X] = \lambda, \quad \mathbb{E}[e^{tX}] = \exp(\lambda(e^t - 1)).$
- Uniform (a,b) :
  $f(x) = \begin{cases} \frac{1}{b-a} \text{ if } a \leq x \leq b \\ 0 \text{ otherwise }. \end{cases}$
  $\mathbb{E}[X] = (a+b)/2, \quad \mathrm{Var}[X] = \frac{(b-a)^2}{12}, \quad \mathbb{E}[e^{tX}] = \frac{e^{tb} - e^{ta}}{t(b-a)}$ if $t \neq 0$.
- Uniform on the square $(a,b) \times (c,d)$:
  $f(x,y) = \begin{cases} \frac{1}{(b-a)(d-c)} \text{ if } a \leq x \leq b, c \leq y \leq d \\ 0 \text{ otherwise }. \end{cases}$
- Normal / Gaussian ($N(\mu, \sigma^2)$):
  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}).$
  $\mathbb{E}[X] = \mu, \quad \mathrm{Var}[X] = \sigma^2, \quad \mathbb{E}[e^{tX}] = \exp(\mu t + \frac{1}{2}\sigma^2 t^2).$
- 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 $t < \lambda$.