AIM Qualifying Review Exam: Probability and Discrete Mathematics

January 6, 2025

There are five (5) problems in this examination.

There should be sufficient room in this booklet for all your work. But if you use other sheets of paper, be sure to mark them clearly and staple them to the booklet.

Problem 1

Find, as function of n, the sum

$$1 + \frac{1}{2}\binom{n}{1} + \frac{1}{3}\binom{n}{2} + \cdots + \frac{1}{n+1}\binom{n}{n}.$$

Hint: try integrating a series for $(1+x)^n$.

Problem 1

Problem 1

Problem 1

Problem 2

An r-combination of a multiset M is an unordered collection of r items in M.

(a) Let T^* be the multiset $\{\infty \cdot a, \infty \cdot b, \infty \cdot c\}$ (infinitely-many a's, b's, and c's). Determine the the number of 10-combinations of T^*. (So, for example, $\overbrace{aaaaaaaaa}^{9}b$ and $b\overbrace{aaaaaaaaa}^{9}$ are different names for the same valid 10-combination, since they consist of a, b, c but have the same number of each letter and differ only in the order.)

(b) Determine the number of 10-combinations in the (size 12) multiset $T = \{3 \cdot a, 4 \cdot b, 5 \cdot c\}$, consisting of three a's, four b's, and five c's.

(c) Determine the number of 10-combinations in $T = \{3 \cdot a, 4 \cdot b, 8 \cdot c\}$, consisting of three a's, four b's, and eight c's.

Problem 2

Problem 2

Problem 2

Problem 3 Let $f(x, y) = 24xy$ on the triangular region $0 \leq x, y \leq x + y \leq 1$.

(a) Show that f is a joint probability density function.

(b) Find $E[X]$, where X is the random variable associated with x under f.

(c) Find $E[Y]$, where Y is the random variable associated with y under f. Solve using the above without any new computation.

(d) Are X and Y independent?

Problem 3

Problem 3

Problem 3

Problem 4 A surveyer is knocking on doors in Ann Arbor, collecting answer to the sensitive question, "are you rooting for Ohio State?". Respondents are to follow this protocol:

- Flip a coin C_1 with heads probability p.
- If C_1 is heads, answer YES or NO truthfully.
- If C_1 is tails, flip another coin, C_2, with heads probability $\frac{1}{2}$, and answer YES or NO according to C_2.

Suppose n people are surveyed and Suppose $cn \leq n$ people have true answer YES. Let X denote the total number of YES answers given to the survey (which is often not exactly cn).

(a) Suppose Alice is surveyed. Let f_{NO} be the probability mass function for X conditioned on Alice's true answer equal to NO and let f_{YES} be the probability mass function for X conditioned on Alice's true answer equal to YES. Find $b = \sum_x |f_{\text{YES}} - f_{\text{NO}}|$ that holds over all possibilities of others' answers. (This protects Alice's privacy.)

(b) Find $\mu = E[X]$. (This and the below insure that the data gathered is useful.)

(c) Find $\sigma^2 = \text{Var}[X] = E[(X - \mu)^2]$.

(d) The Chebyshev inequality says that, for any random variable Y with mean μ and standard deviation σ, and any non-negative k, we have $\Pr(|Y - \mu| > k\sigma) \leq \frac{1}{k^2}$. Given b as above and if we want $\Pr(|X - E[X]| > \frac{1}{10}n) \leq \frac{1}{100}$, find p and n to satisfy requirements, as guaranteed by Chebyshev.

Problem 4

Problem 4

Problem 4

Problem 5

The fruit orange used to be called norange in English like naranja in Spanish; after saying "a norange" many times, English shifted to "an orange."

Suppose we are given a string of letters without spaces, like anorange, and the goal is to insert spaces to maximize the sum of quality scores of the substrings. For example, in modern English, presumably quality(an) + quality(orange) is greater than either quality(a) + quality(norange) or quality(an) + quality(ora) + quality(nge). Note that the number of spaces is not fixed, but optimized, along with the placement of spaces. Qualities may be positive or negative. Ignore any consideration about whether the string of words makes sense; e.g., a/no/range consists of high quality individual words, even if the string of words is less plausible than an/orange.

Give an algorithm that takes a string of n symbols, has access to quality() as a unit-cost black box for single substrings (potential words), and, in time polynomial in n, finds a splitting that maximizes the sum of the qualities. Briefly show correctness and efficiency, including finding a in the runtime $O(n^a)$.

Problem 5

Problem 5

Problem 5