Randomization

CS 4231, Fall 2017 Mihalis Yannakakis

Randomized algorithms

- Make random choices (coin flips, random numbers ..)
- Different random choices are assumed independent
- Outcome of algorithm and running time depends on random choices (besides input)

- Correctness: Show termination and correct answer for all random choices

Basic probability concepts

- Sample space: Set of all possibilities (=sample points).
- We'll deal with finite, or countable sample spaces.

- Examples: Flip 100 coins: 2^{100} sample points=$\{H,T\}^{100}$
  – Inputs of size n=100 numbers in a range $[0,10^{20}]$ (eg. sorting)
  – Set of possible random choices w of a randomized algorithm when run on a specific input

- Probability of each sample point: sum=1
  – Ex. 100 independent unbiased coin flips:
    probability of each sample point (outcome)= 2^{-100}

Basic probability concepts

- Event A = subset of sample space
- Pr (A) = sum { Pr(s) | s $\in$ A }
  - Ex. A = 3 heads in 100 coin flips where coin bias =p:

$$\Pr(A) = \binom{100}{3} p^3 (1-p)^{97}$$

- Independent events A, B:

  Pr(A and B) = Pr(A)Pr(B)

Random variables , Expectation

- Random variable X: Maps Sample space S to R
  - Ex: # heads : binomial distribution
  - Time t(I) of a deterministic algorithm for input I of size n
  - t(I,w) of a randomized algorithm for particular input I, random choices w
- If X is a discrete random variable (eg. #Hs in coin flip, or time *t(I,w)* for particular input I, random choices w) , then $E[X] = \Sigma_x x \text{ Prob}[X=x]$
- Average case (expected) time complexity of deterministic algorithm for an input probability distribution (probability distribution on inputs of size n for every n):

$$\overline{T}(n) = E_{|I|=n}\ t(I)$$

Time complexity of randomized algorithms

- Running time: Depends on input I and random choices w : time $t(I,w)$
- Expected running time for an input I:

  expected time w.r.t. random choices w: $\bar{t}(I) = E_w\, t(I,w)$
- Expected time complexity of the algorithm

  Two versions:

  – Worst-case expected time $T(n) = \max_{|I|=n} E_w\, t(I,w)$

  (worst-case expected time over inputs of size n)

  – Average-case expected time $\bar{T}(n) = E_{|I|=n} E_w\, t(I,w)$

  assumes a probability distribution on inputs of size n, expected time also w.r.t. inputs

Expectation – Indicator r.v.

- Linearity of expectations E[X+Y] = E[X]+E[Y]
- Often it is hard to compute the probabilities Prob[X=x] and do the summation in the defn of E[X]
- If can decompose X as sum $Z_1 + ... + Z_m$ of simpler variables then $E[X] = E[Z_1] + ... + E[Z_m]$
- Indicator random variable I(A): indicates if event A happens: 1 if it happens, 0 if it does not
- E[I(A)] = 1 Prob(A) + 0 Prob(not A) = Prob(A)
- Example: n coin flips with bias p, X = #H's

$$E[X] = \sum_{i=1}^{n} i \binom{n}{i} p^i (1-p)^{n-i} = np$$

- Z_i = I(i-th flip=H). $E(Z_i)=p$
- $X = Z_1 + ... + Z_n \Rightarrow E[X] = np$

Example: Hiring problem

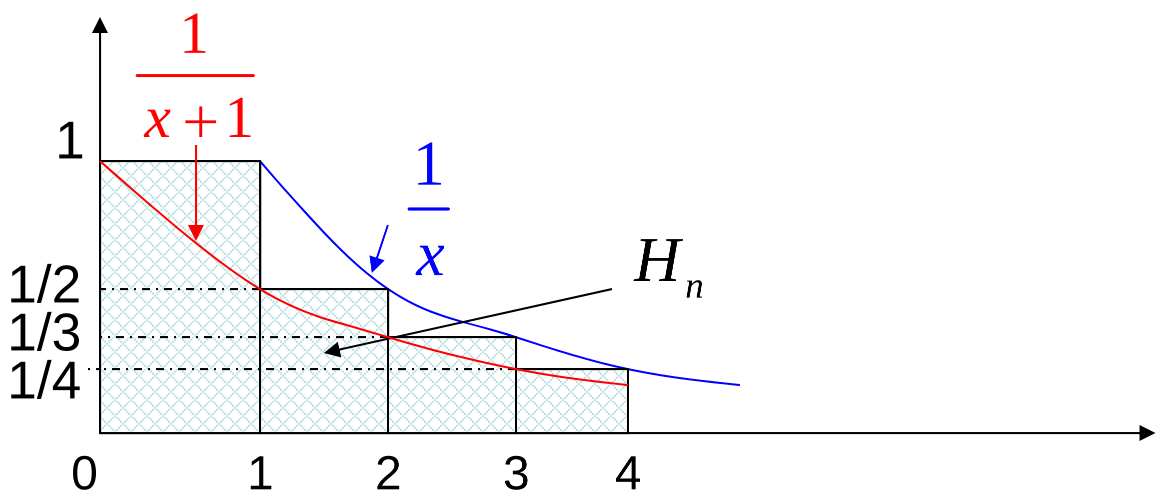
- Interview n candidates for assistant in order, hire candidate i if better than current asst and fire current asst.
- Sample space: orderings of candidates (permutations)
- Probability Distribution: uniform: all same probability =1/n!
- Random variable X: #times hire new asst.

- Indicator variable Z_i = I(hire i-th candidate)
- $E(Z_i)$ = Prob. that the best among candidates {1,…i} is i = 1/i

$$E[X] = \sum_{i=1}^{n} \frac{1}{i} = H_n \approx \ln n$$ *Harmonic series*

Harmonic series

$$E[X] = \sum_{i=1}^{n} \frac{1}{i} = H_n \approx \ln n$$



$$H_n > \int_0^n \frac{1}{x+1} dx = \ln(n+1) > \ln n$$

$$H_n < 1 + \int_1^n \frac{1}{x} dx = 1 + \ln n$$

Birthday Paradox

- How many people do you need until you expect to find two people with the same birthday?
- Answer: << number *n* =365 of days - Grows as $\Theta(\sqrt{n})$, assuming birthdays are independent uniformly random.
- 23 people $\Rightarrow$ at least ½ probability of a birthday coincidence
- Proof: If k people, then prob(no coincidence)=

$$\frac{n(n-1)\cdots(n-k+1)}{n^n} = 1(1-\frac{1}{n})\cdots(1-\frac{k-1}{n})$$

Since $1+x \leq e^x$, probability is $\leq e^{-(1+2+\cdots k-1)/n} = e^{-k(k-1)/2n}$

which is ≤ ½ if $k \geq 1+\sqrt{1+(8\ln 2)n})/2$

For n=365, k ≥ 23 suffices.

Birthday paradox via indicator r.v.

- X_{ij} = I(persons *i* and *j* have the same birthday)
- E[X_{ij}] = Prob(X_{ij})= n/n^2 = 1/n
- E(#pairs with same birthday)= E(ΣX_{ij}) =

  = ΣE(X_{ij}) = $k(k-1)/2n$

$\Rightarrow$ If $k \geq \sqrt{2n} + 1$ then E(#pairs with same birthday) $\geq$ 1

For n=365, k$\geq$28 suffices