Announcements

- Assignments corrected by Peter. Jørgen is correcting now and you should get them soon.
- Pizza meeting (23.), Matalogifest (26)

# of onto func	p. 544
Incl - excl	p. 541
Bernoulli	p. 445
Exp Bernoulli	p. 465
Chernoff (KT)	p. 759

Today

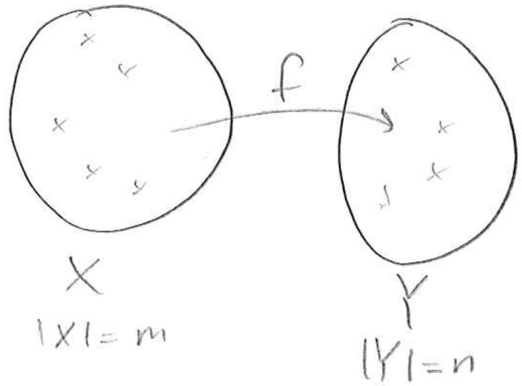
Incl - Excl

SAT, random alg

Expected log n, Harmonic number

Chernoff bounds

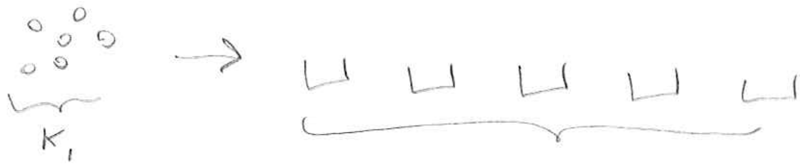
Jan 11.4 # of onto functions, p. 544; Thm 8.6.1:



$$n^m - \binom{n}{1}(n-1)^m + \binom{n}{2}(n-2)^m - \dots + (-1)^{n-1}\binom{n}{n-1}1^m$$

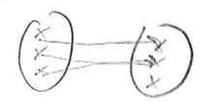
$$= \sum_{i=0}^{n-1}(-1)^i\binom{n}{i}(n-i)^m$$

a) # of ways to



So ≥ 1 box is empty?

e.g.



✓

This is the same as counting all non-onto functions

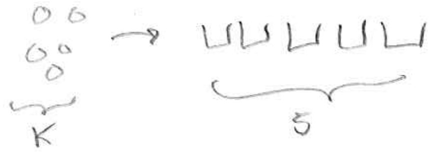
i.e. leave out the n^m term.

Set $m = k$, $n = 5 \Rightarrow \binom{5}{1}(5-1)^k - \binom{5}{2}(5-2)^k + \binom{5}{3}(5-3)^k - \binom{5}{4}(5-4)^k$

$$= 5 \cdot 4^k - 10 \cdot 3^k + 10 \cdot 2^k - 5 \cdot 1^k$$

continued

Jan 11.4 continued

b) # of ways  where exactly <u>one</u> box empty?

Choose one of the boxes to be empty. $\binom{5}{1}=5$

In the remaing 4, find # of ways where <u>no</u> box empty (Use Thm 8.6.1 directly).

$$m = k$$
$$n = 4 \Rightarrow 4^k - \binom{4}{1}(4-1)^k + \binom{4}{2}(4-2)^k - \binom{4}{3}(4-3)^k$$
$$= 4^k - 4\cdot 3^k + 6\cdot 2^k - 4\cdot 1^k$$

Result: $\binom{5}{1}(4^k - 4\cdot 3^k + 6\cdot 2^k - 4\cdot 1^k)$

# of ways with x boxes empty

Thm 8.6.1	0	1	2	3	4	5
a)	0	1	2	3	4	5
b)	0	1	2	3	4	5
c)	0	1	2	3	4	5

not possible

c) # of ways, where ≥ 2 boxes empty?

I.e. exactly 2 or 3 or 4 boxes empty.

a) was ≥ 1 box empty, ie. 1 or 2 or 3 or 4 boxes empty.

b) was $= 1$ box empty

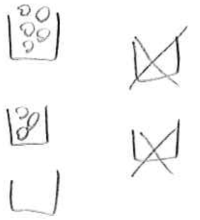
So result is just a) − b).

d) why is c) <u>not</u> $\binom{5}{2}3^k$

ie. choose 2 empty boxes and distribute in rest.

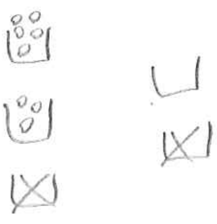
Overcounting!

Consider one particular distribution



$\boxtimes$: chosen to be empty

It is the same as



we do not distinguish between choosing the box to be empty and not putting objs in the box.

Jan 15. 3 Ternary strings

Example 001221

a) # of ternary strs of len n?

3 choices for first
3 choices for second
⋮

$\left.\begin{matrix} \\ \\ \end{matrix}\right\}$ product rule

$\Rightarrow \underbrace{3 \cdot 3 \cdot \cdots 3}_{n} = 3^n$ (just like 2^n for binary str)

b) # of ternary str of len n with $\geq 1 \times \boxed{0}$, $\geq 1 \times \boxed{1}$, $\geq 1 \times \boxed{2}$

We use incl-excl:

Let

P_0 : $< 1 \times \boxed{0}$ i.e. exactly $0 \times \boxed{0}$ i.e. no $\boxed{0}$'s

P_1 : $< 1 \times \boxed{1}$ i.e. exactly $0 \times \boxed{1}$ i.e. no $\boxed{1}$'s

P_2 : $< 1 \times \boxed{2}$ i.e. exactly $0 \times \boxed{2}$ i.e. no $\boxed{2}$'s

Our answer is $N(P_0' P_1' P_2') = \overset{\leftarrow a)}{N} - N(P_0) - N(P_1) - N(P_2)$
$+ N(P_0 P_1) + N(P_0 P_2) + N(P_1 P_2)$
$- N(P_0 P_1 P_2)$

$N = 3^n$ (from a))

$N(P_0)$ = "# of str with only $\boxed{1}$'s and $\boxed{2}$'s" $= 2^n$ (like binary str)

$N(P_1) = N(P_2) = N(P_0) = 2^n$ (same idea)

$N(P_0 P_1)$ = "# of str with only $\boxed{2}$'s" $= 1^n = 1$

$N(P_0 P_2) = N(P_1 P_2) = N(P_0 P_1) = 1$

$N(P_0 P_1 P_2)$ = "no digits" $= 0$

$N(P_0' P_1' P_2') = 3^n - 3 \cdot 2^n + 3 \cdot 1 - 0 = 3^n - 3 \cdot 2^n + 3$

KT 1 ← Kleinberg & Tardos

P. 782

3- coloring graph

Give each node in graph a color $\in \{1, 2, 3\}$
If edge $e = u, v$ has u and v with different colors then e is <u>satisfied</u>.

Give (randomized) alg to get (expected) $\frac{2}{3} c^*$ satisfied

← optimal; of all possible.

Note: Giving colors to nodes corresponds to putting nodes in 3 partitions!
This is just like part of an exercise in the assignment.

Alg: for $v \in V$: assign random color

let $X_e = \begin{cases} 1 & \text{if } e \text{ \underline{satisfied}} \\ 0 & \text{o.w.} \end{cases}$, $X = \sum_{e \in E} X_e$

$E(X_e) = P(X_e = 1)$?

let u's color be picked by Alg already and WLOG the color is 1.
If v ends up with color 1 ⇒ e <u>not</u> <u>satisfied</u> ← $\frac{1}{3}$
otherwise e is <u>sat</u>. Prob $= 1 - \frac{1}{3} = \frac{2}{3}$

$E(X_e) = \frac{2}{3}$

$E(X) = E\left(\sum X_e\right) = \sum E(X_e) = \sum \frac{2}{3} = \frac{2}{3} \sum_{e \in E} 1 = \frac{2}{3} |E| \geq \frac{2}{3} c^*$

(c^* is all edges that are possible to sat which might not be all $|E|$)

KT 7 Max-3-SAT

p. 787

Formulas like $\underbrace{(x_1 \vee \bar{x}_3 \vee x_7)}_{\text{Clause, } c_1} \wedge (\bar{x}_4 \vee x_7 \vee x_8) \wedge \cdots$ ...

Find truth assignment where the most c_i are Sat.

▷ Jørgen showed for Max-3-SAT we have rand alg $\Rightarrow \frac{7}{8} \cdot |C|$ expected.

▷ we now look at Max-SAT (no requirement that len 3 of clauses)

we have $c_1 \dots c_k$ clauses over variables $x_1 \dots x_n$

Example:

$$(x_2 \vee \bar{x}_3) \wedge (x_1) \wedge (x_7 \vee \bar{x}_8 \vee x_9 \vee \bar{x}_{10}) \wedge \cdots$$

▷ a) Use same alg, random assignment (T/F) to x_i with $\frac{1}{2}$.

Show exp clauses sat $\geq \frac{1}{2}k$

let $x_c = \begin{cases} 1 & \text{if } c \text{ is sat} \\ 0 & \text{o.w} \end{cases}$, $X = \sum_{c=1}^{k} x_c$

$E(x_c) = P(x_c = 1)$?

There is always ≥ 1 variable in clause, and that var has $\frac{1}{2}$ of being true $\Rightarrow \frac{1}{2}$ for sat'ing clause

$E(x_c) \geq \frac{1}{2}$, $E(X) \geq \sum_{c=1}^{k} \frac{1}{2} = \frac{1}{2}k$

▷ Give example where no more than half can be sat

$(x_1) \wedge (\bar{x}_1)$

b) Consider all clauses with one term (x_i).

Assume no other clause $(\bar{x}_i)$ (also single term) : $(x_i) \wedge (\bar{x}_i)$ would conflict.

Modify alg to get expected ratio from $\frac{1}{2} = 0{,}5 \to 0{,}6 = \frac{6}{10}$

continued

KT 7 continued

b) let r = # of clauses of size 1

let t = # of clauses of size ≥ 2

(note $t = k - r$)

▷ one would perhaps consider changing the alg to go through and set each size-1-clause to true to sat all of them (potentially getting r sat'ed clause for very little work).

This is a good idea but if all other clauses are like $(\bar{x}_i \vee \bar{x}_j)$ where x_i and x_j each have a size-1-clause (x_i) and (x_j), then all the rest of the clauses cannot be sat.

▷ Instead do this:

If $r \geq \frac{6}{10}k$, do as above, done. (we sat'ed what was required)

else: $r < \frac{6}{10}k$, in that case do random assignment.

Then $E(\text{\# of sat'ed clauses}) \geq \underbrace{\frac{1}{2}r}_{\text{exp, size 1}} + \underbrace{\frac{3}{4}\overbrace{(k-r)}^{t}}_{\text{exp, size} \geq 2}$

$= \frac{10}{20}r + \frac{15}{20}k - \frac{15}{20}r = \frac{15}{20}k - \frac{5}{20}r \underset{r < \frac{6}{10}k}{\geq} \frac{15}{20}k - \frac{5}{20} \cdot \frac{6}{10}k = \frac{15}{20}k - \frac{3}{20}k$

$= \frac{12}{20}k = \frac{6}{10}k$

continued

KT 7 continued

c) Now we have to solve the general SAT (conflicting clauses can occur).

Among the r size-1-clauses we have q conflicting clauses.

There is no hope for us to sat all of them, but we also only have to sat $\frac{6}{10}$ of optimal. So we can only expect to sat at most k − q clauses.

▷ New alg: Go through and remove one of the conflicting clauses for each of the q. New formula has k − q and we can expect to sat $\frac{6}{10}$ of these.

KT 10 Online bidding system
p. 789

There are n bidding agents $\boxed{A_1 | A_2 | \dots | \quad | \quad | A_n}$

Each agent i has bid $b_i > 0$

And all b_i are different.

Auction starts and each agent bids in random order.

They place their bids before hand and can't see each other's bids. $\boxed{b_1 | b_2 | \cdots | b_n}$

b^* : variable storing the highest bid so far.

It is updated each time a bid $b_i > b^*$

old value

What is the expected # of times b^* is updated?

Example: $b_1 = 20$, $b_2 = 25$, $b_3 = 10$

Bids appear b_1, b_3, b_2. b^* is updated for b_1 and b_3 not b_2.

▷ For first bid, what is chance of being better than all previous? (i.e. updating b^*)

↳ 1

▷ For second bid, what is prob?

↳ $\frac{1}{2}$, as it can either be > or < the first bid

▷ Third bid?

↳ $\frac{1}{3}$, as of the 3 first numbers, largest comes last

▷ In general: Bid i

↳ $\frac{1}{i}$, Focus on first i bids

Prob that largest is placed last

▷ Let $X_i = \begin{cases} 1 & \text{if bid } i \text{ updates } b^* \\ 0 & \text{o.w} \end{cases}$ and $X = \sum X_i$ # of times b^* updates

$E(X_i) = \frac{1}{i}$

▷ $E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{1}{i} = H_n \approx \ln n \in O(\log n)$

Harmonic number

Jan 10.3 Multiple choice

2 choices per Question. 10 questions

conditions:

- All Qs must be answered
- wrong answers count towards you.

n_1 is right As, n_2 is wrong As

$n_1 - n_2$ is score.

For 10 Qs and n_1 right, score is $n_1 - (10 - n_1)$

$= 2n_1 - 10$

▷ 3 students

A: $\frac{3}{5}$ for correct, X_A = Score of A

B: $\frac{4}{5}$ for correct, X_B = Score of B

C: $\frac{1}{2}$ for correct, X_C = Score of C

a) $E(X_A)$? we expect $10 \cdot \frac{3}{5}$ to be correct and $10 \cdot \frac{2}{5}$ to be wrong.

$\Rightarrow 10 \cdot \frac{3}{5} - 10 \frac{2}{5} = 2$

$E(X_B) = 10 \cdot \frac{4}{5} - 10 \cdot \frac{1}{5} = 6$

$E(X_C) = 10 \frac{1}{2} - 10 \cdot \frac{1}{2} = 0$

b) $P(X_C > E(X_C))$? i.e. $P(X_C > 0)$? As it is completely symmetrical around 0, $P(X_C > 0) = P(X_C < 0)$, so $P(X_C \neq 0) = 2 \cdot P(X_C > 0)$

$\Rightarrow P(X_C = 0) = 1 - 2 \cdot P(X_C > 0) \Rightarrow P(X_C > 0) = \frac{1 - P(X_C = 0)}{2}$

Now we focus on $P(X_C = 0)$, i.e. prob that $n_1 = n_2$

C outputs a random bit str,

And all $\binom{10}{5}$ ways to choose bitstr with 5 ones (and 5 zeros)

$\Rightarrow$ prob $= \frac{\binom{10}{5}}{2^{10}} = \frac{63}{256}$; Result: $\frac{1 - \frac{63}{256}}{2} = \frac{193}{512}$. continued

Jan 10.3 continued

c) Bar for passing is now 2 points

$p(A \text{ pass})$? $p(B \text{ pass})$? $p(C \text{ pass})$?

Note: Getting 2 points means $n_1 = 6$, $n_2 = 4$

Also: Not possible to get 1 point.

We already know $p(C \text{ pass})$ as it is $p(X_c > 0) = p(X_c \geq 2) = \frac{193}{512}$.

We can use Bernoulli for the rest

[Prob of exactly k success in n trial, where prob of success is p and failure is q $(p+q=1)$ is $\binom{n}{k} p^k q^{n-k}$]

we want 6, 7, 8, 9 or 10 success/rights:

A: $\sum_{k=6}^{10} \binom{10}{k} \left(\frac{3}{5}\right)^k \left(\frac{2}{5}\right)^{10-k} = \frac{6182649}{9765625} \approx 0,633$

B: $\sum_{k=6}^{10} \binom{10}{k} \left(\frac{4}{5}\right)^k \left(\frac{1}{5}\right)^{10-k} = \frac{9445376}{9765625} \approx 0,967$

d) We now have 8 students incl C that are guessing

what is the expected # that pass?

Each is just indep trial $\Rightarrow$ use exp Bernoulli

Chance of success $p = \frac{193}{512}$, Result: $8 \cdot \frac{193}{512} \approx 3,016$

e) Recap: Chernoff bounds

If $X = X_1 + X_2 + \dots + X_n$, and all X_i are indep

we can bound:

$$p(X > (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu}$$

where $\mu \geq E(X)$ (think of it as just $E(X)$)

and $\delta > 0$ is just some chosen number.

continued

Jan 10.3 continued

e) we now consider 3 different tests using same system:

- 100 Qs
- 200 Qs
- 500 Qs

Pass: 20% score

We now want to bound the prob that C passes each.

Let $Y_i = \begin{cases} 1 & \text{if C answers correctly on } Q_i \\ 0 & \text{o.w.} \end{cases}$, $Y = \sum_{i=1}^{n} Y_i$

$E(Y_i) = P(Y_i = 1) = \frac{1}{2}$; $E(Y) = \frac{1}{2} n$

We can now use Chernoff as Y_i indep

$\delta = 0{,}2 \qquad \mu = \frac{1}{2} n$

$$P(Y > \underbrace{(1+0{,}2)\tfrac{1}{2}n}_{\text{20\% more than expected (0 points)}}) < \left(\frac{e^{0{,}2}}{(1+0{,}2)^{(1+0{,}2)}} \right)^{\frac{1}{2}n} \approx 0{,}9814^{\frac{1}{2}n}$$

$n = 100$: $\left(\frac{e^{0{,}2}}{1{,}2^{1{,}2}} \right)^{50} \approx 0{,}3909$

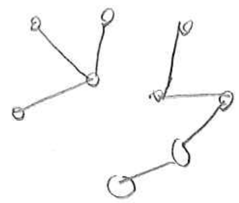
$n = 200$: $\left(\frac{e^{0{,}2}}{1{,}2^{1{,}2}} \right)^{100} \approx 0{,}1528$

$n = 500$: $\left(\frac{e^{0{,}2}}{1{,}2^{1{,}2}} \right)^{250} \approx 0{,}0091$

As expected, the chance of passing decreases as the number of Qs increase.
It gets harder to get away from $E(Y)$!

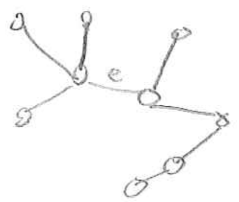
Jan. 10.5 Creating connected graph from connected components.

Example:



This is not a connected graph (you can't get from anywhere to anywhere) but it has 2 connected components.

↓



If we add e, it will join the two components into one. The graph is now connected.

We will look at a randomized alg A to create a connected graph from n nodes with no initial edges.

The alg picks edges from K_n add adds them to G.
↑ complete graph

Note: K_n has $\binom{n}{2}$ edges.

First we have to show:

a) $$\binom{n+1}{2} > \binom{n_1+1}{2} + \binom{n_2+1}{2} + \cdots + \binom{n_k+1}{2}$$

where $\sum_{i=1}^{n} n_i = n$ and $k \geq 2$ (and $k \leq n$)

Note: $$\binom{n+1}{2} = \binom{n}{2-1} + \binom{n}{2} = \binom{n}{1} + \binom{n}{2} = n + \binom{n}{2}$$
↑ Pascal

and $$\binom{n}{2} > \binom{n_1}{2} + \binom{n_2}{2} + \cdots + \binom{n_k}{2}$$

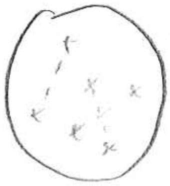
This follows combinatorially from counting ways to pick pairs from sets. See next page

Continued

Jan 10. 5 continued

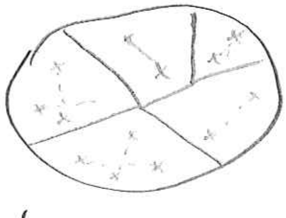
a) $\binom{n}{2} > \binom{n_1}{2} + \binom{n_2}{2} + \cdots + \binom{n_k}{2}$

Consider ~~count~~ all different pairs of a set S



There are $\binom{n}{2}$ ways to do this

Now divide S ~~up~~ into $S_1, S_2, \dots, S_k$ and picking pairs from each individually



Observe that in that case all pairs across S_i and S_j are not counted thus we must have strictly less pairs here.

The formula we should show now follows from these two observations:

$$\binom{n+1}{2} > \binom{n_1+1}{2} + \binom{n_2+1}{2} + \cdots + \binom{n_k+1}{2}$$

$\Downarrow$

$$n + \binom{n}{2} > n_1 + \binom{n_1}{2} + n_2 + \binom{n_2}{2} + \cdots + n_k \binom{n_k}{2}$$

$\Downarrow$

$$\binom{n}{2} + \cancel{n} > \binom{n_1}{2} + \binom{n_2}{2} + \cdots \binom{n_k}{2} + \underbrace{\cancel{\sum_{i=1}^{n} n_i}}_{n} .$$

b) ▷ Why does the alg need $\geq n-1$ iterations?

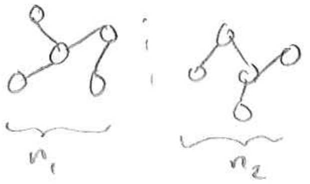
We have n components to start with: 

Every time we can only join at least 2 components.

So we need $n-1$ iterations to join all n.

▷ Give an example where the alg need $\Theta(n^2)$ different edges before terminating.

Imagine alg ends up creating 2 components



But instead of picking an edge that connects ~~them~~, alg keeps adding edges in each component of size n_1 and n_2

This gives $\underbrace{\binom{n_1}{2} + \binom{n_2}{2}}_{\text{all pairs}} \in \Theta(n^2)$; continued

Jan 10. 5 continued

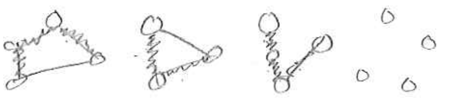
c) we want to analyse # of edges picked by A before it stops

▷ For this we see the alg as going through phases (think coupon collector).
Every time we join 2 components we go to new phase. So we start in phase 0 and end at phase $n-1$

▷ A <u>good</u> edge is one which has one endpoint in one component and another in another component. (so this edge is not present in the graph A builds but it is present in K_n where we pick from.)

▷ An <u>important</u> edge is the good edge picked causing A to go to next phase.
we start with 0 <u>imp</u> edges and end with $n-1$.

▷ Next we do some grouping:
consider the graph in phase i: imp edges are ≀



(8 imp edges so $i = 8$)

with more than 1 node

let k = # of components and let c_j = # of imp edges in component j. we have $c_1 \dots c_k$.
In the drawing $k = 3$, and $c_1 = 3$, $c_2 = 2$, $c_3 = 3$.
Of course $i = \sum_{j=1}^{k} c_j$.

phase

continued

Jan 10. S continued

c) we now have to show that

$$p(\text{picking } \underline{\text{good}} \text{ edge in phase } i) \geq \frac{\binom{n}{2} - \binom{i+1}{2}}{\binom{n}{2}} = 1 - \frac{\binom{i+1}{2}}{\binom{n}{2}}$$

▷ The number of edges in component$_j$ is at most $\binom{c_j+1}{2}$ as the inp edges span tree in comp$_j$ and the # of nodes is c_j+1

▷ This holds for all components, so total # of edges is

$$\binom{c_1+1}{2} + \binom{c_2+1}{2} + \cdots + \binom{c_k+1}{2} \underset{\text{by a)}}{<} \binom{i+1}{2}$$

▷ # of good edges: $\binom{n}{2} - \binom{i+1}{2}$ ← all edges that cross comps.

So the prob is $\dfrac{\binom{n}{2} - \binom{i+1}{2}}{\binom{n}{2}}$

d) Let X_i = # of picked edges in phase i

We have $X_0 \dots X_{n-2}$ (not X_{n-1} as we stop as we enter phase $n-1$)

Let $X = \sum_{i=1}^{n-2} X_i$, Note: $E(X)$ is our expected running time.

▷ First show $E(X_i) \leq \dfrac{\binom{n}{2}}{\binom{n}{2} - \binom{i+1}{2}}$

Here we use geometric distribution!

(Recall we can use this if we have something with $\underbrace{(1-p)^k}_{\text{fails}} \underbrace{p}_{\text{success!}}$

And the prob does not change in entire phase i.

$$E(X_i) = \frac{1}{p} \leq \frac{\binom{n}{2}}{\binom{n}{2} - \binom{i+1}{2}}$$

continued

Jon 10.5 continued

e) Now we compute $E(X)$ by linearity of exp:

$$E(X) = E\left(\sum X_i\right) = \sum E(X_i) \leq \sum_{i=0}^{n-2} \frac{\binom{n}{2}}{\binom{n}{2} - \binom{i+1}{2}} = \binom{n}{2} \sum_{i=0}^{n-2} \frac{1}{\binom{n}{2} - \binom{i+1}{2}}$$

f) Next we show that

$$\left(\frac{n-1}{2}\right)\left(\frac{1}{\binom{n}{2} - \binom{i+1}{2}}\right) \leq \frac{1}{n-i}$$

$$\left(\frac{n-1}{2}\right)\left(\frac{1}{\binom{n}{2} - \binom{i+1}{2}}\right) = \frac{n-1}{2 \cdot \left(\frac{n \cdot (n-1)}{2} - \frac{i \cdot (i+1)}{2}\right)} = \frac{n-1}{n \cdot (n-1) - i \cdot (i+1)}$$

$$\leq \frac{n-1}{n \cdot (n-1) - i(n-1)} = \frac{1}{n-i}$$

as $i \leq n-2 \Rightarrow i+1 \leq n-1$; replacing $i+1$ with $n-1$ only makes it bigger.

g) we can now work further on $E(X)$

$$E(X) \leq \binom{n}{2} \sum_{i=0}^{n-2} \frac{1}{\binom{n}{2} - \binom{i+1}{2}} = n \cdot \frac{n-1}{2} \sum \cdots$$

$$= n \cdot \sum_{i=0}^{n-2} \left(\frac{n-1}{2}\right) \cdot \left(\frac{1}{\binom{n}{2} - \binom{i+1}{2}}\right)$$

$$\leq n \cdot \sum_{i=0}^{n-2} \frac{1}{n-i} = n \sum_{i=2}^{n} \frac{1}{i} \approx n \cdot H_n \approx n \cdot \ln n$$

$$\in O_E(n \log n)$$