SEARCH PROBLEMS

[CONTINUED ...]

ASSIGNMENT/QUIZ 1

- Later this week, you will have assignment 1.
- Release Date: Friday, August 9.
- Due: Monday, August 12.
- Covers Weeks 1 to 3.
- Multiple-choice quiz over Wattle.

Keep an eye on Wattle announcements for further instructions!

EGG DROPPING PROBLEM

Input: There is a building with n floors.

Output: You want to find the **highest** floor that if you drop an egg from it doesn't break.

Assumptions:

- If the egg breaks when dropped from floor k, then it would also have broken from any floor above that.
- If the egg survives a fall, then it will survive any fall below that.

Objective: *Minimize the number of required drop tests (trials).*

SCENARIO 3

What if you are given 2 identical eggs?

Binary Search Approach

- Drop the egg from floor $\lceil \frac{n}{2} \rceil$.
- If it breaks, then we have to apply the strategy with 1 egg for floors $1, \cdots, \left\lceil \frac{n}{2} \right\rceil - 1$.
- This takes $\Omega(n)$ in the worst case.

Can we do better?

SCENARIO 3: A MORE OPTIMAL SOLUTION

- Roughly speaking, we want to minimize our maximum regret.
- Suppose we drop our egg from floor d.
- If it breaks, we step through $(d-1)$ floors one by one.
- If not, instead of jumping another d floors, we should step up just $d-1$ floors because we have one less drop available.
- Thus, the next floor should be $d+(d-1)$.
- If it breaks, we do the floor-by-floor search, otherwise we jump to $d+(d-1)+(d-2)$, and so on.

SCENARIO 3: A MORE OPTIMAL SOLUTION

- The number of drops in this algorithm is at most d by design.
- However, to ensure that we will cover all floors, we need
$$d + (d-1) + (d-2) + \cdots + 1 \geq n$$
- Thus, we are interested in the minimum d which satisfies the above equation.
- The left-hand side is $\frac{d(d+1)}{2}$. It is easy to see that there is a d in $O(\sqrt{n})$, which satisfies this equation.
- Thus, we need $O(\sqrt{n})$ trials.

PROBABILISTIC ANALYSIS
[CLRS SEC. 5.2]

PROBABILISTIC ANALYSIS

Two randomized setups:

- The algorithm is deterministic, but the input is from a random distribution.
- Randomized algorithm: the algorithm itself makes random choices.

Our objective is to analyze the **expected running time** for such algorithms.

TOPICS

- Probability Refresher
- Probabilistic Analysis of Deterministic Algorithms
- Probabilistic Analysis of Randomized Algorithms

PROBABILISTIC MODELS

Probabilistic Modelling Assumptions:

- Experiments with random outcome
- Which outcome is likely to happen is quantifiable

Probabilistic modelling consists of three components:

- Sample Space (S): Set of all possible outcomes.
- Events ($\mathbf{E}$): Subsets of sample space.
- Probability (P): “Quantifies” the likelihood of an event.

PROBABILISTIC MODELS

Probability: A function that maps events to real numbers satisfying the following 3 axioms:

- Non-negativity: $P(\mathrm{A}) \geq 0$ for any event $A \in \mathbf{E}$
- $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$ where $A_i \in \mathbf{E}$ are disjoint (mutually exclusive) events.
- Normalization: $P(S) = 1$.

EXAMPLE

Flipping a fair coin, head (H) or tail (T).

$S = \{H, T\}$

$\mathbf{E} = \{\emptyset, \{H\}, \{T\}, \{H,T\}\}$ and the probabilities are:

$P(\emptyset) = 0$, $P(\{H\}) = P(\{T\}) = 0.5$, $P(\{H,T\}) = 1$

For compactness of writing, we usually do not list zero and one probabilities. Thus, we write the above as

$S = \{H, T\}$; $P(\{H\}) = P(\{T\}) = 0.5$

EXAMPLE

Biased dice

- Each even face is 2 times as likely as each odd face.
- All even faces are equally likely.
- All odd faces are equally likely.

What is the probability that the outcome is less than 4 ?

$P(\{1\}) = P(\{3\}) = P(\{5\}) = x$

$P(\{2\}) = P(\{4\}) = P(\{6\}) = 2x$

$3x + 3 * 2x = 1 \rightarrow x = 1/9.$

$P(\{1\} \cup \{2\} \cup \{3\}) = P(\{1\}) + P(\{2\}) + P(\{3\}) = x$
$2x + x = 4x = \frac{4}{9}.$

RANDOM VARIABLES

Probability assigns likelihood to subsets of the sample space (events).

In many cases, it's easier if the sample space has an ordering structure.

Random variable X is a function that maps sample space S to a number.

RANDOM VARIABLES: DISCRETE

S: Students of COMP3600/6466

P: We choose a student uniformly at random

X: The program a student is enrolled in

Suppose the programs are indexed, e.g., BAC: 0, BIT: 1, Eng: 2, PhB: 3, MCompSci: 4

X(student 1) = 3 ; X(student 2) = 0 ; X(student 3) = 0 ; ….

P(X = 0) means the probability that the randomly chosen student is enrolled in BAC.

$$P(X = 0) = \frac{\#\ BAC\ students}{\#\ students}$$

EXPECTED VALUE

Expected value is the weighted average of possible values of X, where the weight is the probability:

$$E[X] = \sum_{\forall x} x\, P(X = x)$$

$$E[g(X)] = \sum_{\forall x} g(x)\, P(X = x)$$

Linearity of expectation for random variables X and Y and constants a and b:

$$E[aX + b] = aE[X] + b$$
$$E[X + Y] = E[X] + E[Y]$$

INDICATOR RANDOM VARIABLES

Suppose we have sample space S and an event A, then:

- The indicator random variable associated with event A is

$$\mathrm{I}\{A\} = \begin{cases} 1 & A \ occurs \\ 0 & A \ does \ not \ occur \end{cases}$$

- Let $X_A = \mathrm{I}\{A\}$, then $E[X_A] = P(\{A\})$

Proof:

$$E[X_A] = E[\mathrm{I}\{A\}] = 1 * P(\{A\}) + 0 * P(S \backslash \{A\}) = P(\{A\})$$

INDICATOR RANDOM VARIABLES: EXAMPLES

S = {H, T}; P({H}) = P({T}) = 0.5

$$X_H = \mathrm{I}\{\{H\}\} = \begin{cases} 1 & \{H\}\ occurs \\ 0 & \{T\}\ occurs \end{cases}$$

$$E[X_H] = E[\mathrm{I}\{H\}] = 0.5$$

INDICATOR RANDOM VARIABLES: EXAMPLES

Expected number of heads obtained when flipping a fair coin n times?

Let x_i be the indicator random variable that head occurs in the i^{th} flip, and random variable X be the number of heads in n flips.

- $x_i = \mathrm{I}\{\{H_i\}\} = \begin{cases} 1 & head\ occurs\ in\ the\ i^{th} flip \\ 0 & tail\ occurs\ in\ the\ i^{th} flip \end{cases}$
- $E[X] = E[\sum_{i=1}^{n} x_i]$

$$
\begin{aligned}
&= \textstyle\sum_{i=1}^{n} E[x_i] \quad \text{(based on linearity of expectation)} \\
&= \textstyle\sum_{i=1}^{n} 0.5 \quad \text{(based on results in previous slide)} \\
&= 0.5n.
\end{aligned}
$$

NUMBER OF MATCHES

From a randomly shuffled deck of size n, cards are laid out on a table, face up from left to right, and then another shuffled deck is laid out so that each of its cards is beneath a card of the first deck.

What is the expected number of matches of the card above and the card below?

NUMBER OF MATCHES

Each card in the first deck has 1 chance in n of matching its paired card.

Label the cards from 1 to n. Then, define the indicator random variable x_i to be 1 if and only if card number i is matched. Then

$$E[x_i] = P(x_i = 1) = \frac{1}{n}$$

The total number of matches is $X = \sum_{i=1}^{n} x_i$. Thus, the expected number of matches is

$$E[\sum_{i=1}^{n} x_i] = \sum_{i=1}^{n} E[x_i] = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

COUPON COLLECTOR PROBLEM

There are n types of coupons and we want to collect at least one from each type. In each trial, we are given a coupon uniformly at random among all types of coupons.

Question: How many coupons do we need to buy in expectation to have all n types?

COUPON COLLECTOR PROBLEM

Let random variable X be the number of trials we need. We are interested in $E[X]$.

Let x_i be the number of coupons we need to buy to get a new coupon while having $i - 1$ types of coupons. Then, using linearity of expectation

$$E[X] = E\left[\sum_{i=1}^{n} x_i\right] = \sum_{i=1}^{n} E[x_i].$$

COUPON COLLECTOR PROBLEM

If we own $i-1$ coupons, the probability of picking a new coupon is exactly

$$p_i = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$$

Random variable x_i is **geometrically distributed** with parameter p_i. Thus, we have

$$E[x_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$$

COUPON COLLECTOR PROBLEM

$$E[X] = \sum_{i=1}^{n} E[x_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \sum_{j=1}^{n} \frac{1}{j}$$

$\sum_{j=1}^{n} \frac{1}{j}$ is the n-th **Harmonic Number** and equal to $\log n + c$ for some constant c.

Hence, $E[X] = n \log n + cn$.