Probabilistic Analysis

Department of Computer Science, Tsinghua University

PA & RA

- Probabilistic Analysis
  - Using probabilistic models (Appendix C-3, C-4, Ch 5.4.2)
  - Using indicator random variable (Ch 5.2)
  - Bucket sort (Ch 8.4)
  - Hiring problem (Ch 5.1) & on-line hiring problem (Ch 5.4.4)
- Randomized Algorithm
  - Randomized hiring problem (Ch 5.3 p122-p124)
  - Randomized Select (Ch 9.2)

Probabilistic Analysis

- Probabilistic analysis is the use of probability in the analysis of problems.
- We often use probabilistic analysis to analyze the running time of an algorithm.
  - We call it ***expected running time***: it is taken over the distribution of the possible inputs, i.e., averaging the running time over all possible inputs (***average-case running time***).
- Two ways of probabilistic analysis:
  - Using probabilistic models directly
  - Using indicator random variables

Probabilistic Models

- There is a tight relationship between execution of algorithms and experiments:
  - Bernoulli Trails
  - Geometric Distribution
  - Binomial Distribution
  - Balls & Bins
  - …

Jakob Bernoulli

Random Variable

- ***A random variable*** X: is a function from a finite or countably infinite sample space S to the real numbers. It associates a real number with each possible outcome of an experiment.
- ***Expectation***: the expected value of a discrete random variable X is $E[X] = \sum_x xPr\{X = x\}$.

-- *the "average" of the values X takes on.*

- **Example:** The possible outcomes for one coin toss can be described by the sample space *Ω = {heads, tails}.* We can introduce a real-valued random variable Y as follows:
  - $Y(\omega) = \{^{1,\, if\ \omega = head,}_{0,\, if\ \omega = tail.}$
  - $Pr\{Y = y\} = \{^{0.5,\, if\ y = 1,}_{0.5,\, if\ y = 0.}$
  - $E[Y] = 1 \times \Pr\{Y = 1\} + 0\ \times \Pr\{Y = 0\} = 0.5$

Bernoulli Trial

- ***A Bernoulli Trial***: an experiment with only two possible outcomes: success (with a probability of p) and failure ($q = 1 - p$)
- If we treat the outcome of Head as a success and Tail as a failure, coin toss is a Bernoulli trial.

- ***A sequence of Bernoulli trials***:
  - Each with a probability of p for success
  - Q: How many trials occur before we obtain a success?
- Let X be the number of trials needed to obtain a success.
  - $Pr\{X = k\} = q^{k-1}p$ (geometric distribution)
  - $E[X] = \sum_{k=1}^{\infty} kq^{k-1}p = \frac{p}{q}\sum_{k=0}^{\infty} kq^k = \frac{p}{q} \times \frac{q}{(1-q)^2} = \frac{1}{p}$

Balls & Bins

- Balls and Bins
  - Toss identical balls into b bins.
  - Tosses are independent.
  - Each toss the ball is equally likely ($= \frac{1}{b}$) to end up in any bin.
  - Useful model for analyzing hashing (Ch 11) and searching (problem 5.2) etc.

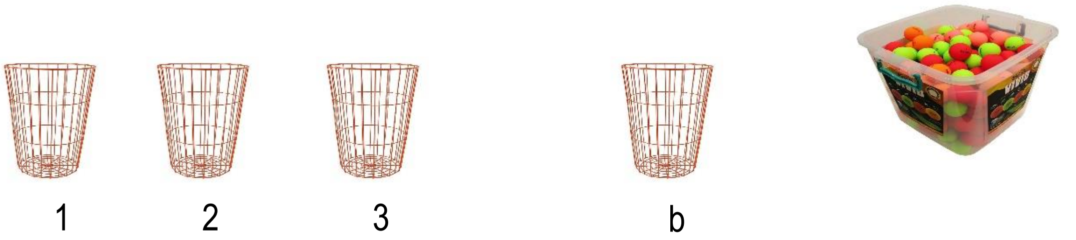
1 2 3 b

Balls & Bins

- How many trials we need to have one ball fall in a given bin?
  - Let X be the number of trials needed to hit a given bin.

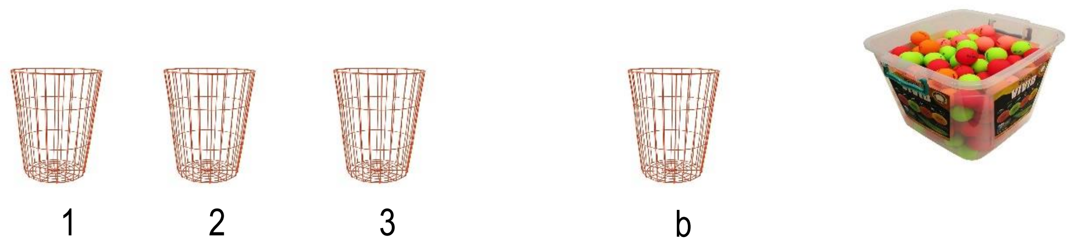
Balls & Bins

- How many balls must be tossed until every bin contains at least one ball?
  - Consider b stages:



Balls & Bins

- How many balls must be tossed until every bin contains at least one ball?
  - Consider b stages
  - At the i^{th} stage, $i-1$ bins contains balls & $b-i+1$ bins are empty, let n_i be the number of tosses to hit a new bin.
  - The probability of a success toss at the i^{th} stage is:
  - $E[n_i] =$
  - $E[n_1 + n_2 + \cdots + n_b] = \sum_{i=1}^{b} E[n_i] =$



In-class exercise

- 5.2 (b) and (d): Searching for a value x in an unsorted array A consisting of n elements.
- Strategy: pick a random index i into A. If $A[i] = x$, then we terminate; otherwise we continue the search by picking another index, until we find an index j such that $A[j] = x$ or we have checked all elements in A.
- Q1: Suppose there is exactly one index i such that $A[i] = x$. What is the expected number of indices into A we must pick before we find x?
- Q2: Suppose there are no indices i such that $A[i] = x$. What is the expected number of indices into A that we must pick before we have checked all elements of A?

Indicator Random Variable

- *Definition*
  - Given a sample space S and an event A. the indicator random variable $I\{A\}$ or X_A associated with event A is defined as: $I\{A\} = X_A = \begin{cases} 1 & if\ A\ occurs, \\ 0 & otherwise. \end{cases}$
  - $E[X_A] = 1 * Pr\{A\} = Pr\{A\}$
  - The random variable Y in the toss coin example is also an indicator random variable.

- *Probabilistic Analysis*
  - Expected (Average-case) running time or the number of executions of certain statements
  - $E[X+Y] = E[X] + E[Y]$ (X *and* Y can be *dependent*)
  - Simplified the calculation by using indicator random variables

Inversions

- **Example:** What is the expected number of inversions in an input array of size n for insertion sort? Assuming all permutations of the input array are equally likely.
- Why?
  - The running time of insertion sort is in proportion to the number of inversions.
- Let X be the number of total inversions, $X_{ij} = \mathrm{I}\{i, j$ is a pair of inversion$\}$ (i.e., $i < j$ and $A[i] > A[j]$).
- $X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}$
- $E[X_{ij}] = \frac{1}{2}$
- $E[X] =$

Group Exercise: 5.2 (e)

- Search A for x in order, consider $A[1], A[2], \cdots, A[n]$ until either find $A[i] = x$ or reach to the end of the input array.
- Q: suppose there is exactly one index i such that $A[i] = x$. What is the average-case running time of this algorithm?