EECS 126 Probability and Random Processes — UC Berkeley, Fall 2023

Jiantao Jiao — October 31, 2023

Midterm 2

Last Name	First Name	SID

Left Neighbor First and Last Name	**Right Neighbor** First and Last Name

Rules.

- **Unless otherwise stated, all your answers need to be justified and your work must be shown. Answers without sufficient justification will get no credit.**
- **All work you want graded can be on both the front and back of the sheets in the space provided. Both sides will be scanned/graded.**
- You have 10 minutes to read the exam and 70 minutes to complete the exam. (DSP students with $X\%$ time accommodation should spend $10 \cdot X\%$ time on reading and $70 \cdot X\%$ time on completing the exam).
- This exam is closed-book. You may reference one double-sided handwritten sheet of paper. No calculator or phones allowed.
- Collaboration with others is strictly prohibited. If you are caught cheating, you may fail the course and face disciplinary consequences.

Problem	out of
Problem 1	21
Problem 2	21
Problem 3	21
Problem 4	21
Problem 5	21
Problem 6	21
Total	126

1 Concentration Bounds [21 points]

Jesse is not happy with his knowledge of concentration inequalities. Let us teach Jesse some more about concentration inequalities!

(a) Write down your SID on the top right corner to get 3 points. (3 points)

(b) Consider $Y \sim \mathsf{Bin}(n, p)$. Let's say we are interested in the probability of Y realizing a value too far off from the mean np. Find the exact expression for $P(Y - np \geq \delta)$. You may assume that np and δ are positive natural numbers and $np + \delta < n$. You may leave your answer in the form of a summation. (2 points)

(c) Use Chebyshev inequality to upper bound $P(Y - np \geq \delta)$. (1 point)

(d) Now prove $P(Y - np \geq \delta) \leq \frac{np(1-p)}{np(1-p)+\delta^2}$. Compare this bound to the one in part (c). (5+1 point)

Hint: for a zero-mean random variable Z, $P(Z \geq \lambda) = P(Z + u \geq \lambda + u)$ $\forall u, \lambda \geq 0$. Try applying Markov inequality.

(e) Let us shift our focus and try to use Chernoff bound to get a different concentration bound.

Suppose for zero-mean and independent random variables $X_1, \ldots, X_n$, you are given the constants $\sigma_1, \ldots, \sigma_n$ such that $\mathbb{E}[e^{sX_i}] \leq e^{s^2\sigma_i^2/2}$ $\forall i \in \{1, \ldots, n\}, s \in \mathbb{R}$.

(i) For $t > 0$, prove that $P(X_i \geq t) \leq e^{-t^2/2\sigma_i^2}$. (2 points)
Hint: Recall the Chernoff bound done in discussion.

(ii) For $t > 0$, prove that $P(\sum_i X_i \geq t) \leq e^{-t^2/2\sum_i \sigma_i^2}$. (3 points)

(iii) Note that $Y \sim \mathsf{Bin}(n, p)$ can be written as $Y = \sum_{i=1}^n X_i$ where X_i are i.i.d. $\mathsf{Bern}(p)$ random variables. Use the idea of Chernoff bound present above to upper bound $P(Y - np \geq \delta)$. (4 points)
Hint: You may use without proof that for $X \sim \mathsf{Bern}(p)$, we have $\mathbb{E}[e^{s(X-p)}] \leq e^{s^2/8}$ $\forall s$.

(a) SID (3 points).

(b) Define the set $S = \{np + \delta, np + \delta + 1, \ldots, n\}$ (1 point), then

$$P(Y - np \geq \delta) = \sum_{i \in S} \binom{n}{i} p^i (1-p)^{n-i}. \text{(1 point)} \tag{1}$$

(c) $P(Y - np \geq \delta) \leq P(|Y - np| \geq \delta) \leq \frac{np(1-p)}{\delta^2}$. (1 points)

(d) Let the variance of Z be σ^2.

$$\begin{aligned} P(Z \geq \lambda) &= P(Z+u \geq \lambda+u) \quad \forall u \geq 0 \\ &\leq P((Z+u)^2 \geq (\lambda+u)^2) \quad \text{(1 points)} \\ &\leq \frac{\sigma^2+u^2}{(\lambda+u)^2} \quad \text{(1 points)} \end{aligned}$$

Since the above is true for all $u \geq 0$, we can optimize the RHS to find the $u \geq 0$ that minimizes it. (1 points)

That is given by $u = \sigma^2/\lambda$ and we get the bound $P(Z \geq \lambda) \leq \frac{\sigma^2}{\sigma^2+\lambda^2}$. (1 points)

Considering $Z = Y - np$ and $\delta = \lambda$ in the above, we get the desired inequality. (1 points)

This inequality is stronger than the Chebyshev bound found in part (c). (1 points)

(e) (i) $P(X_i \geq t) = P(e^{sX_i} \geq e^{st}) \leq \frac{e^{s^2\sigma_i^2/2}}{e^{st}} \quad \forall s > 0$. (1 point)
Choosing $s = t/\sigma_i^2$, we get $P(X_i \geq t) \leq e^{-t^2/2\sigma_i^2}$. (1 points)

(ii) $P(\sum_i X_i \geq t) = P(e^{s\sum_i X_i} \geq e^{st}) \leq \frac{e^{s^2\sum_i \sigma_i^2/2}}{e^{st}} \quad \forall s > 0$ by independence of X_is. (2 point)
By steps similar to previous part we get $P(\sum_i X_i \geq t) \leq e^{-t^2/2\sum_i \sigma_i^2}$. (1 points)

(iii) $P(Y - np \geq \delta) = P(\sum_i (X_i - p) \geq \delta)$. (1 points)
In the above result, we can use $X_i - p$ as our zero-mean random variables with $\sigma_i^2 = \frac{1}{4}$ as given in the hint. (1 points)
Therefore, $P(\sum_i (X_i - p) \geq \delta) \leq e^{-2\delta^2/n}$. (2 points)

2 Convergence [21 points]

Jesse is not happy with his knowledge of convergence of random variables. Let us teach Jesse some more about convergence!

(a) Write down your SID on the top right corner to get 3 points. (3 points)

(b) Let $X_1, \ldots, X_n$ be i.i.d. random variables from the gaussian distribution $\mathcal{N}(\mu, \sigma^2)$. Jesse does not know μ or σ^2, and only wants to find out the probability that the random variables take non-negative values. Jesse constructs the sequence of random variables

$$Z_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_i \geq 0\},$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. If Z_n convergences to a limit almost surely, then find it; otherwise, explain why it does not convergence almost surely. (5 points)

(c) (Independent of part (b)) Jesse sees a sequence of Bernoulli random variables $X_1, X_2, \ldots$ and assumes they are i.i.d. In order to find the mean of the Bernoulli distribution, he considers the empirical mean and thinks that by SLLN, it shall converge to the actual mean. However, he does not know that the sequence was generated by Walter White. Walter generates $X_i \sim \mathsf{Bern}(p_1)$ independently for all indices i that are even, and generates $X_i \sim \mathsf{Bern}(p_2)$ independently for all indices i that are odd. In other words, the sequence is comprised of alternating samples from $\mathsf{Bern}(p_1)$ and $\mathsf{Bern}(p_2)$ distribution independently.

This violates the condition for SLLN to work, so will the empirical mean that Jesse is considering converge to something almost surely? If so, find it. Else, explain why not. (5 points)

(d) (Same setting as part (c)) Jesse starts to suspect that Walter is up to some nefarious deed. Jesse considers the sequence of random variables $Y_i = \mathbb{I}\{X_{2i} > X_{2i-1}\} - \mathbb{I}\{X_{2i} < X_{2i-1}\}$. He considers looking at the sample mean $\frac{1}{n}\sum_{i=1}^{n} Y_i$. Argue how this will help Jesse figure out that the X_i are not i.i.d. (assume $p_1 \neq p_2$). (5 points)

(e) Can you provide an intuitive reason why the binomial pmf looks like a Gaussian distribution for large n with the help of CLT? (3 points)

(a) SID (3 points).

(b) The random variables $\mathbb{I}\{X_i \geq 0\}$ are

- independent (1 point)
- identically distributed (1 point)

with mean $P(X \geq 0)$ for $X \sim \mathcal{N}(\mu, \sigma^2)$. (1 points).

By SLLN, $Z_n \to P(X \geq 0)$ a.s. (2 points)

(c) Consider $Y_i = (X_{2i-1} + X_{2i})/2$. Then $\frac{1}{2n}\sum_{i=1}^{2n} X_i = \frac{1}{2}\sum_{i=1}^{n} Y_i$. (3 point)

Note that Y_i has mean $(p_1 + p_2)/2$. (1 point)

By SLLN, $\frac{1}{2n}\sum_{i=1}^{2n} X_i = \frac{1}{2}\sum_{i=1}^{n} Y_i$ converges to $(p_1 + p_2)/2$ almost surely. (1 point)

(d) Note that $E[Y_i] = p_1 - p_2 \neq 0$ (2 point)
and Y_i are i.i.d. (1 point).
Thus, by SLLN, it converges to $p_1 - p_2 \neq 0$. (1 point)
If X_i were i.i.d, then it would have converged to 0 instead. (1 point)

(e) $Y \sim \mathsf{Bin}(n, p)$ can be seen as $Y = \sum_{i=1}^{n} X_i$ where X_i are i.i.d. $\mathsf{Bern}(p)$. (1 point)
By CLT, we can expect $Y \approx \mathcal{N}(np, np(1-p))$. (2 point)
Therefore, Y looks like a scaled Gaussian.

3 Information Theory [21 points]

(a) Write down your SID on the top right corner to get 3 points. (3 points)

(b) Jesse has a biased coin and tosses it n times to observe $X_1, \ldots, X_n$ i.i.d. If $X_i \sim \mathsf{Bern}(p)$, then find the entropy of the sequence, i.e., $H(X_1, X_2, \ldots, X_n)$. (3 points)

(c) Jesse sees the realization (X_1, X_2, X_3, X_4) but he does not know the bias of the coin p. Unfortunately, Jesse does not have a fair coin and wants to generate coin tosses that are $\mathsf{Bern}(0.5)$ from the observation of the Xs. Based on the realization of (X_1, X_2, X_3, X_4), Jesse wants to output K i.i.d. $\mathsf{Bern}(0.5)$ random variables, where K may depend on the realization of (X_1, X_2, X_3, X_4) and for some realizations, Jesse is fine with not outputting anything. Help him accomplish this task; there are multiple ways to achieve this and you may get more marks for giving schemes that provide more bits K. (6 points)

Hint: Jesse may not output anything if he sees the realization $(1, 1, 1, 1)$.

(d) Jesse and Walter are having a cook-off (tournament) with best of 3 games. The games stop as soon as the winner is decided, i.e., if Jesse wins the first 2 games, then the last game is not played. Jesse and Walter are equally matched and the probability of winning for either is 0.5 for a match.

(i) Let X denote the outcome of the tournament. For example, a possible realization of X can be (J, J) denoting Jesse won the first and second game. Find $H(X)$. (3 points)

(ii) Let Y denote the number of games that were played in the tournament. Find $H(Y)$. (3 points)

(iii) Find the joint entropy $H(X, Y)$. (3 points)

(a) SID (3 points).

(b) Since X_i are independent, $H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i)$, (2 point)
and $H(X_i) = -p \log_2(p) - (1-p)\log_2(1-p)$. (1 point)

(c) Since Jesse doesn't know p, the only thing he can use is that the sequence is i.i.d. and permutation of the same string is equiprobable. That is $(X_1, \ldots, X_4) = (1, 0, 0, 0)$ and $(X_1, \ldots, X_4) = (0, 1, 0, 0)$ are both equally likely. Thus, he can use the following mapping:

$$
\begin{aligned}
&0000 \to \Lambda;\\
&0001 \to 00,\ 0010 \to 01,\ 0100 \to 10,\ 1000 \to 11;\\
&0011 \to 00,\ 0110 \to 01,\ 1100 \to 10,\ 1001 \to 11;\\
&0101 \to 0,\ 1010 \to 1;\\
&1110 \to 00,\ 1101 \to 01,\ 1011 \to 10,\ 0111 \to 11;\\
&1111 \to \Lambda;
\end{aligned}
$$

where Λ is the null output.

(d) (i) X can have the realizations $JJ, JWJ, WJJ, WW, WJW, JWW$ (1 point)
with respective probabilities being $\frac{1}{2^2}, \frac{1}{2^3}, \frac{1}{2^3}, \frac{1}{2^2}, \frac{1}{2^3}, \frac{1}{2^3}$. (1 point)
Thus $H(X)$ is 2.5. (1 point)

(ii) Based on the above, Y can have values $2, 3$ (1 point)
with equal probabilities. (1 point)
Thus, $H(Y) = 1$. (1 point)

(iii) $H(X, Y) = H(X) + H(Y|X)$ (1 point)
and Y is a deterministic function of X so $H(Y|X) = 0$. (1 point)
Thus, $H(X, Y) = 2.5$. (1 point)

4 Basics of Markov Chain [21 points]

Consider the Markov chain with state space $\{0, 1, 2, 3, 4, 5, 6\}$ and transition probability matrix

$$\begin{pmatrix} 1/2 & 0 & 1/8 & 1/4 & 1/8 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2/9 & 5/9 & 2/9 \\ 0 & 0 & 0 & 0 & 5/6 & 0 & 1/6 \\ 0 & 0 & 0 & 0 & 2/3 & 1/3 & 0 \end{pmatrix}$$

(a) Write down your SID on the top right corner to get 3 points. (3 points)

(b) Find $\mathbb{P}(X_6 = 4 | X_3 = 2, X_2 = 0)$. (3 points)

(c) Find the recurrent and transient classes. (6 points)

(d) Consider the Markov chain restricted on the recurrent class, is it reversible? Prove your claim. (6 points)

(e) Suppose the initial state is 0, find the expected total number of visits to state 3, i.e.,

$$\mathbb{E}\left[\sum_{n=1}^{\infty} \mathbb{1}(X_n = 3) | X_0 = 0\right]$$

(3 points)

(a) SID (3 points).

(b) We have

$$\mathbb{P}(X_6 = 4 | X_3 = 2, X_2 = 0) = \mathbb{P}(X_6 = 4 | X_5 = 0) = 1/8. \text{ (3 points)}$$

(c) Transient class: (0,1,2,3). (3 points)
Recurrent class: (4,5,6). (3 points)

(d) Yes (2 points). The invariant distribution is $(1/2, 1/3, 1/6)$(2 points), which verifies the detailed balance condition. (2 points)

(e) By first-step equation, we have $\mathbb{P}(\tau_3 < \infty | X_0 = 0) = 3/4$. It follows that

$$\mathbb{E}\left[\sum_{n=1}^{\infty} \mathbb{1}(X_n = 3) | X_0 = 0\right] = 3. \text{ (3 points)}$$

.

5 Ehrenfest's diffusion model [21 points]

In Ehrenfest's diffusion model, a container is separated by a permeable membrane in the middle and filled with a total of K particles. At each time $n = 1, 2, \ldots$, one particle is picked uniformly in random among the K particles and placed into the other part of the container. Let X_n be the number of particles in the left part of the container.

(a) Write down your SID on the top right corner to get 3 points. (3 points)

(b) What is the transition probability $\mathbf{P}$ of Ehrenfest's diffusion model? (6 points)

(c) Find the stationary distribution. (6 points)

(d) Suppose initially all K particles are located at the right part of the container. For each of the following sequences of random variables, study its convergence. (6 points)

That is, you are expected to answer: (i) if the sequence converges; (ii) if so, what is the limit, and what is the notion of convergence (e.g. almost surely, in probability, or in distribution); (iii) if not, briefly explain the reason.

- $(X_n)_{n=1,2,\ldots}$
- $(Y_n)_{n=1,2,\ldots}$ where $Y_n = \frac{1}{n}\sum_{j=1}^{n} X_j^2$

(a) SID (3 points).

(b) The transition probability is given by

$$p_{ij} = \begin{cases} \frac{i}{K}, & j = i - 1 \text{ (2 points)} \\ \frac{K-i}{K}, & j = i + 1 \text{ (2 points)} \\ 0, & \text{else (2 points)} \end{cases}$$

(c) We have the following relation:

$$\pi_i = \pi_{i-1} \cdot p_{i-1,i} + \pi_{i+1} \cdot p_{i+1,i}, \ \forall i = 1, 2, \ldots, K-1 \text{ (3 points)}$$

Solving this gives $\pi_i = \binom{K}{i}\pi_0$, yielding $\pi_i = \binom{K}{i}2^{-K}$ (3 points).

(d) By (c), the Markov chain is irreducible. It is further obvious that it is periodic. It follows that

- X_n does not converge(2 points), because we notice that X_n must have the same parity with n(1 points).
- Let $z \sim \text{Binom}(K, 1/2)$, then $Y_n \to \sum_{i=1}^{K} i^2 \cdot \pi_i = \text{Var}(z) + \mathbb{E}[z]^2 = \frac{K+K^2}{4}$ almost surely (3 points, give 2 points if the convergence notion is wrong.).

6 Bus Problem [21 points]

Let the number of bus arrivals follow a Poisson process $\{N(t) : t > 0\}$ with rate $\lambda > 0$.

(a) Write down your SID on the top right corner to get 3 points. (3 points)

(b) Define $T_i = \inf\{t > 0 : N(t) = i\}$ as the time of i-th arrival. Write out (no need to prove) the mean and variance of T_i. (3 points)

(c) Find the conditional distribution of the exact time at which the bus arrived, given exactly one bus has arrived by time t, i.e., find the distribution of $T_1|N(t) = 1$. (6 points)

(d) If every arriving bus is with probability η blue and with probability $1-\eta$ green, independently from each other and $\{N(t) : t > 0\}$. What is the probability that exactly one blue bus and one green bus arrive the during time interval $(0, T]$? (3 points)

(e) Let Y_i denote the CO2 emission for each i-th arrival, such that $\mu = \mathbb{E}[Y_i]$ for each i, and each Y_i are independent from each other and $\{N(t) : t > 0\}$. What is the expected total CO2 emission at $t = T$? (6 points)

(a) SID (3 points).

(b) We know that $T_i \sim \text{Erlang}(i, \lambda)$. Therefore $\mathbb{E}[T_i] = i/\lambda$(2 points), $\text{Var}[T_i] = i/\lambda^2$(1 points).

(c) For any $s < t$,

$$\begin{aligned}
\mathbb{P}(T_1 < s|N(t) = 1) &= \frac{\mathbb{P}(N(s) = 1, N(t) - N(s) = 0)}{\mathbb{P}(N(t) = 1)} \text{ (3 points)} \\
&= \frac{\mathbb{P}(N(s) = 1)\mathbb{P}(N(t) - N(s) = 0)}{\mathbb{P}(N(t) = 1)} \\
&= \frac{\lambda s e^{-\lambda s} e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} \\
&= \frac{s}{t} \text{ (3 points).}
\end{aligned}$$

It follows that $T_1|N(t) = 1 \sim \text{Unif}(0, t)$.

(d) By Poisson splitting, blue bus arrival and green bus arrival are independent Poisson processes with parameter $\lambda\eta$ and $\lambda(1-\eta)$. It follows that the probability that exactly one blue bus and one green bus the during time interval $(0, T]$ is

$$\mathrm{e}^{-\lambda\eta T}\lambda\eta T \cdot \mathrm{e}^{-\lambda(1-\eta)T}\lambda(1-\eta)T \text{ (3 points).}$$

(e) The total CO2 emission up till time t is given by

$$W(T) = \sum_{i=1}^{N(T)} Y_i.$$

Notice that

$$\mathbb{E}[W(T)|N(T)] = \mu N(T). \text{ (3 points)}$$

By law of iterative expectation

$$\begin{aligned}
\mathbb{E}[W(T)] &= \mathbb{E}[\mathbb{E}[W(T)|N(T)]] \\
&= \mathbb{E}[\mu N(T)] \\
&= \lambda \mu T. \text{ (3 points)}
\end{aligned}$$