This midterm consists of 2 questions for a total of 100 points + 15 bonus points for 2(g). The duration of the exam is 2 hours. Please show your work for all the questions.

Problem 1 (35 points)

A traveler flies from City 1 to City N following the route 1 → 2 → ⋯ → N. At each airport (from 1 to N − 1), the luggage is lost independently with probability p. Once the luggage is lost, it is not retrieved.

(a) [10 points] What is the probability that the luggage reaches N successfully?

(b) [15 points] Given that the luggage does not reach city N, find the probability that it was lost in city i, i = 1, ⋯ , N − 1. Where is it most likely to have been lost?

(c) [10 points] Now suppose that you’re allowed to choose one city from 1, ⋯ , N − 1 where the probability of the luggage being lost will become 0 (while the probability of loss in the other cities remains p). Which city should you pick to maximize the probability of the luggage reaching N successfully?

Problem 2 (65 points + 15 bonus points)

Hash tables are data structures that allow faster element lookup than arrays or search trees. A hash table assigns each element to a bucket, where a single bucket can store any number of elements. For fast retrieval, the number of elements per bucket should be small (ideally 1). Here we consider a hash table with n buckets (1, ⋯ , n) and m elements (1, ⋯ , m). Each element is independently put into a bucket uniformly at random. We define random variables X_1, ⋯ , X_m and Y_1, ⋯ , Y_n, where X_i is the index of bucket containing the i th element and Y_j is the number of elements in the j th bucket.

(a) [10 points] What are the probability mass functions of X_i, i = 1, ⋯ , m and Y_j, j = 1, ⋯ , n? We do NOT require the joint pmfs here, just write the pmfs for the individual random variables.

(b) [10 points] What is the expectation and variance of the number of elements in the j th bucket, j = 1, ⋯ , n?

(c) [10 points] Find the expected number of buckets with

(i) no elements,

(ii) exactly one element.

(iii) more than one element.

(d) [10 points] Use the union bound to find an upper bound on the probability that at least one bucket has more than 1 element.

(e) [10 points] What is the probability that elements 1 and 2 go into the same bucket? What is the probability that at least one other element goes into the same bucket as 1?

(f) [15 points] Consider (for this part only) an unlimited source of elements which are inserted into the buckets (uniformly at random) one after the other until the first time an element is inserted into a non-empty bucket. Let W be the number of elements inserted (including the element inserted into the non-empty bucket). Find the probability mass function of W.

(g) [15 bonus points] Suppose the m elements are sequentially inserted into the buckets, and we say that a collision occurs whenever an element is inserted into a non-empty bucket. Let Z denote the number of collisions. Find E[Z].