MSAI Probability Home Assignment 2

1 Introduction

We have prepared 16 problems for you. You only need to solve 7 of them to obtain full score for this homework. This time, you should solve at least 3 problems from the discrete list and at least three from the continuous list. Problem that you correctly solve extra to those 7 will be a bonus. Bonuses count towards exemption from exam and project.

2 Random variables and distributions

Problem 1. Consider two independent random variables $X \sim F_X$ and $Y \sim F_Y$. Find the CDF of random variables $Z_1 = \max(X, Y)$ (it means that for every outcome w we have $Z_1(w) = \max(X(w), Y(w))$, so Z_1 jumps between values of X and Y) and $Z_2 = \min(X, Y)$ (same reasoning applies).

3 Discrete random variables

Problem 2. (Greedy employer) According to the labor laws at city M, the employer must provide **all** employees with a vacation if **at least one** of them has a birthday on that day. Apart from these vacation days, there are none, and the employees work 365 days. The employer asks you to find the optimal number of employees such that the number of working man-days is maximized. What is this number? How many working man-days does it yield?

Problem 3. (LOTUS for expectation) For $X \sim Pois(\lambda)$, find $\mathbb{E}\left[2^X\right]$, if it is finite.

Problem 4. (Papers please) You have n enumerated letters and n enumerated envelopes. You randomly put letters into envelopes. What is the expected value of the number of coinciding numbers of the letter and its envelope?

Problem 5. (Fundamental bridge) n people enter the elevator at the first floor of 17-floor building. Any person may live on any of the 16 floors (because you don't need an elevator if you live on the fist floor) with equal probability and independent of others. Let X be the number of stops of the elevator. Find the expected value and variance for X.

Problem 6. (Poisson approximation to Binomial) A group of n people play "Secret Santa". They do it as follows: each puts their name on a slip of paper in a hat, then picks a name randomly from the hat **without replacement**, and then buys a gift for that person. Unfortunately, they overlook the possibility of drawing one's own name. Assume $n \geqslant 2$. Find the expected value of the number X of people who pick their own names. What is the approximate distribution of X if n is large (specify distribution and parameters)?

Problem 7* . (It's not as hard as it seems) The legendary physicist Richard Feynman posed the following problem about how to decide what to order at a certain restaurant. You plan to make multiple visits to a restaurant, where you have never eaten before. In total, you plan to eat m meals. Each visit, you will order only one dish.

The restaurant has n dishes on the menu, with $n \geqslant m$. Assume that if you had tried all the dishes, you would have a definite ranking of them from 1 (your least favorite) to n (your favorite). If you knew which your favorite was, you would be happy to order it always (you never get tired of it).

Before you've eaten at the restaurant, this ranking is completely unknown to you. After you've tried some dishes, you can rank those dishes amongst themselves, but don't know how they compare with the dishes you haven't yet tried. There is thus an **exploration-exploitation**

tradeoff: should you try new dishes, or should you order your favorite among the dishes you have tried before?

A natural strategy is to have two phases in your series of visits to the restaurant: an exploration phase, where you try different dishes each time, and an exploitation phase, where you always order the best dish you obtained in the exploration phase. Let k be the length of the exploration phase (so $(m-k)$ is the length of the exploitation phase). Your goal is to maximize the expected sum of the ranks of the dishes you eat there. What length k should you pick?

We will now show that the answer to this question is $k = \sqrt{2(m+1)} - 1$.

1. Let X be the true rank of the best dish that you find in the exploration phase. Find the expected sum of the ranks of all the dishes you eat (including both phases), in terms of k, n and $\mathbb{E}[X]$.

2. Find the PMF of X as a simple expression in terms of binomial coefficients.

3. Show that
$$\mathbb{E}[X] = \frac{k(n+1)}{k+1}$$

4. Optimize $\mathbb{E}[X]$ for k.

Problem 8* . (Poisson process) Let $X_1, X_2, \ldots \sim Be(p)$ be a series of independent random variables. Let $N \sim Pois(\lambda)$. Find the PMF of $Y = \sum_{i=1}^{N} X_i$.

4 Continuous random variables

Problem 9. (Universality of the uniform, backwards) Consider random variable $X \sim U[0,1]$. Consider any cumulative distribution function F, not necessarily continuous. Then define **quantile function** as
$$F^{-1}(y) = \inf\{x \text{ such that } F(x) \geqslant y\}$$

Prove that $F^{-1}(X) \sim F$.

Problem 10. (Inverse transform sampling) Suppose you have a computer and the only permitted operation for you is to read random chunks of its memory (zeros and ones) of size N bits. You can assume that contents of these chunks is also random. So you perform a read operation and obtain a new random size-N array of zeros and ones every time. Propose a way to generate an exponential random variable using this tool.

Problem 11. (Cauchy distribution) Consider 2D real plane $\mathbb{R}^2$. Consider a point with coordinates $(0, d)$. Select an angle φ uniformly distributed in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Issue a ray from $(0, d)$ at the angle φ to the y-axis. Denote the point where it intersects x-axis as $(X, 0)$. The distribution of X is called the Cauchy distribution. Find its CDF and PDF. Explain, why it does not have an expected value.

Problem 12. (Half-life of a particle) Let T be the time until a radioactive particle decays, and suppose that $T \sim Exp(\lambda)$. Find half-life of the particle, i.e. time t such that $\mathbb{P}(T > t) = \frac{1}{2}$.

Problem 13. (Hazard function) Let T be the lifetime of a certain person (how long that person lives), and let T have CDF F and PDF f. The **hazard function** of T is defined as
$$h(t) = \frac{f(t)}{1 - F(t)}$$

1. Explain, why $h(t)$ is the probability density for death at time t

2. Show that an Exponential r.v. has constant hazard function.

Problem 14. (Properties of exponential) Consider independent random variables $X \sim Exp(\lambda_X)$ and $Y \sim Exp(\lambda_Y)$.

1. Find the CDF of $Z = \min\{X, Y\}$.

2. Let $\lambda_X = \lambda_Y$. Find the CDF of $Z = X + Y$. This will be a (very simple case of) **Gamma distribution**. Hint: use the convolution formula.

Note that both results work for any number of random variables, as long as they are mutually independent.

Problem 15* **.** (A failing system) Consider a chandelier with n electric light bulbs, where every light bulb i burns out independently after random time $T_i \sim Exp(\lambda_i)$. Burnt out light bulb is immediately replaced with an identical new one (it has same life expectancy distribution). Show that the number of burn outs of light bulb i until time s is distributed as $Pois(\lambda_i s)$. Hint: use result from previous problem.

Problem 16* **.** (Gumbel distribution) Consider $X \sim Exp(1)$ and $Y = -\log X$. Then Y has the distribution called **Gumbel distribution**.

1. Find its CDF.

2. Consider $X_1, X_2, \ldots, X_n \sim Exp(1)$ and define $M_n = \max\{X_1, \ldots, X_n\}$. Show that the CDF of $Y = (M_n - \log n)$ converges to the CDF Gumbel distribution as $n \to \infty$ (this is what we call convergence of random variables in distribution).