CMPT 210 – Practice Problems Based on Lectures 14–16

Problem 1: Basic Discrete Distributions

Consider the following experiments. For each random variable, identify its distribution (Bernoulli, Uniform, Binomial, or Geometric), specify its support, and write down its PDF. You may introduce parameters as needed (for example, n or p).

(a) You toss a biased coin where $\mathbb{P}(\text{Heads}) = p$. Let X be 1 if the outcome is Heads and 0 otherwise.

(b) You roll a standard k-sided fair die with faces $\{1, 2, \ldots, k\}$. Let Y be the number that appears.

(c) You toss the same biased coin independently n times with $\mathbb{P}(\text{Heads}) = p$ on each toss. Let Z be the total number of Heads in the n tosses.

(d) You repeatedly toss the biased coin with $\mathbb{P}(\text{Heads}) = p$, and stop as soon as you get the first Heads. Let T be the number of tosses required to obtain the first Heads.

(e) For each r.v. above, write its expectation $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\mathbb{E}[Z]$, and $\mathbb{E}[T]$, and its variance $\text{Var}[X]$, $\text{Var}[Y]$, $\text{Var}[Z]$, and $\text{Var}[T]$ in terms of the parameters (k, n, p).

Problem 2: Defective Disks and Money-Back Offers

A company produces computer disks. Each disk is independently defective with probability p. Disks are sold in packages of n disks.

Let D be the random variable equal to the number of defective disks in a package.

(a) What distribution does D follow? State the parameters and write its PDF.

(b) Compute $\mathbb{E}[D]$ and $\mathrm{Var}[D]$.

(c) The company advertises the following money-back guarantee: if a package contains *more than one* defective disk, the customer can return the package for a full refund. Let G be the indicator random variable for the event "package is returned".

i. Express G in terms of D.

ii. Express $\mathbb{E}[G]$ as a probability in terms of D.

iii. Compute $\mathbb{E}[G]$ explicitly as a function of n and p.

(d) Suppose each package sells for \$10 and costs the company c to manufacture. When a package is returned, the company loses the sale (so the revenue is \$0), but still pays the manufacturing cost.

i. Define a random variable R for the revenue from a single package.

ii. Define a random variable P for the profit from a single package.

iii. Compute $\mathbb{E}[R]$ and $\mathbb{E}[P]$ in terms of n, p, c.

(e) (Future section: Tail inequalities) Using Markov's inequality, derive an upper bound on $\mathbb{P}(D \geq 2)$ in terms of n and p and compare this bound to the *exact* probability you found in part (c).

Problem 3: Crashing Programs and Indicator Variables

A server runs n independent programs in parallel. Each program i crashes during a given hour with probability p_i, independently of the others. Define indicator random variables

$$X_i = \begin{cases} 1, & \text{if program } i \text{ crashes in this hour,} \\ 0, & \text{otherwise.} \end{cases}$$

Let $C = \sum_{i=1}^{n} X_i$ be the total number of crashes in that hour.

(a) Identify the distribution of each X_i and compute $\mathbb{E}[X_i]$ and $\mathrm{Var}[X_i]$.

(b) Express $\mathbb{E}[C]$ and $\mathrm{Var}[C]$ in terms of the p_i.

(c) Suppose $p_i = p$ for all i.

   i. Simplify your expressions for $\mathbb{E}[C]$ and $\mathrm{Var}[C]$.

   ii. Identify the distribution of C.

(d) The system administrator defines a random variable

$$T = \begin{cases} 1, & \text{if at least one program crashes (i.e., } C \geq 1\text{),} \\ 0, & \text{otherwise.} \end{cases}$$

   i. Express $\mathbb{P}[T = 1]$ in terms of the p_i.

   ii. Compute $\mathbb{E}[T]$ in terms of the p_i.

(e) (Future section: Markov/Chebyshev) Suppose $p_i = p$ for all i. Derive an upper bound on $\mathbb{P}(C \geq \lambda)$ for a threshold $\lambda > 0$ using:

   i. Markov's inequality.

   ii. Chebyshev's inequality.

   Express your bounds in terms of n, p, λ.

Problem 4: Coupon Collector Variant

A coffee shop has a promotion with n different coupon types (colors). Each time you buy a coffee, you receive one coupon. Each coupon's type is chosen independently and uniformly at random from the n types.

(a) Let T be the total number of coupons you must collect in order to obtain at least one coupon of each type. Define random variables $X_1, \ldots, X_n$ such that

$$T = X_1 + X_2 + \cdots + X_n,$$

where X_k is the number of additional coupons you must collect after having obtained exactly $(k-1)$ distinct types until you obtain a new coupon type. Clearly describe the distribution of each X_k and its parameter(s).

(b) Compute $\mathbb{E}[X_k]$ for each $k \in \{1, \ldots, n\}$.

(c) Use linearity of expectation to express $\mathbb{E}[T]$ as a sum and simplify it as far as you can.

(d) Show that

$$\mathbb{E}[T] \leq n(1 + \ln n).$$

(e) (Future section: Big-O notation and asymptotics) Explain why $\mathbb{E}[T] = \Theta(n \log n)$ in asymptotic notation.

Problem 5: Random Guest Lists and Enemy Pairs

Batman has $2n$ friends $V = \{v_1, \dots, v_{2n}\}$. Some pairs of friends are secret enemies. Let E be the set of enemy pairs, so

$$E \subseteq \{(v_i, v_j) : 1 \leq i < j \leq 2n\}, \qquad |E| = m.$$

Batman does not know which pairs are enemies; he only knows that there are m such pairs.

Alfred suggests inviting a random subset of exactly n friends: each possible subset of size n is chosen with equal probability.

Let X be the number of enemy pairs among the invited guests.

(a) For each enemy pair $e = (v_i, v_j) \in E$, define an indicator random variable I_e that is 1 if both v_i and v_j are invited, and 0 otherwise. Express X in terms of the I_e.

(b) Compute $\mathbb{E}[I_e]$ for a fixed enemy pair e. (Carefully argue the probability that both v_i and v_j are invited under this random process.)

(c) Use linearity of expectation to compute $\mathbb{E}[X]$ in terms of n and m.

(d) Show that there exists *some* guest list of size n such that the number of enemy pairs in that guest list is at most $\mathbb{E}[X]$.

(e) (Future section: Variance) Suppose you are additionally given that $\mathrm{Var}[X] \leq \sigma^2$ for some known σ^2. Using Chebyshev's inequality, derive a lower bound on the probability that a random guest list of size n contains at most $\mathbb{E}[X] + t$ enemy pairs, where $t > 0$ is a parameter you choose.

Problem 6: Load Balancing and Chernoff-Type Reasoning

A website receives N independent requests in a single time interval. Each request is assigned uniformly at random to one of m identical servers, independently of other requests.

For each request $i \in \{1, \dots, N\}$ and each server $j \in \{1, \dots, m\}$, define

$$X_{i,j} = \begin{cases} 1, & \text{if request } i \text{ is assigned to server } j, \\ 0, & \text{otherwise.} \end{cases}$$

Let $L_j = \sum_{i=1}^{N} X_{i,j}$ be the number of requests assigned to server j.

(a) What is the distribution of each $X_{i,j}$? What is $\mathbb{E}[X_{i,j}]$?

(b) Show that $L_j \sim \mathrm{Bin}(N, 1/m)$ and compute $\mathbb{E}[L_j]$ and $\mathrm{Var}[L_j]$.

(c) Using Markov's inequality, give an upper bound on $\mathbb{P}(L_j \geq c\,\mathbb{E}[L_j])$ for some $c \geq 1$.

(d) (Future section: Chernoff) Suppose N is large and m is such that $\mathbb{E}[L_j]$ is at least a few hundred. Sketch how a Chernoff bound of the form

$$\mathbb{P}(L_j \geq c\,\mathbb{E}[L_j]) \leq \exp(-\beta(c)\,\mathbb{E}[L_j])$$

with $\beta(c) > 0$ would provide a much tighter bound than Markov's inequality. You do *not* need to derive the exact function $\beta(c)$; just compare the "shape" of the two bounds.

(e) Use the union bound to express an upper bound on the probability that *some* server is assigned at least $c\,\mathbb{E}[L_j]$ requests.

Problem 7: Counting Patterns in Coin Tosses

You are given a biased coin that shows Heads with probability p on every toss, independently of other tosses. You toss the coin n times, where $n \geq 3$.

Let H_i denote the outcome of toss i, where $H_i = 1$ if the i-th toss is Heads and $H_i = 0$ otherwise. We are interested in the number of occurrences of the pattern HTH in the n-toss sequence.

(a) For each position $i \in \{1, 2, \ldots, n-2\}$, define an indicator random variable X_i that is 1 if the pattern HTH starts at position i, i.e.,

$$(H_i, H_{i+1}, H_{i+2}) = (1, 0, 1),$$

and 0 otherwise. Write down X_i explicitly as a function of H_i, H_{i+1}, and H_{i+2}.

(b) Let $T = \sum_{i=1}^{n-2} X_i$ be the total number of occurrences of HTH. Express $\mathbb{E}[T]$ using linearity of expectation.

(c) Compute $\mathbb{E}[X_i]$ explicitly in terms of p. (Hint: use independence of the coin tosses.)

(d) Use your previous answers to obtain a closed-form expression for $\mathbb{E}[T]$ in terms of n and p.

(e) (Future section: Covariance and variance of sums) Briefly explain why the X_i's are *not* mutually independent. Identify at least one pair X_i, X_j that are dependent and one pair that are independent.