SI 140A: Probability and Statistics for EECS
Spring 2025, Instructor: Xavier Lagorce, TA: Junye Wang, Chuyan Zhou

Homework 4

Due: 23:59 (CST), Mar 28, 2025
Full Mark: 50 points, 5 problems including 1 bonus.

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- You are required to write down all the major steps towards making your conclusions; otherwise you may obtain limited points of the problem.
- Write your homework in English; otherwise you will get no points of this homework.
- Any form of plagiarism will lead to 0 point of this homework.
- If there are any descriptions unclear, we recommend you to ask on Piazza.

Problem 1

(LOTUS *5'+5'***)**

(a) Use LOTUS to show that for $X \sim \text{Pois}(\lambda)$ and any function g, $E[Xg(X)] = \lambda E[g(X+1)]$. This is called the *Stein-Chen identity* for the Poisson.

(b) Find the third moment $E[X^3]$ for $X \sim \text{Pois}(\lambda)$ by using the identity from (a) and a bit of algebra to reduce the calculation with the fact that X has mean λ and variance λ.

Problem 2

(Hypergeometric? *3'+3'+3'+3'+3'***)**

Let a random variable X be Hypergeometric with parameters w, b, n: $X \sim \text{HGeom}(w, b, n)$.

(a) Find $E[X]$ by the method of indicator random variables.

(b) Find $E[\binom{X}{2}]$ by the method of (pairs of) indicator random variables.

(c) Find $Var[X]$ from the conclusions of (b).

Now consider the background story of the Hypergeometric distribution: An urn contains w white balls and b black balls, which are randomly drawn one by one without replacement. Let X denote the number of black balls drawn before drawing r $(1 \leq r \leq w)$ white balls.

(d) Find the PMF of X.

(e) Find $E[X]$.

Problem 3

(Elevator *5'+5'***)**

A building has n floors, labeled $1, 2, \ldots, n$. At the first floor, k people enter the elevator, which is going up and is empty before they enter. Independently, each decides which of the floors $2, 3, \ldots, n$ to go to and presses that button (unless someone has already pressed it).

(a) Assume for this part only that the probabilities for floors $2, 3, \ldots, n$ are equal. Find the expected number of stops the elevator makes on floors $2, 3, \ldots, n$.

(b) Generalize (a) to the case that floors $2, 3, \ldots, n$ have probabilities $p_2, \ldots, p_n$ (respectively); you can leave your answer as a finite sum.

Problem 4

(Database Bug *3'+4'+4'+4'***)**

Alvin's database of friends contains n entries, but due to a software bug, the addresses correspond to the names in a totally random fashion. Alvin writes a holiday card to each of his friends and sends it to the (software-corrupted) address. Let X denote the number of friends of him who will get the correct card.

(a) Find $E[X]$.

(b) Find $Var[X]$.

(c) Find the PMF of X.

(d) When $n \to +\infty$, show that the distribution of X converges to a Poisson distribution.

Problem 5

(Bonus: Entropy *5'+5'***)**

(a) (**Maximum Entropy**) Find the probability mass function $p(x)$ that maximizes the entropy $H(X)$ of a nonnegative integer-valued random variable X (whose **support is** $\mathbb{N}$) subject to the constraint

$$E[X] = \sum_{n=0}^{+\infty} np(n) = A.$$

for a fixed value $A > 0$. Evaluate this maximum $H(X)$. Hint: *You may regard the distribution* $(p(x), x \in \mathbb{N})$ *as a group of variables. You may use Lagrange Multiplier Method studied in Calculus.*

Note that in this subproblem we are using the notion of entropy for probability distributions on a countably infinite set. Also, in this subproblem note that the entropy is a concave function of any underlying probability distribution $(p(x), x \in \mathcal{X})$. You should make convenience for yourself of this. This fact is a consequence of the convexity of the function $x \mapsto x \log_2 x$, defined for $x \geq 0$.

(b) (**Entropy of a Mis-sorted File**)A deck of cards in order $1, 2, \ldots, n$ is provided. One card is removed at random, then replaced at random. What is the entropy of the resulting deck?

Note that every outcome is equally probable, and the resulting deck is a RV whose states can indicate distinct events, each of which contains some outcomes coming from different configurations and appearing the same in the result sequence.