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]

Let $(S_n)_{n\geq 0}$ be a simple symmetric random walk on $\mathbb{Z}^d$ for $d \geq 1$. Prove that the trajectory of S_n intersects itself infinitely often almost surely.

Q 2 [10 points]

Let $\{X_k\}_{k\geq 1}$ be a sequence of i.i.d. random variables with Pareto(α) distribution with $\alpha > 0$. This means that $\mathbb{P}(X_1 \geq x) = x^{-\alpha}$ for all $x \geq 1$. Let $M_n = \max\{X_1, \dots, X_n\}$ be the running maximum of the family of random variables $\{X_k\}_{k\geq 1}$.

(a) Find a suitable function $f_\alpha(n)$ such that the rescaled random variable $F_n = \frac{M_n}{f_\alpha(n)}$ converges weakly to a non-degenerate random variable F^α. Describe also the law of F^α.

(b) Is it true that

$$\limsup_n \frac{X_n}{f_\alpha(n)} < \infty \qquad \text{a.s.?}$$

Motivate your answer.

Q 3 [15 points]

Let $(X_i)_{i\geq 1}$ be a sequence of i.i.d. integer valued random variables with symmetric probability mass function, i.e.

$$\mathbb{P}\left[X_1 = k\right] = \mathbb{P}\left[X_1 = -k\right] > 0 \qquad \forall\ k \in \mathbb{Z}_{\geq 0}.$$

(a) Define the characteristic function $\phi(t) = \mathbb{E}\left[e^{\mathrm{i}tX_1}\right]$. Show that $|\phi(t)| < 1$ for all $t \notin 2\pi\mathbb{Z}$.

(b) Define the symmetric random walk $(S_n)_{n\geq 0}$ as

$$S_n = X_1 + \cdots + X_n, \quad \text{for } n > 0 \qquad \text{and} \qquad S_0 = 0.$$

Assuming that $\mathrm{Var}(X_1) < \infty$, show that S_n is recurrent.

Q 4 [15 points]

Consider the 2×2 symmetric random matrix

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \qquad a, c \sim \mathcal{N}(0,2),\ b \sim \mathcal{N}(0,1),$$

with a, b, c independent. Let $\lambda_1 \geq \lambda_2$ be the (real) ordered eigenvalues of A.

(a) Compute the joint pdf $\rho(\lambda_1, \lambda_2)$ of the ordered eigenvalues of A.
Hint: express $A = O\Lambda O^T$ *where* $O = \left(\begin{smallmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{smallmatrix}\right)$ *and* $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2)$ *and compute the Jacobian of the change of variables.*

(b) Define $\mathsf{s} = |\lambda_1 - \lambda_2|$ the spacing between the eigenvalues of A. Compute the pdf of the random variable s.

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 $\pi_M^\star$ 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 $\pi_M^\star$ can be different from $\pi_{M'}^\star$.

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

Note: this statement holds whenever $\operatorname{rank}(\mathbf{K}) = \operatorname{rank}(\mathbf{W})$, 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 $t < \lambda.$

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]

Let $(S_n)_{n\geq 0}$ be a simple symmetric random walk on $\mathbb{Z}^d$ for $d \geq 1$. Prove that the trajectory of S_n intersects itself infinitely often almost surely.

Q 2 [10 points]

Let $\{X_k\}_{k\geq 1}$ be a sequence of i.i.d. random variables with Pareto(α) distribution with $\alpha > 0$. This means that $\mathbb{P}(X_1 \geq x) = x^{-\alpha}$ for all $x \geq 1$. Let $M_n = \max\{X_1, \dots, X_n\}$ be the running maximum of the family of random variables $\{X_k\}_{k\geq 1}$.

(a) Find a suitable function $f_\alpha(n)$ such that the rescaled random variable $F_n = \frac{M_n}{f_\alpha(n)}$ converges weakly to a non-degenerate random variable F^α. Describe also the law of F^α.

(b) Is it true that

$$\limsup_n \frac{X_n}{f_\alpha(n)} < \infty \qquad \text{a.s.?}$$

Motivate your answer.

Q 3 [15 points]

Let $(X_i)_{i\geq 1}$ be a sequence of i.i.d. integer valued random variables with symmetric probability mass function, i.e.

$$\mathbb{P}[X_1 = k] = \mathbb{P}[X_1 = -k] > 0 \qquad \forall\, k \in \mathbb{Z}_{\geq 0}.$$

(a) Define the characteristic function $\phi(t) = \mathbb{E}\left[e^{\mathrm{i}tX_1}\right]$. Show that $|\phi(t)| < 1$ for all $t \notin 2\pi\mathbb{Z}$.

(b) Define the symmetric random walk $(S_n)_{n\geq 0}$ as

$$S_n = X_1 + \dots + X_n, \quad \text{for } n > 0 \qquad \text{and} \qquad S_0 = 0.$$

Assuming that $\mathrm{Var}(X_1) < \infty$, show that S_n is recurrent.

Q 4 [15 points]

Consider the 2×2 symmetric random matrix

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \qquad a, c \sim \mathcal{N}(0, 2),\ b \sim \mathcal{N}(0, 1),$$

with a, b, c independent. Let $\lambda_1 \geq \lambda_2$ be the (real) ordered eigenvalues of A.

(a) Compute the joint pdf $\rho(\lambda_1, \lambda_2)$ of the ordered eigenvalues of A.
Hint: express $A = O\Lambda O^T$ *where* $O = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ *and* $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2)$ *and compute the Jacobian of the change of variables.*

(b) Define $\mathsf{s} = |\lambda_1 - \lambda_2|$ the spacing between the eigenvalues of A. Compute the pdf of the random variable s.

Q 5 [Revisiting exponential kernel: PDS or not] **[10 points]**

For any $p > 0$, let K_p be the kernel defined over $\mathbb{R}_+ \times \mathbb{R}_+$ by

$$K_p(x, y) = e^{-(x+y)^p}.$$

Show that K_p is positive definite symmetric (PDS) if and only if $p \leq 1$.

Q 6 [Locally linear embedding] **[10 points]** The *locally linear embedding* (LLE) algorithm also aims to find a low-dimensional representation that preserves neighborhood relations as measured by a weight matrix $\mathbf{W}$. The algorithm works as follows:

(a) Find t nearest neighbors for each point.

(b) Construct $\mathbf{W}$, a sparse, symmetric $m \times m$ matrix, whose ith row sums to one and contains the linear coefficients that optimally reconstruct $\mathbf{x}_i$ from its t neighbors. More specifically, if we assume that the ith row of $\mathbf{W}$ sums to one, then the reconstruction error is

$$\left(\mathbf{x}_i - \sum_{j \in \mathcal{N}_i} W_{ij}\mathbf{x}_j\right)^2 = \left(\sum_{j \in \mathcal{N}_i} W_{ij}(\mathbf{x}_i - \mathbf{x}_j)\right)^2 = \sum_{j,k \in \mathcal{N}_i} W_{ij} W_{ik} C'_{jk}$$

where $\mathcal{N}_i$ is the set of indices of the neighbors of point $\mathbf{x}_i$ and $C'_{jk} = (\mathbf{x}_i - \mathbf{x}_j)^\top(\mathbf{x}_i - \mathbf{x}_k)$ is the local covariance matrix. Minimizing this expression with the constraint $\sum_j W_{ij} = 1$ gives the solution

$$W_{ij} = \frac{\sum_k (C'^{-1})_{jk}}{\sum_{st}(C'^{-1})_{st}}.$$

Note that the solution can be equivalently obtained by first solving the system of linear equations $\sum_j C'_{kj} W_{ij} = 1$, for $k \in \mathcal{N}_i$, and then normalizing so that the weights sum to one.

(c) Find the k-dimensional representation that best obeys neighborhood relations as specified by $\mathbf{W}$, i.e.,

$$\mathbf{Y} = \operatorname*{arg\,min}_{\mathbf{Y}'} \sum_i \left(\mathbf{y}'_i - \sum_j W_{ij}\mathbf{y}'_j\right)^2. \qquad (1)$$

The solution to the minimization in (1) is $\mathbf{Y} = \mathbf{U}^\top_{\mathbf{M},k}$, where $\mathbf{M} = (\mathbf{I} - \mathbf{W}^\top)(\mathbf{I} - \mathbf{W}^\top)$ and $\mathbf{U}^\top_{\mathbf{M},k}$ are the bottom k singular vectors of $\mathbf{M}$, excluding the last singular vector corresponding to the singular value 0.

LLE coincides with KPCA used with a particular kernel matrix $\mathbf{K}_{LLE}$ whereby the output dimensions are normalized to have unit variance.

Show the connection between LLE and KPCA by deriving the expression for $\mathbf{K}_{LLE}$.

Q 7 [Reweighting subgroups] **[15 points]** Problems can arise when certain types of examples are underrepresented in the data. One natural solution is to *reweight* the data. Suppose we have a training dataset D for linear regression that is the union of two (disjoint) datasets A and B, where $|A| \gg |B|$. For example, A and B might represent data from two different groups of individuals. Let $(x_a^{(i)}, y_a^{(i)})$ denote the i-th example in A and $(x_b^{(i)}, y_b^{(i)})$ denote the i-th example in B.

The reweighted training loss is defined as

$$L(w) = \frac{1}{|A|}\sum_{i=1}^{|A|}(w^\top x_a^{(i)} - y_a^{(i)})^2 + \frac{1}{|B|}\sum_{i=1}^{|B|}(w^\top x_b^{(i)} - y_b^{(i)})^2,$$

where $w \in \mathbb{R}^d$ is the weight vector parameter for linear regression (in this problem, we will omit the bias term).

(a) (5 points) Explain why this loss function is more likely to promote more equal treatment of individuals in dataset A and individuals in dataset B, compared with running normal linear regression on D.

(b) (5 points) This loss function can be optimized by gradient descent. Compute the gradient of $L(w)$ with respect to w.

(c) Suppose you run gradient descent on this loss function. You achieve 0 error on training data from group A as well as on training data from group B. However, test error is much higher on individuals from group B than from group A.

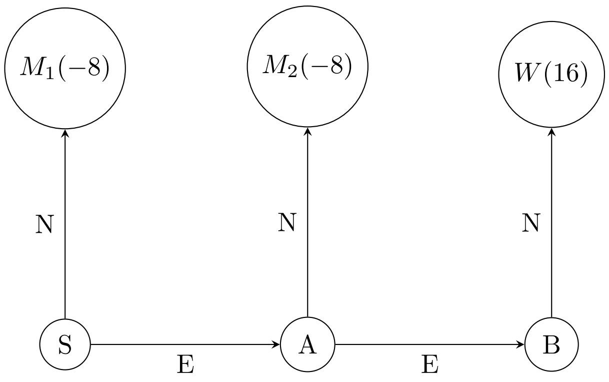
  (i) (3 points) Explain why it would make sense that test error is higher on group B than group A, when both training errors are 0.

  (ii) (2 points) Suggest a change to the loss function that could help improve test error on group B. Explain your reasoning.

Q 8 [MDP and Reinforcement Learning] **[15 points]**

You are living in the year 2050 and are given the task to find water on Mars. You remember the material studied at NUS Math and design a robot that uses the MDP shown below to represent the state space on Mars. Martians live in M_1 and M_2 and water can be found in state W. The states S, A, and B are useless and provide 0 reward. Once your robot reaches one of the M and W states its only option is to send a boolean value back to Earth representing whether or not it found water, and it's not getting any additional reward for doing that.

The robot should stay away from the Martians, so the robot receives a reward of -8 when entering an M state, and a reward of 16 when entering the W state (as indicated in the figure). The only actions it can take is to either go East (E) or go North (N) (it can only go North from state B). You are the best at designing robots and hence, there is no noise when transitioning to different states and actions always succeed.



(a) (5 points) What are the optimal values, V^*, of each state if $\gamma = 0.5$?

(b) (5 points) What are the Q-values for the state S if $\gamma = 0.5$?

(c) (5 points) What is the optimal policy, π^*, for this MDP?

—— **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 $t \ne 0$.
- 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 ($N(\mu,\sigma^2)$):
  $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2}).$
  $\mathbb{E}[X] = \mu, \quad \mathrm{Var}[X] = \sigma^2, \quad \mathbb{E}[e^{tX}] = \exp(\mu t + \frac{1}{2}\sigma^2 t^2).$
- 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 $t < \lambda$.