Announcements

Count	DMSSI	9	16-11-16
#9		

(Cormen) Hire Assistant	p. 115
Ind. var.	weekly note 4
(Cormen) Coupon coll.	p. 134
Geo. dist.	p. 470
(TK) Chernoff	p. 759

Today

Analysis of different randomized algs.

Use of indicator variables.

Chernoff bounds.

Cormen 5.2-1 Hire-Assistant
p. 122

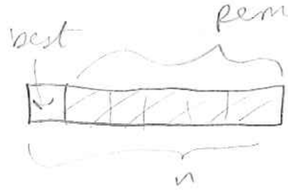
▷ Recall Hire-Assistant: want the best candidate. Interview in random order and hire if better than previously best cand. How many times do we hire and fire the previous? (expected)

Last time we saw online bidding agents and it is essentially the same. Here we saw that we exp to "hire" $O(\log n)$ times.

▷ Now: Prob of hiring exactly once?

↳ Means best is first,

▷ How many ways can that happen?



$\Rightarrow (n-1)!$

▷ $\text{prob} = \frac{(n-1)!}{n!} = \frac{1}{n}$

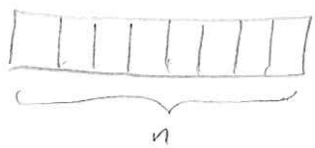
▷ Prob of hiring n times?

↳ Means they come in ascending order

▷ How many ways? 1 out of $n!$ $\Rightarrow$ $\text{prob} = \frac{1}{n!}$

Cormen 5.2-5

Let $A[1..n]$, array of distinct numbers



If for some $i < j$, $A[i] > A[j]$, then (i,j) is an <u>inversion</u>.

A contains random perm of $\{1, \dots, n\}$

Use ind. vars. to compute # of inversions.

▷ what should we create ind. vars. for? $X_{?} = \begin{cases} 1 & \text{if } ? \\ 0 \end{cases}$

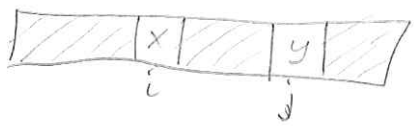
We try $X_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ is an inv} \\ 0 \end{cases}$,

▷ How many X_{ij} do we have?

we only create X_{ij}'s for $i < j$, so for $i = 1$ we have $n-1$, for $i = 2$ we have $n-2$ ..., upto $i = n-1$ where we have 1. This can be expressed as $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 1 = \frac{n \cdot (n-1)}{2} = \binom{n}{2}$

▷ For each X_{ij}, what is $P(X_{ij} = 1) = E(X_{ij})$?

↳ consider some i, j



$x, y \in \{1, \dots, n\}$. If $x > y$, (i,j) is an inv.

x can take on n different values, and

y can take on $n-1$ of the remaining values. $\Big\} \Rightarrow n \cdot (n-1)$ total outcomes

We are interested in those cases where $x > y$ which there are $\frac{n \cdot (n-1)}{2}$ of (same logic as above).

$E(X_{ij}) = \frac{\frac{n \cdot (n-1)}{2}}{n (n-1)} = \frac{1}{2}$

▷ $X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}$, counts all inversions

▷ $E(X) = \sum \sum E(X_{ij}) = \frac{1}{2} \sum \sum 1 = \frac{1}{2} \frac{n \cdot (n-1)}{2} = \frac{n \cdot (n-1)}{4}$

▷ Fun fact: The running time of insertion sort and bubble sort is the same as the number of inversions in A as they undo all of these step by step.

Cormen 5.3-4 Algs for random perm
p. 129

Previously we have seen ways to create a random perm of array A, for instance RANDOMIZ-IN-PLACE:

```
n = len(A)
for i = 1 to n:
    swap A[i] and A[random(i, n)]
```

random number between i and n.

▷ Now Prof. Armstrong suggests:

```
PERM-BY-CYCLIC(A):
  n = len(A)
  B : new array[1.. n]
  off set = random(1, n)
  for i=1 to n
      dest = i + off set
      if dest > n          } modulo
          dest = dest - n
      B[dest] = A[i]
  return B
```

▷ Understanding alg: It picks a random offset and shifts everything to the right (with wrap around) to get B.

▷ Does this give prob $\frac{1}{n!}$ for each possible perm?

↳ No. For some given A we can only get n diff B's where we should be able to get n!

all shifts.

▷ But show that each A[i] has $\frac{1}{n}$ prob of ending up anywhere in B.

↳ A[i] can be shifted by 1 up to n, so it can end at n new positions (which is all of them) and it is with $\frac{1}{n}$ prob

(technically n-1 new and 1 old.)

Cormen 5.3-5 Perm by sorting
p. 129

$n = \text{len}(A)$
P: new array[1.. n]
for $i = 1$ to n:
    $P[i] = \text{random}(1, n^3)$
Sort A using P as sort keys

▷ Show that prob of all elems in P unique is $1 - \frac{1}{n}$.

▷ What is prob that first pick is unique?
↳ 1

▷ Prob of second being unique?
↳ $\frac{n^3 - 1 \leftarrow \text{left}}{n^3 \leftarrow \text{total}} = 1 - \frac{1}{n^3}$

▷ Third?
↳ $1 - \frac{2}{n^3}$

▷ So we get $1 \cdot \left(1 - \frac{1}{n^3}\right) \cdot \left(1 - \frac{2}{n^3}\right) \cdots \left(1 - \frac{n-1}{n^3}\right)$

$$\geq \underbrace{1 \cdot \left(1 - \frac{n}{n^3}\right) \cdots \left(1 - \frac{n}{n^3}\right)}_{n}$$ (can only get worse by using $\frac{n}{n^3}$ everywhere)

$\frac{n}{n^3} = \frac{1}{n^2}$

$$\geq \left(1 - \frac{1}{n^2}\right)^n$$

$$\geq 1 - \frac{1}{n}$$ (as $(1-x)^n \geq 1 - n \cdot x$)

Cormen 5.3-7 Random Sample
p. 129

▷ We want random sample of size m of $\{1, 2, \ldots, n\}$.
Each subset of size m should be likely. ← $\frac{1}{\binom{n}{m}}$

▷ We could: RANDOM-IN-PLACE, take first m elems.
This is n calls to random()
If m is much smaller than n, this could be expensive.

▷ Suggested alg: note $\{1, \ldots, n\}$ is implicit

```
RAND-SAMPLE(m, n):
  if m = 0:
    return ∅        } base case
  else:
    S = RAND-SAMPLE(m-1, n-1)
    i = random(1, n)
    if i ∈ S:
      S' = S ∪ {n}      ← S cannot contain n already as
    else:
      S' = S ∪ {i}
  return S'
```

▷ Understanding alg: we assume we can create a sample of size $m-1$ (each one eq likely). We pick a new candidate to be included in S (prob $\frac{1}{n}$). If it is already present, we add the only elem we know cant be present: n. Otherwise we add the candidate.

▷ Show alg creates subset with $\frac{1}{\binom{n}{m}}$ chance
Induction on m: base case $m = 0$: $\frac{1}{\binom{n}{0}} = \frac{1}{1} = 1$
There is only 1 subset of $\emptyset$, namely $\emptyset$.
Now assume we have sample of $\{1, \ldots, n-1\}$ of size $m-1$ where each sample has prob $= \frac{1}{\binom{n-1}{m-1}}$.

continued

Cormen 5.3-7 continued

First of all, let us write out $\binom{n-1}{m-1} = \frac{(n-1)!}{(m-1)!(n-1-(m-1))!}$

$= \frac{(n-1)!}{(m-1)!\,(n-m)!}$, so prob for S was $\frac{1}{\binom{n-1}{m-1}} = \frac{(m-1)!\,(n-m)!}{(n-1)!}$

▷ Now determine if n becomes part of sample S' or not.

▷ n is included if i is already in S or i was chosen to be n.

The prob of including n is $\frac{\overset{\text{not clear in } S}{m-1} + \overset{\text{the } n \text{ outside } S}{1}}{n} = \frac{m}{n}$

The total prob of S': $\frac{(m-1)! \cdot (n-m)!}{(n-1)!} \cdot \frac{m}{n} = \frac{m!\,(n-m)!}{n!}$

$= \frac{1}{\binom{n}{m}}$

▷ n is not included if i is not already in S.

prob of that is just $1 - \frac{m}{n} = \frac{n-m}{n}$

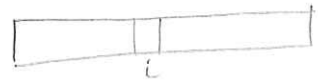
But in this case, the prob of the recursive sample is $\frac{1}{\binom{n-1}{m}} = \frac{m!\,(n-m-1)!}{(n-1)!}$ as i is in the sample already.

The total prob:

$$\frac{m!\,(n-m-1)!}{(n-1)!} \cdot \frac{n-m}{n} = \frac{m!\,(n-m)!}{n!} = \frac{1}{\binom{n}{m}}$$

Cormen 5-2 Searching unsorted array
p. 143

3. random algs

search: x , A: 

a) First alg idea: Pick rand i in $\{1, \dots, n\}$ ← $|A|$
check if $A[i] = x$, keep going until all i tried or
x is found

▷ Create pseudo code:

```
search(A, x):
    C = Ø      ← checked indices
    while |C| < n:
        i = random(1, n)
        if A[i] = x
            return i
        else
            C = C ∪ {i}
    return "Not found"
```

b) Suppose x is present one place in A.
Exp # of tries before x is found?

▷ Bernoulli! we have indep trials and $p = \frac{1}{n}$
(Geometric distribution)

$$\frac{1}{p} = \frac{1}{\frac{1}{n}} = n$$

c) Suppose x appears $k \geq 1$ times.
Exp # of tries?

$$p = \frac{k}{n} \Rightarrow \frac{1}{p} = \frac{n}{k}$$

d) Suppose x not present. Exp # of tries?
similar to coupon collector:
Move to next phase each time a new index is tested

$\Rightarrow O(n \log n)$

Continued

Cormen 5-2 continued

e) We now look at deterministic search:
go through $A[1], A[2], \dots, A[n]$ until x is found.

▷ First, suppose x appears <u>once</u> in A
(Each permutation of A is equally likely $\frac{1}{n!}$)
- What is worst-case running time?
↳ n comparisons (x is last element)

▷ What is average-case running time?

↳ Here we have to compute the average cmp over all inputs

There are $(n-1)!$ cases where it take 1 

$(n-1)!$ cases where it take 2 

⋮

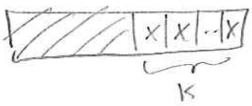
up to $(n-1)!$ cases where it takes 

Divide all these by the total number: $n!$
we get

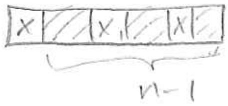
$$\frac{\sum_{i=1}^{n} i \cdot (n-1)!}{n!} = \frac{(n-1)! \sum_{i=1}^{n} i}{n} = \frac{\cancel{(n-1)!}\, n(n+1) \cdot \frac{1}{2}}{\cancel{n!}} = \frac{n+1}{2}$$

f) Suppose now that x appears k times

▷ Worst-case?

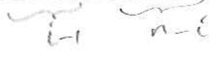
↳  , we hit first x at $n-k+1$

▷ Average-case?

\# of perm w. cost 1: 

Choose $k-1$ spots for the rest of the x's and perm all $n-k$:

$\binom{n-1}{k-1}(n-k)!$

\# of perm w cost i:  $\binom{n-i}{k-1}(n-k)!$

Total # of perm $\frac{n!}{k!}$ (all x's are identical)

continued

$$\frac{\sum i \binom{n-i}{k-1}(n-k)!}{\frac{n!}{k!}} \underset{\text{By Maple}}{=} \frac{n+1}{k+1}$$

Cormen 5-2 continued

g) suppose x not present

▷ worst case?

↳ n, we have to check all

▷ Avg case?

↳ n, all instances will cost n. $\left(\text{could write } \frac{\sum_{i=1}^{n} n}{n} = n \right)$

h) Now consider Scramble-Search:

Shuffle A then run deterministic search

▷ For x present once, k-times and none, what is exp running time?

↳ It is just the same as average for the deterministic.

i) which alg would you use?

(No universal right answer, discussion)

Peter: I would go with deterministic.
It has avg case like scramble,
but does not pay the price of shuffle.
Cache-friendly.

KT 2

p. 782

100 000 people vote in election.

2 candidates: D, R

80 000 go to poll and intend to vote D

20 000 ——— " ——— R

Ballot confusing, $\frac{1}{100}$ prob of voting opposite of intention.

Let $X = \#$ of votes for D

▷ What is $E(X)$?

Let $X_i = \begin{cases} 1 & \text{if person } i \text{ voted D} \\ 0 & \text{"} \end{cases}$, $X = \sum_{i=1}^{100000} X_i$

For $i \leq 20\,000$: $E(X_i) = \frac{99}{100} \cdot 0 + \frac{1}{100} \cdot 1 = \frac{1}{100}$

For $i > 20\,000$: $E(X_i) = \frac{1}{100} \cdot 0 + \frac{99}{100} \cdot 1 = \frac{99}{100}$

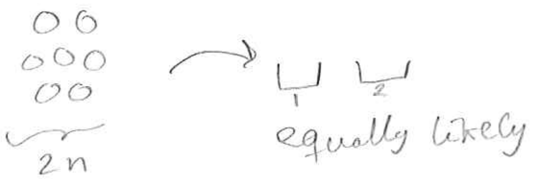
By linearity of exp:

$$E(X) = \sum_{i=1}^{20000} E(X_i) + \sum_{i=20001}^{100000} E(X_i) = 20000 \cdot \frac{1}{100} + 80000 \cdot \frac{99}{100}$$

$$= 79400$$

KT 13 Balls-and-bins experiment

p. 790



, exp # in each box is n.

Let X_1 = # balls in bin 1, $E(X_1) = n$

X_2 = # balls in bin 2, $E(X_2) = n$

▷ We want to look at how big $X_1 - X_2$ (or $X_2 - X_1$) is likely to be.

▷ Specifically prove: For $\varepsilon > 0$, $\exists c > 0 : P(X_1 - X_2 > c\sqrt{n}) \leq \varepsilon$.

Hint: Use Chernoff bounds. $P(X > (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$

First obs: $X_2 = 2n - X_1$

So $X_1 - X_2 > c\sqrt{n} \Rightarrow X_1 - (2n - X_1) > c\sqrt{n} \Rightarrow 2X_1 - 2n > c\sqrt{n}$

$\Rightarrow X_1 > \frac{c\sqrt{n}}{2} + n$

We want on the form $(1+\delta)\mu$, and we know $\mu = E(X_1) = n$

So

$$(1+\delta)n = \frac{c\sqrt{n}}{2} + n \Rightarrow \cancel{n} + \delta n = \frac{c\sqrt{n}}{2} \cancel{+n}$$

$$\Rightarrow \delta = \frac{c\sqrt{n}}{2n} = \frac{c}{2\sqrt{n}} \quad \left(\text{as } \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}\right)$$

We can now plug into Chernoff bound

$$P(X_1 - X_2 > c\sqrt{n}) = P\left(X_1 > \left(1 + \frac{c}{2\sqrt{n}}\right)\overset{E(X_1)}{n}\right) \leq$$

$$\underbrace{\left(\frac{e^{\frac{c}{2\sqrt{n}}}}{\left(1+\frac{c}{2\sqrt{n}}\right)^{\left(1+\frac{c}{2\sqrt{n}}\right)}}\right)^{n}}, \quad \text{and } e^{x} < (1+x)^{(1+x)} \quad \forall x$$

So for a fixed n, if we want } to be $\leq \varepsilon$, we can choose a large enough c.

Exercise on weekly note: Multiple choicese

n Questions

4 Answers to each, one of which is correct

Students can answer one of the 4, or leave blank.

+1 for correct answer, $-\frac{1}{3}$ for incorrect, 0 if blank

Pass: $\geq \frac{n}{2}$ score.

Def: A challenged student knows $\leq 40\%$ of answers, and guess the rest ($\frac{1}{4}$ for each choice)

Def: A test is good if the prob that a challenged student passes is $\leq 5\%$

Let $m = \frac{3}{5}n$ ← The amount a chall stud guesses.

Let $X_i = \begin{cases} 1 & \text{if } i\text{'th guess correct} \\ 0 & \text{o.w.} \end{cases}$, $X = \sum_{i=1}^{m} X_i$

a) $E(X)$?

▷ First of all, what is $E(X_i)$?

↳ $E(X_i) = P(X_i = 1) = \frac{1}{4}$

$E(X) = \underset{\text{lin of exp}}{\sum E(X_i)} = \sum_{i=1}^{m} \frac{1}{4} = \frac{1}{4}m = \frac{3}{20}n$

Note this is not the exp score of 1 Q. It would be $S_i = \begin{cases} 1 \\ -\frac{1}{3} \end{cases}$ which is not an ind. var.

$E(S_i) = \frac{3}{4}\left(-\frac{1}{3}\right) + \frac{1}{4} \cdot 1$
$= -\frac{1}{4} + \frac{1}{4} = 0$

b) Show that a chall stud only passes if $X \geq \frac{3}{2}E(X)$

continued

Exercise on weekly note: continued

what he actually answers correct

b) show chall stud only pass if $X \geq \frac{3}{2} E(X)$

i.e. show the corresponding score is $\geq \frac{n}{2}$.

▷ what could be the first thing to do?

↳ Expand $\frac{3}{2} E(X) = \frac{3}{2} \cdot \frac{3}{20} n = \frac{9}{40} n$

▷ Next?

compute score where $X \geq \frac{9}{40} n$

First: The stud gets $\frac{2}{5} n$ points for known answers.

And the stud gets $\frac{9}{40} n$ for the guesses.

But $\frac{3}{5} n - \frac{9}{40} n = \frac{24}{40} n - \frac{9}{40} n = \frac{15}{40} n = \frac{3}{8} n$ are incorrect.

The score for these: $-\frac{1}{3} \cdot \frac{3}{8} n = -\frac{1}{8} n$

Total score:

$$\frac{2}{5} n + \frac{9}{40} n - \frac{1}{8} n = \frac{16}{40} n + \frac{9}{40} n - \frac{5}{40} n = \frac{20}{40} n = \frac{1}{2} n$$

c) Use chernoff to find n where prob of passing chall stud is $\leq 0{,}05$ (5%). I.e. we want to bound $P(X > \frac{3}{2} E(X))$

If we choose $\delta = \frac{1}{2}$ (as $1 + \frac{1}{2} = \frac{3}{2}$) and $\mu = E(X) = \frac{3}{20} n$ we get

$$P\left(X > \frac{3}{2} \cdot \frac{3}{20} n\right) < \underbrace{\left(\frac{e^{\frac{1}{2}}}{\left(\frac{3}{2}\right)^{\left(\frac{3}{2}\right)}}\right)}_{\text{just a constant}}{}^{\frac{3}{20} n} \approx 0{,}897^{\frac{3}{20} n} \approx 0{,}984^{n}$$

Solve $0{,}984^n = 0{,}05$ for n gives → $n = 185$ (rounded up)

So we need 185 Qs to ensure $\leq 5\%$ prob.