భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

AI5090: Stochastic Processes

Mid-Term Exam

Date: 02 April 2025

Instructions:

- **Fill in your name and roll number on each page of the answer booklet (to be provided separately).**
- **This exam is for a total of 40 MARKS, and will count towards 30% of your total grade.**
- **You may use any result covered in class directly without proof.**
- **Hints are provided for some questions.**
  However, it is NOT mandatory to solve the question using the approach in the hints.
  If you think you have a better approach in mind than the one given in the hint, feel free to present your approach.
- **Show all your working clearly.**
  We want to see your thought process, and possibly provide partial credit for the intermediate logical steps.
- **Plagiarism will NOT be entertained at any length.**
  If you are caught cheating during the exam, your answer script will NOT be evaluated.

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Assume that all random variables appearing in the questions below are defined with respect to $\mathscr{F}$.

Basic Formulae

1. We say that $X_n \xrightarrow{\text{a.s.}} X$ if

$$\mathbb{P}(A_{\text{lim}}) = 1, \qquad \text{where} \quad A_{\text{lim}} \coloneqq \left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1.$$

2. We say that $X_n \xrightarrow{\text{m.s.}} X$ if

$$\lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] = 0.$$

3. We say that $X_n \xrightarrow{\text{p.}} X$ if

$$\forall \varepsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}\Big(|X_n - X| > \varepsilon\Big) = 0.$$

4. We say that $X_n \xrightarrow{\text{d.}} X$ if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \qquad \forall x \in C_{F_X},$$

   where C_{F_X} = set of continuity points of F_X.

5. The moment generating function of a random variable X is defined as the function

$$M_X(t) = \mathbb{E}[e^{tX}], \qquad t \in \mathbb{R}.$$

6. The characteristic function of a random variable X is defined as the function

$$C_X(s) = \mathbb{E}[e^{jsX}], \qquad s \in \mathbb{R}.$$

7. <u>Taylor's series for C_X:</u>
   If $\mathbb{E}[|X|^k] < +\infty$ for some $k \in \mathbb{N}$, then

$$C_X(s) = \sum_{m=0}^{k} \mathbb{E}[X^m]\, \frac{(js)^m}{m!} + o(s^k).$$

8. <u>Strong law of large numbers:</u>
   If $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables with finite mean $\mu = \mathbb{E}[X_1]$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} \mu.$$

1. Imagine you work at a major research lab developing a next-generation Large Language Model (LLM). During training, you use **mini-batch stochastic gradient descent (SGD)** to update your model parameters. Each mini-batch is sampled randomly from a massive corpus. Because the mini-batches vary across iterations, the model's loss at each iteration n is inherently random.

   Given the sheer size of data and the possibility of widely differing mini-batches, the gradient noise–the randomness stemming from sampling different subsets of data–plays a critical role in determining how the model's parameters evolve. Many researchers have observed that when batches are large enough and come from diverse data, the gradient noise can often be approximated as Gaussian due to the Central Limit Theorem-like effects (i.e., summing many small, independent contributions from individual examples in the mini-batch).

   A common, simplified model for the loss at iteration $n \in \mathbb{N}$, say L_n, is given as

   $$L_n = \frac{V}{n} + \frac{Y_n}{\sqrt{n}},$$

   where V has finite variance σ^2, and typically captures the initial variance in the model parameters, and $Y_1, Y_2, \cdots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ model the instantaneous gradient noise at iteration n, independent of V.

   (a) **(3 Marks)**
   Prove that $L_n \xrightarrow{\text{a.s.}} 0$.

   Hint:
   For any $\varepsilon > 0$, find a summable upper bound on $\mathbb{P}(|L_n - 0| > \varepsilon)$, and use Borel–Cantelli lemma.

   (b) **(2 Marks)**
   For each $n \in \mathbb{N}$, let $M_n = \sqrt{n}\, L_n$. Prove that

   $$M_n \xrightarrow{\text{d.}} M, \qquad M \sim \mathcal{N}(0,1).$$

   Hint:
   Show that the characteristic function of M_n converges to that of a standard Gaussian.

---

2. **(5 Marks)**
   Let $\{X_n\}_{n\in\mathbb{N}}$ be any given sequence of random variables. Show that

   $$X_n \xrightarrow{\text{p.}} 0 \qquad \text{if and only if} \qquad \lim_{n\to\infty} \mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right] = 0.$$

   Hint:
   Let $Z_n \coloneqq \frac{|X_n|}{1+|X_n|}$. For the if part, fix $\varepsilon > 0$, and write

   $$\mathbb{P}(|X_n| > \varepsilon) = \mathbb{E}[\mathbf{1}_{\{|X_n|>\varepsilon\}}] = \mathbb{E}\left[\frac{1}{1+|X_n|}\,\mathbf{1}_{\{|X_n|>\varepsilon\}}\right] + \mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\,\mathbf{1}_{\{|X_n|>\varepsilon\}}\right].$$

   Upper bound the right-hand side of the above relation carefully and show that it goes to 0 as $n \to \infty$.

   For the only if part, fix $\varepsilon > 0$, and write

   $$\mathbb{E}[Z_n] = \mathbb{E}[Z_n\,\mathbf{1}_{\{|X_n|>\varepsilon\}}] + \mathbb{E}[Z_n\,\mathbf{1}_{\{|X_n|\le\varepsilon\}}].$$

   Upper bound the right-hand side of the above relation carefully to show that it can be made negligible as $n \to \infty$.

---

3. For $n \in \mathbb{N}$, let S_n be the position of the 1-D random walk on the set of integers, starting from $S_0 = 0$.
   At any given time, the random walk moves one integer to the right with probability p, or one integer to the left with probability $1-p$. Let $X_n \coloneqq S_n - S_{n-1}$.

(a) **(3 Marks)**
If $p \neq \frac{1}{2}$, show that $\mathbb{P}(S_n \text{ i.o.}) = 0$.

Hint:
What can you say about the convergence of $\left\{\frac{S_n}{n}\right\}_{n \in \mathbb{N}}$?
Use this finding to evaluate the limsup/liminf of any typical sample path of $\{S_n\}_{n \in \mathbb{N}}$, and show that infinitely many of the S_n's cannot be 0.

(b) **(2 Marks)**
Let $\{A_n\}_{n \in \mathbb{N}}$ be a collection of events belonging to $\mathscr{F}$ such that $\mathbb{P}(A_n) = 1 - \frac{1}{2^n}$.
Compute $\mathbb{P}(A_n \text{ i.o.})$.

---

4. Let $\{\mathscr{F}_n\}_{n \in \mathbb{N}}$ be a filtration, i.e., $\mathscr{F}_n \subseteq \mathscr{F}_{n+1} \subset \mathscr{F}$ for all $n \in \mathbb{N}$.
For each $n \in \mathbb{N}$, let Y_n be a random variable with respect to the σ-algebra $\mathscr{F}_n$.
Let N be a stopping time with respect to the above filtration, and let

$$\mathscr{F}_N = \{A \in \mathscr{F} : \ A \cap \{N = n\} \in \mathscr{F}_n \ \forall n \in \mathbb{N}\}.$$

(a) **(2 Marks)**
Show that Y_N is a random variable with respect to $\mathscr{F}_N$.

(b) **(3 Marks)**
Let $Y_1, Y_2, \cdots \overset{\text{i.i.d.}}{\sim} \text{Unif}\{1, \ldots, 10\}$. Let

$$N := \inf\{n \geq 1 : Y_n \in \{1, 3, 5, 7\}\}.$$

Show that N is a stopping time with respect to the natural filtration of the process $\{Y_n\}_{n \in \mathbb{N}}$, and compute $\mathbb{E}[N]$.

---

5. **(5 Marks)**

Suppose there are 100 distinct papers in a drawer. In every round, you draw a paper and sign it (if it is not already signed), and then place the paper back into the drawer. If an already signed paper is obtained in any draw, you simply place the paper back into the drawer. If any paper is equally likely to be drawn each time, independent of all other rounds, evaluate the expected number of rounds to sign all 100 papers. You may leave the final answer in the form of a summation.

---

6. Let $\{X_t : t \in \mathbb{R}\}$ be a stochastic process defined via

$$X_t = 2A + Bt, \qquad t \in \mathbb{R},$$

where $A, B \overset{\text{i.i.d.}}{\sim} \text{Unif}\{-1, +1\}$.

(a) **(3 Marks)**
Compute $\mathbb{P}(X_t \geq 0)$ for all $t \in \mathbb{R}$.

(b) **(2 Marks)**
Compute $\mathbb{P}\Bigg(X_t \geq 0 \ \text{ for all } t \in \mathbb{R}\Bigg)$.

---

7. Let $X_1, X_2, \cdots \overset{\text{i.i.d.}}{\sim} \text{Exponential}(\lambda)$ for some fixed $\lambda > 0$.
For each n, let $T_n := \sum_{i=1}^{n} X_i$.
Imagine that cars arrive at a traffic junction at times $T_1, T_2, \ldots$, where T_n denotes the arrival instant of nth car.
Consider a continuous-time process $\{N(t) : t \geq 0\}$, where for each $t \geq 0$,

$$N(t) := \max\{n \in \mathbb{N} : T_n \leq t\}$$

denotes the number of cars seen up to time t.

(a) **(2 Marks)**
It is a well-known fact that for each $n \in \mathbb{N}$, the random variable $T_n \sim \text{Gamma}(n, \lambda)$, with the PDF

$$f_{T_n}(s) = \frac{\lambda^n \, e^{-\lambda\, s} \, s^{n-1}}{(n-1)!}, \qquad s > 0.$$

Using this, derive the PMF of $N(t)$ for each $t \in \mathbb{R}_+$.

Hint: Observe that

$$\mathbb{P}(N(t) = n) = \mathbb{P}(T_n \leq t, \; T_{n+1} > t).$$

(b) **(1 Mark)**
Derive the moment generating function of $N(t)$ for any given $t \in \mathbb{R}_+$.

(c) **(2 Marks)**
Consider a new discrete-time process $\{Y_m\}_{m \in \mathbb{N}}$ defined as $Y_m = N(m)$, $m \in \mathbb{N}$.
Prove that

$$\frac{Y_m}{m} \xrightarrow{\text{m.s.}} \lambda.$$

(d) **(2 Marks)**
Show that for any $\varepsilon > 0$,

$$\mathbb{P}\left(\frac{Y_m}{m} > \lambda + \varepsilon\right) \leq \exp\left(-m\left[(\lambda+\varepsilon)\,\log\left(\frac{\lambda+\varepsilon}{\lambda}\right) - (\lambda+\varepsilon) + \lambda\right]\right).$$

Hint: Use Chernoff bound.

(e) **(2 Marks)**
Show that for any $\varepsilon > 0$,

$$\mathbb{P}\left(\frac{Y_m}{m} < \lambda - \varepsilon\right) \quad \begin{cases} = 0, & \varepsilon \geq \lambda, \\ \leq \exp\left(-m\left[(\lambda-\varepsilon)\,\log\left(\frac{\lambda-\varepsilon}{\lambda}\right) - (\lambda-\varepsilon) + \lambda\right]\right), & \varepsilon < \lambda. \end{cases}$$

Hint: Use Chernoff bound, but first think of a way to convert '$<$' in the event $\left\{\frac{Y_m}{m} < \lambda - \varepsilon\right\}$ to '$>$'.

(f) **(1 Mark)**
Using the results of part (d) and part (e), show that

$$\frac{Y_m}{m} \xrightarrow{\text{a.s.}} \lambda.$$