6.1200J/18.062J *Mathematics for Computer Science* — Wednesday 1st May, 2024
Massachusetts Institute of Technology, Spring 2024
Z. Abel, B. Chapman, E. Demaine
revised Tuesday 30th April, 2024

Problem Set 10

- **Due Date:** 11:59pm on **Friday 10th May, 2024**
- **Days Covered**: 20, 21, and 22 (including Lecture, Warm-Up, and Recitation)

Problem 1. Random Pixels [10 points]

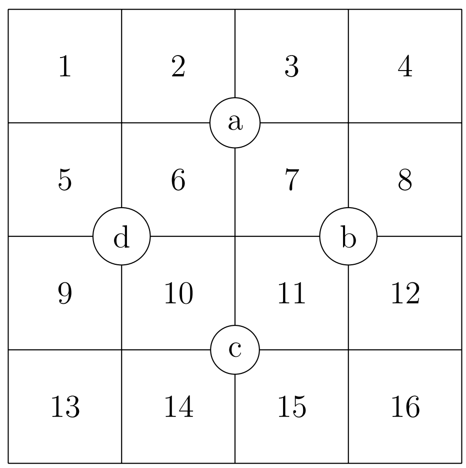

Figure 1: Graph G consisting of four squares.

Let G be the grid shown in figure 1. Each of the numbered cells of the grid will be randomly assigned one of the colors cyan, magenta, yellow, or black, with probability $1/4$ for each color. The assignments of colors to cells are mutually independent. We call a vertex in this grid *monochromatic* if all four adjacent cells have the same color; let A be the event that a is monochromatic, and similarly with events B, C, and D referring to b, c, and d respectively.

(a) [2 pts] What is $\Pr[A]$, i.e., the probability that all four cells around a have the same color?

(b) [6 pts] Prove carefully that events A, B, and C are mutually independent. (Since the same is true for the other sets of 3 squares, this will show that events A, B, C, and D are 3-wise independent.)

Hint: To verify that A, B, and C are mutually independent, there are 4 formulas that need to be checked.

(c) [2 pts] Prove that A, B, C, and D are **not** mutually independent.

Problem 2. Median vs Maximum [20 points]

A hat is filled with 100 Raboots, numbered 1 through 100 (with no repeats). Three Raboots are pulled out of the hat uniformly at random without replacement[1]; let D be the **median** of their three numbers, and let X be the **maximum** of their three numbers.

Note: for this problem, please express your answers in closed form: arithmetic, exponents, factorials, and binomial coefficients are fine, but your final answers should not use $\sum$-notation, ellipses, or the like.

(a) [6 pts] What are the values of $\text{PMF}_D(k)$ and $\text{PMF}_X(k)$ for each $1 \leq k \leq 100$? Please answer in closed form, and explain your answers.

(b) [6 pts] What is the value of $\text{CDF}_D(k)$ for each $1 \leq k \leq 100$? Please answer in closed form, and explain your answer.

Hint: Recall that your answer should not use $\sum$ notation—in fact, we recommend not using summations for this problem at all! Instead, think about how many of the three numbers are $\leq k$. There are two cases to consider.

(c) [2 pts] Observe that D can take any value between 2 and 99, while X can take any value between 3 and 100. What is $\Pr[D = 10 \text{ AND } X = 5]$? Explain your answer.

(d) [2 pts] Are random variables D and X independent? Why or why not?

(e) [4 pts] For which real numbers k is $\text{PMF}_{X-D}(k)$ greater than 0? (You do not need to compute the exact value of $\text{PMF}_{X-D}(k)$.) Please explain your answer.

Problem 3. Probabilistic Method [10 points]

Let $x_1, \ldots, x_n$ be a collection of n boolean variables. Recall that a *literal* is either a variable x_i or its negation $\overline{x_i}$.

Let ϕ be a k-CNF formula with m clauses: that is, a formula of the form

$$\phi = C_1 \wedge C_2 \wedge \cdots \wedge C_m$$

where each clause C_j is of the form $(\ell_1 \vee \ell_2 \vee \ldots \vee \ell_k)$ where $\ell_1, \ell_2, \ldots, \ell_k$ are literals. For this problem, assume that each clause has literals from k **distinct** variables.

Imagine that we assign a value in {True, False} to each variable x_i by flipping a fair coin; the n coin flips are made independently.

(a) [2 pts] Consider a single clause of the form $(\ell_1 \vee \ell_2 \vee \ldots \vee \ell_k)$. What is the probability that it is satisfied, i.e. that at least one of its literals is True?

[1] "Without replacement" means that the same Raboot cannot be chosen twice.

(b) [2 pts] Let U be the random variable which counts the total number of unsatisfied clauses. Express U as a sum of indicator random variables, with one indicator random variable for each of the m clauses. Be careful to write down an expression for U, not $\text{Ex}[U]$.

(c) [2 pts] What is the expected number of unsatisfied clauses, $\text{Ex}[U]$?

Hint: Use linearity of expectation.

(d) [2 pts] Using the previous part, prove that if $k = 5$ and $m = 31$ then $\Pr[U = 0] > 0$.

(e) [2 pts] Conclude that if $k = 5$ and $m = 31$ then there exists a satisfying assignment: some way of assigning values {True, False} to the variables so that all of the clauses in ϕ are satisfied.

Problem 4. Teeter Dance [10 points]

An over-caffeinated Spinda teeters along the (actually-)Infinite Corridor. With each step, Spinda randomly moves East or West with equal probability.

We designate Spinda's initial position in Lobby 10 as location 0, with successive positions to the East labeled $1, 2, \ldots,$ and positions to the West labelled $-1, -2, \ldots$. Let L_t be the random variable giving Spinda's location after t steps. Before it starts, Spinda is known to be at location 0, so

$$\text{PMF}_{L_0}(n) = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases}$$

After one step, Spinda is equally likely to be at location 1 or -1, so

$$\text{PMF}_{L_1}(n) = \begin{cases} 1/2 & \text{if } n = \pm 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) [2 pts] Give the distributions PMF_{L_t} for $t = 2, 3, 4$ by filling in the table of probabilities below, where omitted entries are 0. For each row, write all the nonzero entries so they have the same denominator.

	location								
	-4	-3	-2	-1	0	1	2	3	4
initially					1				
after 1 step				1/2	0	1/2			
after 2 steps			?	?	?	?	?		
after 3 steps		?	?	?	?	?	?	?	
after 4 steps	?	?	?	?	?	?	?	?	?

(b) [8 pts] Help Spinda regain its bearings by answering the following questions. Provide your derivations and reasoning. (All answers should be functions of t and i.)

(i) What is the final location s of a t-step walk that steps East exactly i times?

(ii) How many different length-t walks end at this same location s?

(iii) What is the probability that Spinda ends at this same location s after t steps?

(iv) Let $B_t ::= (L_t + t)/2$. Conclude that B_t follows the binomial distribution $\text{Bin}_t(1/2)$.