Problem Set 9

2n个 number, n个 disjoint pigeon hole, n+1 pigeons. so must has 2 pigeons on same ....

Problem 1. [**10 points**]

① 证明存在相邻 ② gcd(k, k+1) = 1.

(a) [5 pts] Show that of any $n+1$ distinct numbers chosen from the set $\{1, 2, \ldots, 2n\}$, at least 2 must be relatively prime. (*Hint:* $\gcd(k, k+1) = 1$.)

(b) [5 pts] Show that any finite connected undirected graph with $n \geq 2$ vertices must have 2 vertices with the same degree.

max degree = (n−1).

n ≥ 0, n个鸽子, n−1个窝.

so ⇒ must ...

Problem 2. [**10 points**] **Under Siege!**

Fearing retribution for the many long hours his students spent completing problem sets, Prof. Leighton decides to convert his office into a reinforced bunker. His only remaining task is to set the 10-digit numeric password on his door. Knowing the students are a clever bunch, he is not going to pick any passwords containing the forbidden consecutive sequences "18062", "6042" or "35876" (his MIT extension).

$10! - 6! - 7! - 6! + 3!$

How many 10-digit passwords can he pick that don't contain forbidden sequences if each number $0, 1, \ldots, 9$ can only be chosen once (i.e. without replacement)?

Problem 3. [**50 points**] Be sure to show your work to receive full credit. In this problem we assume a standard card deck of 52 cards.

$\binom{13}{1} \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot 4^3$

(a) [4 pts] How many 5-card hands have a single pair and no 3-of-a-kind or 4-of-a-kind?

(b) [4 pts] For fixed positive integers n and k, how many nonnegative integer solutions $x_0, x_1, \ldots, x_k$ are there to the following equation?

$$\sum_{i=0}^{k} x_i = n$$

$\binom{n+k-1}{k}$

(c) [4 pts] For fixed positive integers n and k, how many nonnegative integer solutions $x_0, x_1, \ldots, x_k$ are there to the following equation?

$$\sum_{i=0}^{k} x_i \leq n$$

$\sum_{i=0}^{n} \binom{i+k-1}{k}$

$\Updownarrow$

$\sum^{k+1} x_i = n$

$i=0$

$= \binom{n+k}{k+1}$

(E, X) $\frac{n!}{2!}$

(d) [4 pts] How many simple undirected graphs are there with n vertices? $\Longrightarrow 2^{\binom{n}{2}}$.

(e) [4 pts] How many directed graphs are there with n vertices (self loops allowed)? $2^{(n^2)}$ 一点有n个方向. $n \times n$.

(f) [4 pts] How many tournament graphs are there with n vertices? $2^{\frac{(n-1)\cdot n}{2}}$ $2^{\binom{n}{2}}$ 有 $\binom{n}{2}$ 场比赛, 然后每场两种结果 $2^{\binom{n}{2}}$.

(g) [4 pts] How many acyclic tournament graphs are there with n vertices? $n!$

acyclic, 意味 $x \to y$, $y \to z$, $z \not\to x$, $\therefore x \to y$ 是可传递的,

所有一个顺序 $x_0 \to x_1 \to \cdots x_n$, n 的全排列

(h) [4 pts] How many numbers are there that are in the range [1..700] which are divisible by 2, 5 or 7?

$350 + 140 + 100 - 70 - 20 - 50 + 10$

(i) [9 pts] In how many ways can you arrange n books on k bookshelf (assuming the order of books on a shelf matters?) $(n+k-1)!$

(j) [9 pts] How about if there has to be at least 1 book at each bookshelf? $(n-k-1)!$

Problem 4. [15 points] Give a combinatorial proof of the following theorem:

$$n2^{n-1} = \sum_{k=1}^{n} k\binom{n}{k}$$

(*Hint: Consider the set of all length-n sequences of 0's, 1's and a single *.*)

Problem 5. [15 points] At a congressional hearing, there are $2n$ members present. Exactly n of them are Democrats and n of them are Republicans. The members want to select a smaller subcommittee of size n from within those present at the hearing. However, since the Democrats currently hold majority, they want there to be more Democrats then Republicans in the committee. In how many ways can you select such a committee? (*Hint: Consider two cases: n odd and n even.*)

$< n.$

n为奇数时. $\binom{2n}{n} \cdot \frac{1}{2}$. 选出n个人, > 才 < 各 $\frac{1}{2}$.

n为偶数. $\left[\binom{2n}{n} - \binom{n}{\frac{n}{2}} \cdot \binom{n}{\frac{n}{2}}\right] \times \frac{1}{2}$.

6.042J / 18.062J Mathematics for Computer Science
Fall 2010

6.042/18.062J Mathematics for Computer Science — November 5, 2010

Tom Leighton and Marten van Dijk

Problems for Recitation 16

1 Combinatorial Proof

A ***combinatorial proof*** is an argument that establishes an algebraic fact by relying on counting principles. Many such proofs follow the same basic outline:

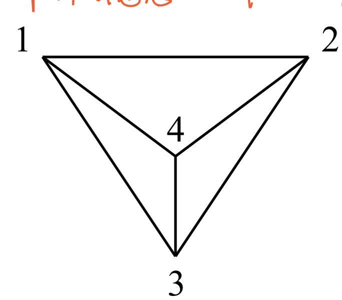
1. Define a set S.
2. Show that $|S| = n$ by counting one way.
3. Show that $|S| = m$ by counting another way.
4. Conclude that $n = m$.

2 Triangles

Let $T = \{X_1, \ldots, X_t\}$ be a set whose elements X_i are themselves sets such that each X_i has size 3 and is $\subseteq \{1, 2, \ldots, n\}$. We call the elements of T "triangles". Suppose that for all "edges" $E \subseteq \{1, 2, \ldots, n\}$ with $|E| = 2$ there are exactly λ triangles $X \in T$ with $E \subseteq X$.

For example, if we might have the triangles depicted in the following diagram, which has $\lambda = 2$, $n = 4$, and $t = 4$:

什么意思. $\{1, 3\}$. $|E| = 2$.



In this example, each edge appears in exactly two of the following triangles:

$$\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}$$

|S|?

三角形的边

Prove

$$\lambda \cdot \frac{n(n-1)}{2} = 3t$$

by counting the set

$$C = \{(E, X) : X \in T, E \subseteq X, |E| = 2\}$$

in two different ways.

Key is 理解这个set的含义, E, X 共同构成了集合.

同一个边, 在不同的三角形中是不同的, 同一个三角形中, 不同的边是不同的。所以两种不同的方法是.

E有 $\binom{n}{2}$ 种, 每一种中有λ个三角形.

1. 计算所有边所属的三角形个数.
2. 计算所有三角形边的个数.

又.

∴ 三角形个数为 $\dfrac{\lambda \cdot \frac{n(n-1)}{2}}{3}$.

3 Counting, counting, counting

Learning to count takes practice! Briefly justify your answers to the following questions. Not every problem can be solved with a cute formula; you may have to fall back on case analysis, explicit enumeration, or ad hoc methods. Do as many problems as you can and save the rest to study for Quiz II. You may leave factorials and binomial coefficients in your answers.

1. How many different arrangements are there of the letters in $BANANA$?

$$\frac{6!}{2! \times 3!}$$

2. How many different paths are there from point $(0, 0, 0)$ to point $(10, 20, 30)$ if every step increments one coordinate and leaves the other two unchanged?

$$\frac{60!}{10! \times 20! \times 30!}$$

3. Find the number of 5-card hands with exactly three aces.

$$\binom{4}{1} \cdot \binom{48}{2}$$

4. Find the number of 5-card hands in which every suit appears at most twice.

$$\frac{\binom{52}{5} - \binom{4}{1}\cdot\binom{13}{3}\cdot\binom{39}{2} - \binom{4}{1}\cdot\binom{13}{4}\cdot\binom{39}{1}}{\binom{4}{1}\cdot\binom{13}{2}\cdot(13)^3 + \binom{4}{2}\cdot\binom{13}{2}^2\cdot 13\times 2}$$

5. There are 15 sidewalk squares in a row. Suppose that a ball is thrown down the row so that it bounces on 0, 1, 2, or 3 distinct sidewalk squares. How many different throws are possible? Two throws are considered to be equivalent if they bounce on the same squares in a different order.

没看懂题...

$f(n) = f(n-1) + f(n-2) + f(n-3)$

$r^3 - r^2 - r - 1 = 0.$

6. In how many different ways can the numbers shown on a red die, a green die, and a blue die total up to 15? Assume that these are ordinary, 6-sided dice.

$5+5+5. \Rightarrow 1$

$4+5+6 \Rightarrow \binom{3}{1}\cdot\binom{2}{1} = 10$

$6+6+3 \Rightarrow \binom{3}{1}$

7. In how many ways can 20 indistinguishable pre-frosh be stored in four different crates if each crate must contain an *even* number of pre-frosh?

$\binom{13}{3}$

8. How many paths are there from point $(0,0)$ to $(50,50)$ if every step increments one coordinate and leaves the other unchanged and there are impassable boulders sitting at points $(10,10)$ and $(20,20)$?

$$\frac{\frac{100!}{50!\times 50!} - \frac{20!}{10!\cdot 10!}\times\frac{80!}{40!\cdot 40!} - \frac{40!}{20!\cdot 20!}\times\frac{60!}{30!\times 30!} + \left(\frac{20!}{10!\cdot 10!}\right)^2\cdot\frac{60!}{30!\cdot 30!}}{\binom{100}{50} - \binom{20}{10}\cdot\binom{80}{40} - \binom{40}{20}\cdot\binom{60}{30} + \binom{20}{10}\binom{20}{10}\binom{60}{30}}$$

9. In how many ways can the 180 students in 6.042 be divided into 36 groups of 5?

$$\frac{180!}{(5!)^{36}\cdot 36!}$$

10. In how many different ways can 10 indistinguishable balls be placed in four distinguishable boxes, such that every box gets 1, 2, 3, or 4 balls?

$1 \le b_i \le 4$ 6. 10 − 4. 先分配一人一个.

$\binom{4}{1}$ 3, 3 0 0. 3 1 1 1 $\binom{4}{1}$

4! 3 2 1 0

2 2 2 0 (4)

2 2 1 1 $\binom{4}{2}$

11. In how many different ways can Blockbuster arrange 64 copies of *Cat in the Hat*, 96 copies of *Matrix Revolutions*, and 1 copy of *Amelie* on 5 shelves?

$\binom{5}{1}\cdot\binom{68}{4}\times\binom{100}{4}$ ✓

$\frac{(96+64+1+4)!}{96!\cdot 64!\cdot 4!\cdot 1!}$

$10\times9\times8\times7\times6\times5$, 20.

$\frac{10!}{4!\times3!\times2!}$ 6 × 8 · 1, 3, 4, 3

$3\times\binom{5}{2}\times\binom{6}{2}$

$3\times10\times15$

1, 3

3.

$\frac{6!}{2!\cdot3!}$ 60

$3\times\binom{6}{2}$

4 There's more than one way...

In the beginning of today's recitation, we gave a combinatorial proof of the following theorem:

Theorem.

$$\sum_{i=0}^{n}\binom{k+i}{k}=\binom{k+n+1}{k+1}$$

We can also prove this theorem using induction. Give such a proof.

$n=0$ ✓.

$P(n)$: $\sum_{i=0}^{n}\binom{k+i}{k}=\binom{k+n+1}{k+1}$

Strong induction, for $n\ge0$, … $n=1$, $n=2$, 有 $P(n)$.

证 $P(n+1)=\sum_{i=0}^{n+1}\binom{k+i}{k}=\sum_{i=0}^{n}\binom{k+i}{k}+\binom{k+(n+1)}{k}$

$=\binom{k+n+1}{k+1}+\binom{k+n+1}{k}$

$=\frac{(k+n+1)!}{(k+1)!\,(n)!}+\binom{k+n+1}{k}$

$\frac{(k+n+1)!}{(k+1)!\cdot n!}+\frac{(k+n+1)!}{k!\cdot(n+1)!}$

$\frac{(n+1)\,(k+n+1)!+(k+1)\,(k+n+1)!}{(k+1)!\cdot(n+1)!}$

$=(k+n+2)!$

$$= \frac{(k+n+2)!}{(k+1)! \cdot (n+1)!} \quad \checkmark \quad = \binom{k+n+2}{k+1} \quad \frac{(k+(n+2))!}{((k+1)!\,(n+1)!)}$$

Problems for Recitation 17

The Four-Step Method

This is a good approach to questions of the form, "What is the probability that ——?" Intuition *will* mislead you, but this formal approach gives the right answer every time.

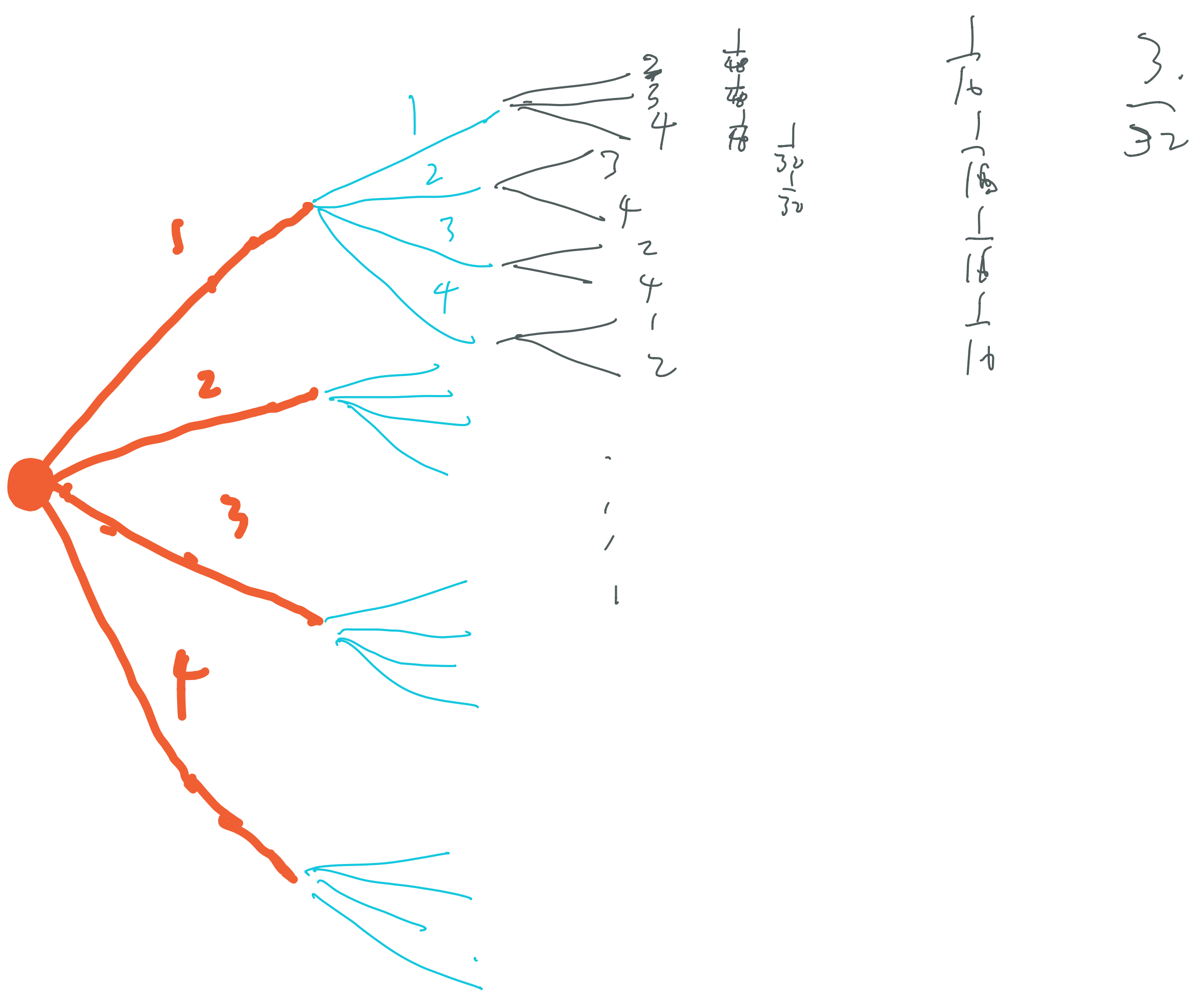
1. Find the sample space. (Use a tree diagram.)
2. Define events of interest. (Mark leaves corresponding to these events.)
3. Determine outcome probabilities:
   (a) Assign edge probabilities.
   (b) Compute outcome probabilities. (Multiply along root-to-leaf paths.)
4. Compute event probabilities. (Sum the probabilities of all outcomes in the event.)

1 The Four-Door Deal

Suppose that *Let's Make a Deal* is played according to different rules. Now there are four doors, with a prize hidden behind one of them. The contestant is allowed to pick a door. The host must then reveal a different door that has no prize behind it. The contestant is allowed to stay with his or her original door or to pick one of the other two that are still closed. If the contestant chooses the door concealing the prize in this second stage, then he or she wins.

1. Contestant Stu, a sanitation engineer from Trenton, New Jersey, stays with his original door. What is the probability that he wins the prize? 1/4.

   The tree diagram is awkwardly large. This often happens; in fact, sometimes you'll encounter *infinite* tree diagrams! Try to draw enough of the diagram so that you understand the structure of the remainder.



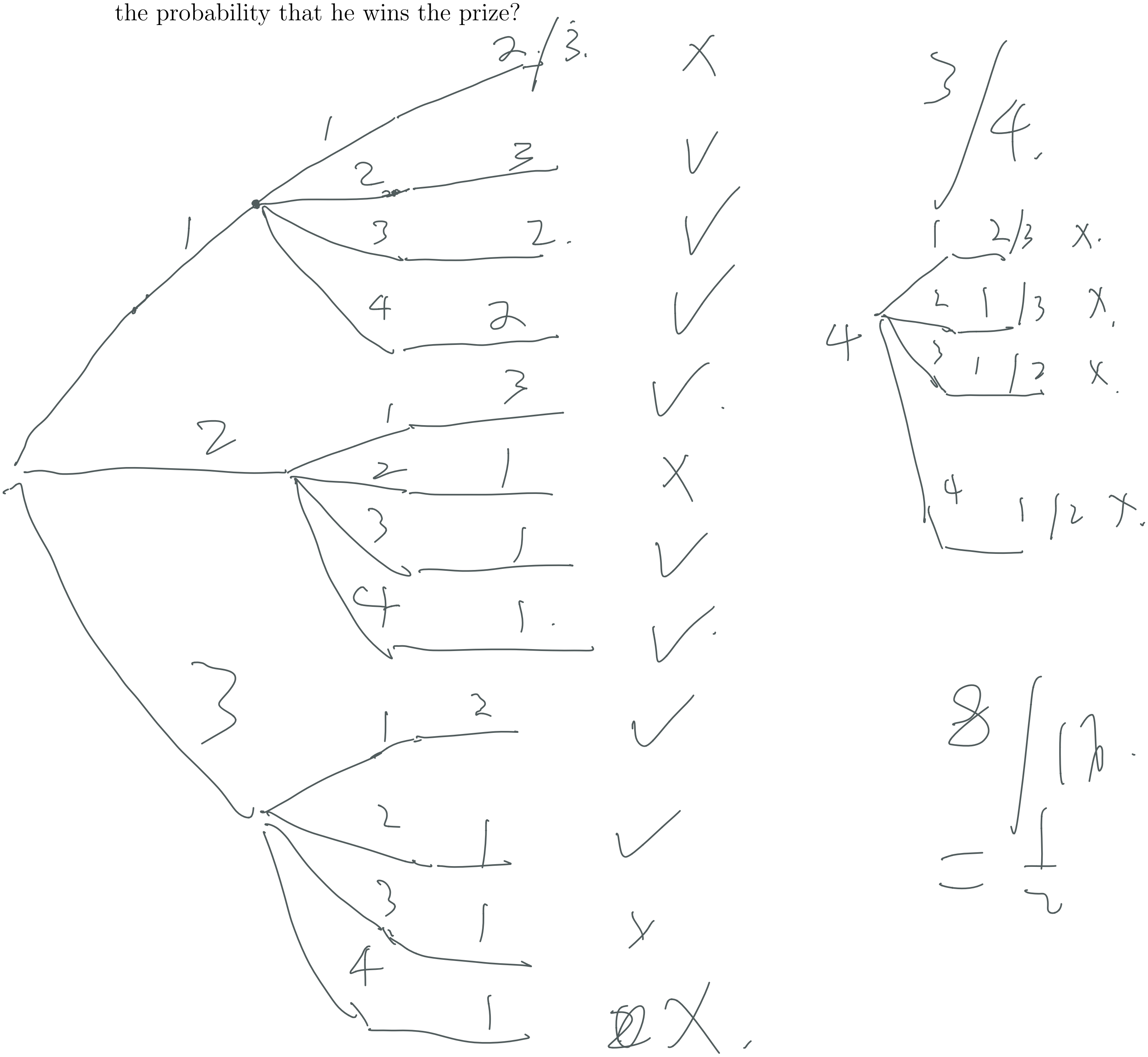
2. Contestant Zelda, an alien abduction researcher from Helena, Montana, switches to one of the remaining two doors with equal probability. What is the probability that she wins the prize?

$3/8$.

2 Earliest Door

Let's consider another variation of the four-doors problem. Say the doors are labeled A, B, C, and D. Suppose that Carol always opens the *earliest* door possible (the door whose label is earliest in the alphabet) with the restriction that she can neither reveal the prize nor open the door that the player picked.

This gives contestant Mergatroid— an engineering student from Cambridge, MA— just a little more information about the location of the prize. Suppose that Mergatroid always switches to the earliest door, excluding his initial pick and the one Carol opened. What is the probability that he wins the prize?

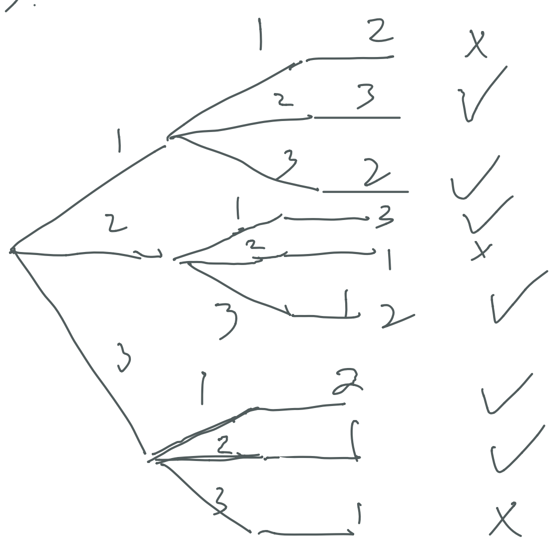


3 The 3 doors version revisited

3.1 Carol picks the smallest door

Suppose we are in the original game show with 3 doors. In our original analysis we assumed Carol picked the door randomly. In this case suppose Carol picks the smallest door, while still making sure of both i) it contains a goat and ii) it is not the contestants first choice. The contestant follows the switching strategy. What is the probability the contestant wins?

2/3.



$\frac{1}{9}$

2/3.

3.2 Carol picks the smallest door with probability p

This time, when Carol has a choice she chooses the smallest possible door with probability p and the other remaining door with probability $1 - p$. The contestant still follows the switching strategy. What is the probability the contestant wins, in terms of p?

2/3.

3.3 Optimal strategery

So far we assumed the contestant always switches. We also know from lecture another strategy: the contestant always sticks to her original choice. We determined that the probability of winning with the "always stay" strategy is simple to calculate from the probability of winning with the "always switch" strategy, and that switching was better.

What if the contestant decides whether to switch or not on a case by case basis? That is, suppose the contestant makes a decision of whether to switch or to stay based on 1) Her original choice, and 2) Carol's choice of door. Suppose the doors are labelled A, B and C. Show "always switching" is optimal. (Hint: a strategy can be seen as a mapping that assigns a pair (D_1, D_2) of observations to a decision: switch to D_3 or stay in D_1. The strategy needs to be defined for all pairs (A, B), $(A, C) \cdots$. You can optimize the reaction for each observation individually.)

6.042J / 18.062J Mathematics for Computer Science
Fall 2010

Problems for Recitation 18

1 Nerditosis

There is a rare and deadly disease called *Nerditosis* which afflicts about 1 person in 1000. One symptom is a compulsion to refer to everything— fields of study, classes, buildings, etc.— using numbers. It's horrible. As victims enter their final, downward spiral, they're awarded a degree from MIT. Two doctors claim that they can diagnose Nerditosis.

1. Doctor X received his degree from Harvard Medical School. He practices at Massachusetts General Hospital and has access to the latest scanners, lab tests, and research. Suppose you ask Doctor X whether you have the disease.

   - If you have Nerditosis, he says "yes" with probability 0.99.
   - If you don't have it, he says "no" with probability 0.97.

   Let D be the event that you have the disease, and let E be the event that the diagnosis is erroneous. Use the Total Probability Law to compute $\Pr\{E\}$, the probability that Doctor X makes a mistake. A: event say yes. B: say no.

   $$\Pr(E) = \Pr(B \cap D) + \Pr(A \cap \neg D)$$
   $$= \underline{0.01 \times \frac{1}{1000}} + \frac{999}{1000} \times \cancel{0.97}\ 0.03 \approx 3\%$$

2. "Doctor" Y received his genuine degree from a fully-accredited university for \$49.95 via a special internet offer. He knows that Nerditosis strikes 1 person in 1000, but is a little shaky on how to interpret this. So if you ask him whether you have the disease, he'll helpfully say "yes" with probability 1 in 1000 regardless of whether you actually do or not.

   Let D be the event that you have the disease, and let F be the event that the diagnosis is faulty. Use the Total Probability Law to compute $\Pr\{F\}$, the probability that Doctor Y made a mistake.

   $$\Pr(F) = \Pr(D \cap \neg Yes) + \Pr(\neg D \cap Yes)$$
   $$= \frac{1}{1000} \cdot \frac{999}{1000} + \frac{999}{1000} \cdot \frac{1}{1000} \approx \frac{1}{500} \approx 0.2\%$$

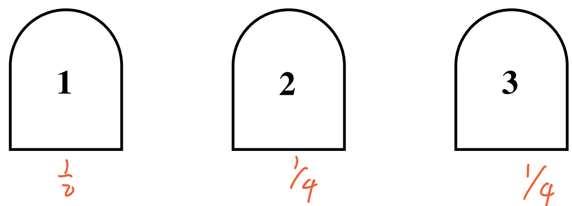
3. Which doctor is more reliable?

   ~~Yes as. X more reliable.~~ ✓

   No as Y more reliable.

2 Barglesnort

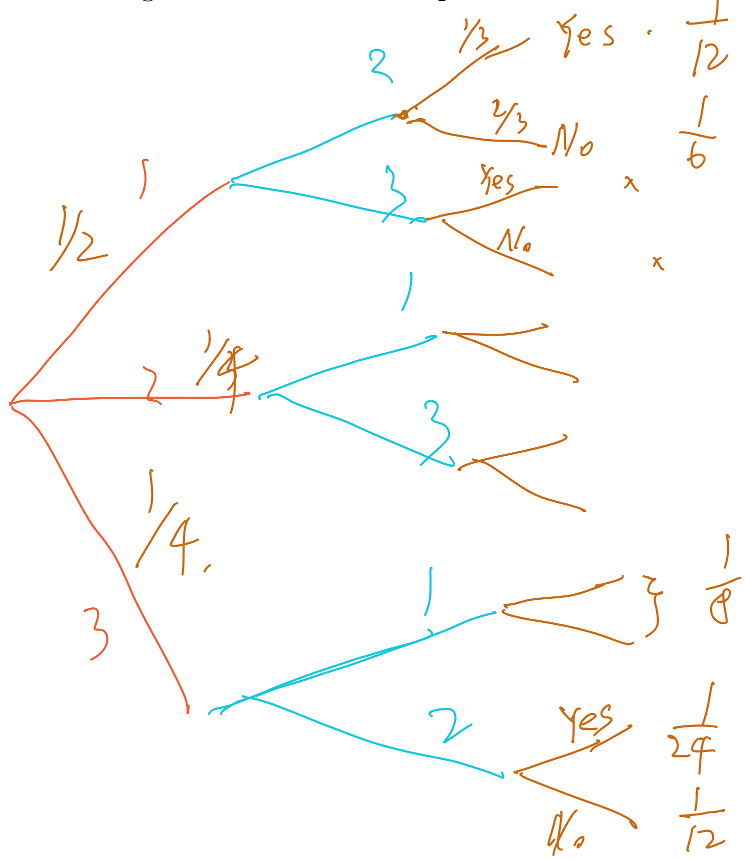
A Barglesnort makes its lair in one of three caves:



1/3.

The Barglesnort inhabits cave 1 with probability $\frac{1}{2}$, cave 2 with probability $\frac{1}{4}$, and cave 3 with probability $\frac{1}{4}$. A rabbit subsequently moves into one of the two unoccupied caves, selected with equal probability. With probability $\frac{1}{3}$, the rabbit leaves tracks at the entrance to its cave. (Barglesnorts are much too clever to leave tracks.) What is the probability that the Barglesnort lives in cave 3, given that there are no tracks in front of cave 2?

Use a tree diagram and the four-step method.



$$Pr(In_3 \mid Without\ 2) = \frac{\frac{1}{8} + \frac{1}{12}}{\left(1 - \frac{1}{12} - \frac{1}{24}\right)} \quad \frac{5}{24}$$

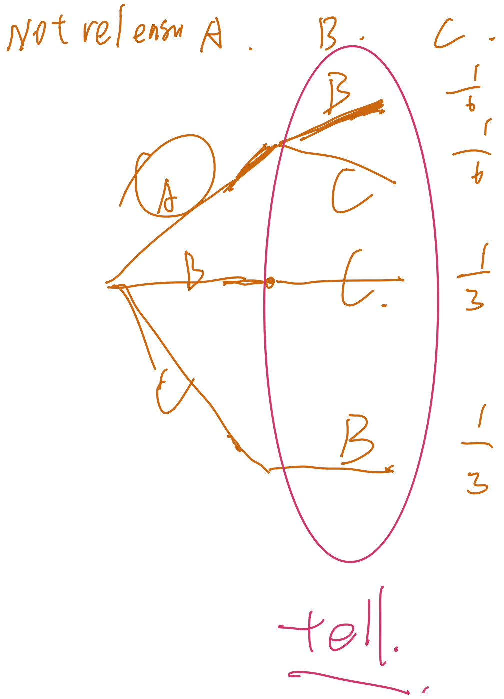
$$= \frac{5}{21}.$$

3 Prisoners

There are three prisoners in a maximum-security prison for fictional villains: the Evil Wizard Voldemort, the Dark Lord Sauron, and Little Bunny Foo-Foo. The parole board has declared that it will release two of the three, chosen uniformly at random, but has not yet released their names. Naturally, Sauron figures that he will be released to his home in Mordor, where the shadows lie, with probability $\frac{2}{3}$.

A guard offers to tell Sauron the name of one of the other prisoners who will be released (either Voldemort or Foo-Foo). However, Sauron declines this offer. He reasons that if the guard says, for example, "Little Bunny Foo-Foo will be released", then his own probability of release will drop to $\frac{1}{2}$. This is because he will then know that either he or Voldemort will also be released, and these two events are equally likely.

Using a tree diagram and the four-step method, either prove that the Dark Lord Sauron has reasoned correctly or prove that he is wrong. Assume that if the guard has a choice of naming either Voldemort or Foo-Foo (because both are to be released), then he names one of the two uniformly at random.



$$\Pr(\text{Foo} \mid B) = \frac{\frac{1}{3}}{\frac{1}{3}+\frac{1}{6}} = \frac{2}{3} \quad \text{Too}$$

$$\Pr(\text{Foo} \mid C) = \frac{2}{3} \quad \text{too}$$

没有影响.

6.042J / 18.062J Mathematics for Computer Science
Fall 2010

Problems for Recitation 19

1 Bayes' Rule

Bayes' Rule says that if A and B are events with nonzero probabilities, then:

$$\Pr\{A \mid B\} \cdot \Pr\{B\} = \Pr\{B \mid A\} \cdot \Pr\{A\}$$

a. Prove Bayes' Rule.

$$\Pr\{A \mid B\} = \frac{\Pr\{B \mid A\} \cdot \Pr\{A\}}{\Pr\{B\}}$$

b. A weatherman walks to work each day. Some days it rains:

$$\Pr\{\text{rains}\} = 0.30$$

Sometimes the weatherman brings his umbrella. Usually this is because he predicts rain, but he also sometimes carries it to ward off bright sunshine.

$$\Pr\{\text{carries umbrella}\} = 0.40$$

As a weatherman, he usually doesn't get caught out in a storm without protection:

$$\Pr\{\text{carries umbrella} \mid \text{rains}\} = 0.80$$

Suppose you see the weatherman walking to work, carrying an umbrella. What is the probability of rain? Use Bayes' Rule.

$$\Pr(\text{rains} \mid \text{car}\ldots) = \frac{0.80 \times 0.30}{0.40}$$

$$= 0.6$$

2 DNA Profiles

Suppose that we create a a national database of DNA profiles. Let's make some (overly) simplistic assumptions:

- Each person can be classified into one of 20 billion different "DNA types". (For example, you might be type #13,646,572,661 and the person next to you might be type #2,785,466,098.) Let $T(x)$ denote the type of person x.
- Each DNA type is equally probable.
- The DNA types of Americans are mutually independent.

a. A congressman argues that there are only about 300 million Americans, so even if a profile for every American were stored in the database, the probability of even one coincidental match would be very small.

Recall from lecture that if there are N days in a year and m people in a room, then the probability that no two people in the room have the same birthday is about $e^{-m^2/(2N)}$. Using this fact, what is the probability that two people's DNA profiles would match if every person's profile were stored in the database?

$1 - e^{-m^2/2N}.$

3~~00,000,000~~

20,0~~00,000,000~~

so ~~small~~'

$= 1 - e^{-\frac{9}{4}\cdot 10^6}$

b. After this database is implemented, some DNA is found at a crime scene. The DNA is sequenced and a person with matching DNA is found through the database and accused of the crime. At the trial the defense attorney argues that, by the birthday principle, the probability that there are multiple people whose DNA is identical is a virtual certainty, and so the jury cannot conclude beyond a reasonable doubt that the defendant is the criminal.

What is the flaw in this argument? Under what circumstances could you conclude based on DNA evidence alone that there is no doubt that the defendant committed the crime? (assuming no errors in the DNA tests, a comprehensive database, etc. etc.)

3 The Immortals

There were n Immortal Warriors born into our world, but in the end *there can be only one.* The Immortals' original plan was to stalk the world for centuries, dueling one another with ancient swords in dramatic landscapes until only one survivor remained. However, after a thought-provoking discussion of probabilistic independence, they opt to give the following protocol a try:

1. The Immortals forge a coin that comes up heads with probability p.
2. Each Immortal flips the coin once.
3. If *exactly one* Immortal flips heads, then he or she is declared The One. Otherwise, the protocol is declared a failure, and they all go back to hacking each other up with swords.

a. One of the Immortals (the Kurgan from the Russian steppe) argues that as n grows large, the probability that this protocol succeeds must tend to zero. Another (McLeod from the Scottish highlands) argues that this need not be the case, provided p is chosen *very carefully*. What does your intuition tell you?

Kurgan right.

b. What is the probability that the experiment succeeds as a function of p and n?

$$Pr = \frac{1}{n} \cdot (p) \cdot (1-p)^{n-1}$$

c. How should p, the bias of the coin, be chosen in order to maximize the probability that the experiment succeeds? (You're going to have to compute a derivative!)

$$f(p) = \frac{1}{n} \cdot p \cdot (1-p)^{n-1} \qquad p = \frac{1}{5} \qquad p = 1$$

$$f'(p) = \frac{1}{n} \cdot (1-p)^{n-1} + \frac{p}{n} \cdot (n-1) \cdot (1-p)^{n-2} \cdot (-1).$$

$$= \frac{1}{n}(1-p)^{n-1} - \frac{p \cdot (n-1)}{n} \cdot (1-p)^{n-2}$$

$$(1-p)^{n-2} \left(\frac{(1-p)}{n} - p(n-1) \right) \qquad n = 2.$$

$> 0.$ if n $< 0.$

d. What is the probability of success if p is chosen in this way? What quantity does this approach when n, the number of Immortal Warriors, grows large?

$$f(p) = \frac{1}{n} \cdot n \cdot \left(1 - \frac{1}{n}\right)^{n-1}$$

$$= \frac{\left(1 - \frac{1}{n}\right)^{n}}{n}.$$

$$\lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^{n} \approx \frac{1}{e}$$

$$\left(\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^{n} = e^{x} \right)$$

Problems for Recitation 20

Suppose that a coin that comes up heads with probability p is flipped n times. Then for all $\alpha < p$

$$\Pr\{\#\text{ heads} \leq \alpha n\} \leq \frac{1-\alpha}{1-\alpha/p} \cdot \frac{2^{nH(\alpha)}}{\sqrt{2\pi\alpha(1-\alpha)n}} \cdot p^{\alpha n}(1-p)^{(1-\alpha)n}$$

where:

$$H(\alpha) = \alpha \log_2 \frac{1}{\alpha} + (1-\alpha)\log_2 \frac{1}{1-\alpha}$$

1 Approximating the Cumulative Binomial Distribution Function

A coin that comes up heads with probability p is flipped n times. Find an upper bound on

$$\Pr\{\#\text{ heads} \geq \beta n\}$$

where $\beta > p$. Think about the number of tails and plug into the monster formula above.

$\Pr\{\# T \leq (n - \beta n)\}$ $\beta > p$

$\Pr\{\# T \leq n(1-\beta)\}$. $-p > -\beta$

$1-p > 1-\beta$.

So. $\dfrac{1-(1-\beta)}{1-\dfrac{(1-\beta)}{1-p}} \cdot \dfrac{2^{n \cdot H(1-\beta)}}{\sqrt{2\pi\beta \cdot (1-\beta) \cdot n}} \cdot \cdot \cdot \cdot$

2 Gallup's Folly

A Gallup poll found that 45% of the adult population of the United States plan to vote Republican in the next election. Gallup polled 640 people and claims a margin of error of 3 percentage points.

Let's check Gallup's claim. Suppose that there are m adult Americans, of whom pm plan to vote Republican and $(1-p)m$ do not. Gallup polls n Americans selected uniformly and independently at random. Of these, qn plan to vote Republican and $(1-q)n$ do not. Gallup then estimates that the fraction of Americans who plan to vote Republican is q.

Note that the only randomization in this experiment is in who Gallup chooses to poll. So the sample space is all sequences of n adult Americans. The response of the i-th person polled is "yes" with probability p and "no" with probability $1-p$ since the person is selected uniformly at random. Furthermore, the n responses are mutually independent.

a. Give an upper bound on the probability that the poll's estimate will be 0.04 or more too low. Just write the expression; don't evaluate yet!

$\Pr\{qn \leq (p - 0.04)n\}$

b. Give an upper bound on the probability that the poll's estimate will be 0.04 or more too high. Again, just write the expression.

c. The sum of these two answers is the probability that Gallup's poll will be off by 4 percentage points or more, one way or the other. Unfortunately, these expressions both depend on p— the unknown fraction of voters planning to vote Republican that Gallup is trying to estimate!

   However, the sum of these two expressions is maximized when $p = 0.5$. So evaluate the sum with $p = 0.5$ and $n = 640$ to upper bound the probability that Gallup's error is 0.04 or more. Pollsters usually try to ensure that there is a 95% chance that the actual percentage p lies within the poll's error range, which is $q \pm 0.04$ in this case. Is Gallup's poll properly designed?

3 Noisy Channel

Suppose we are transmitting packets of data across a noisy channel. Each packet has probability .01 of being lost. Now suppose we are transmitting $10,000$ packets. What is the probability that at most 2% of the packets are lost?

6.042J / 18.062J Mathematics for Computer Science
Fall 2010

Problems for Recitation 21

Problem 1. [**points**] Here's yet another fun 6.042 game! You pick a number between 1 and 6. Then you roll three fair, independent dice.

- If your number never comes up, then you lose a dollar.
- If your number comes up once, then you win a dollar.
- If your number comes up twice, then you win two dollars.
- If your number comes up three times, you win *four* dollars!

What is your expected payoff? Is playing this game likely to be profitable for you or not?

$$\frac{5}{6} \cdot \frac{5}{6} \times \frac{5}{6} \times -1 = -\frac{125}{216}$$

$$+ \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} \times 3 = \frac{75}{216}$$

$$+ \frac{1}{6} \cdot \frac{1}{6} \times \frac{5}{6} \times 3 \times 2 = \frac{30}{216}$$

$$+ \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \times 4 = \frac{4}{216}$$

$E(\cdots)$

$$= -\frac{16}{216}$$

$$= -\frac{2}{17}$$

Problem 2. [**points**] The number of squares that a piece advances in one turn of the game Monopoly is determined as follows:

- Roll two dice, take the sum of the numbers that come up, and advance that number of squares.
- If you roll *doubles* (that is, the same number comes up on both dice), then you roll a second time, take the sum, and advance that number of additional squares.
- If you roll doubles a second time, then you roll a third time, take the sum, and advance that number of additional squares.
- However, as a special case, if you roll doubles a third time, then you go to jail. Regard this as advancing zero squares overall for the turn.

first ?

(a) [pts] What is the expected sum of two dice, given that the same number comes up on both?

$$Ex(\text{sum} \mid \text{same}) = \left[7 + 7 + \tfrac{1}{6}\cdot(7)\right] \cdot \frac{35}{36}$$

(b) [pts] What is the expected sum of two dice, given that different numbers come up? (Use your previous answer and the Total Expectation Theorem.)

$$Ex(\text{sum} \mid \text{diff}) = 7.$$

(c) [pts] To simplify the analysis, suppose that we always roll the dice three times, but may ignore the second or third rolls if we didn't previously get doubles. Let the random variable X_i be the sum of the dice on the i-th roll, and let E_i be the event that the i-th roll is doubles. Write the expected number of squares a piece advances in these terms.

$$Ex(sum) = \Big[X_1 + Pr(E_1) \cdot X_2 + Pr(E_2) \cdot P(E_1) \cdot X_3 \Big] \Big(1 - Pr(E_1) \cdot Pr(E_2) \cdot Pr(E_3) \Big)$$

(d) [pts] What is the expected number of squares that a piece advances in Monopoly?

Problems for Recitation 23

Theorem 1. *Let $E_1, \ldots, E_n$ be events, and let X be the number of these events that occur. Then:*

$$\text{Ex}(X) = \Pr\{E_1\} + \Pr\{E_2\} + \ldots + \Pr\{E_n\}$$

Theorem 2 (Markov's Inequality)**.** *Let X be a nonnegative random variable. If $c > 0$, then:*

$$\Pr\{X \geq c\} \leq \frac{\text{Ex}(X)}{c}$$

Theorem 3 (Chebyshev's Inequality)**.** *For all $x > 0$ and any random variable R,*

$$\Pr\{|R - \text{Ex}(R)| \geq x\} \leq \frac{\text{Var}[R]}{x^2}$$

Theorem 4 (Union Bound)**.** *For events $E_1, \ldots, E_n$:*

$$\Pr\{E_1 \cup \ldots \cup E_n\} \leq \Pr\{E_1\} + \ldots + \Pr\{E_n\}$$

Theorem 5 ("Murphy's Law")**.** *If events $E_1, \ldots E_n$ are mutually independent and X is the number of these events that occur, then:*

$$\Pr\{E_1 \cup \ldots \cup E_n\} \geq 1 - e^{-\text{Ex}(X)}$$

Theorem 6 (Chernoff Bounds)**.** *Let $E_1, \ldots, E_n$ be a collection of mutually independent events, and let X be the number of these events that occur. Then:*

$$\Pr\{X \geq c\,\text{Ex}(X)\} \leq e^{-(c \ln c - c + 1)\,\text{Ex}(X)} \qquad \textit{when } c > 1$$

1 Getting Dressed

Sometimes I forget a few items when I leave the house in the morning.

For example, here are probabilities that I forget various pieces of footwear:

left sock	0.2
right sock	0.1
left shoe	0.1
right shoe	0.3

a. Let X be the number of these that I forget. What is $\text{Ex}(X)$?

$$\text{Ex}(X) = 0.7.$$

b. Upper bound the probability that I forget one or more items. Make no independence assumptions.

$$\Pr(X \geq 1) = \Pr\left(X \geq \frac{10}{7} \cdot \text{Ex}(X)\right) \leq \frac{7}{10} \cdot X$$

use upper bond: $\Pr(X \geq 1) \leq 0.2 + 0.1 + 0.1 + 0.3 = 0.7$

(发生至少一件)

c. Use the Markov Inequality to upper bound the probability that I forget 3 or more items.

$$\Pr(X \geq 3) \leq \frac{0.7}{3} = \frac{7}{30}$$

d. Now suppose that I forget each item of footwear independently. Use Chebyshev's Inequality to upper bound the probability that I forget two or more items.

$$\text{Var}(X) = \text{Var}(X_1) + \cdots + \text{Var}(X_4)$$
$$= 0.2 \times 0.8 + 0.1 \times 0.9 \times 2 + 0.3 \times 0.7$$
$$= 0.16 + 0.18 + 0.21$$
$$= 0.55.$$

$$\Pr(X \geq 2) = \Pr(X - \text{Ex}(X) \geq 1.3)$$

$= Pr\left((X - Ex) \geq \frac{1.3}{\sqrt{0.55}}\right) = \frac{1}{c^2+1}$

$\frac{0.55}{1.3^2+1}$ $\quad \frac{1}{2.6+1}$ 3.7

e. Use Theorem 5 to lower bound the probability that I forget one or more items.

$Pr \geq 1 - e^{-0.7}$

g. I'm supposed to remember many other items, of course: clothing, watch, backpack, notebook, pencil, kleenex, ID, keys, etc. Let X be the total number of items I remember. Suppose I remember items mutually independently and $\text{Ex}(X) = 36$. Use Chernoff's Bound to give an upper bound on the probability that I remember 48 or more items.

$Pr(X \geq 48) = Pr\left(X \geq \frac{4}{3} Ex\right)$

$c = \frac{4}{3} \quad k = c \cdot \ln c + c - 1 =$

$Pr \leq e^{-k \cdot 36}$

h. Give an upper bound on the probability that I remember 108 or more items.

$c = 3.$

$Pr \leq e^{-(3 \cdot \ln 3 + 3 - 1) \cdot 36}$

= very small.

2 A Financial Crisis

There are a lot of foreign words here, but don't be scared! We will be trying to understand why the subprime mortgage collapse happened!

For a more complete story of how the crisis happened, refer to section 19.5.3 of the text. The following is a set of vocabulary that we will be using:

- A **loan** is money lent to a borrower. If the borrower does not pay on the loan, the loan is said to be in **default**, and collateral is seized. In the case of mortgage loans, the borrower's home is used as collateral.
- A **bond** is a collection of loans, packaged into one entity. A bond can be divided into **tranches**, in some ordering, which tell us how to assign losses from defaults. Suppose a bond contains 1000 loans, and is divided into 10 tranches of 100 bonds each. Then, all the defaults must fill up the lowest tranche before the affect others. For example, suppose 150 defaults happened. Then, the first 100 defaults would occur in tranche 1, and the next 50 defaults would happen in tranche 2.
- The lowest tranche of a bond is called the **mezzanine tranche**.
- We can make a "super bond" of tranches called a **collateralized debt obligation (CDO)** by collecting mezzanine tranches from different bonds. This super bond can then be itself separated into tranches, which are again ordered to indicate how to assign losses.

Armed with this knowledge, we can now solve problems about the crisis!

1. Suppose that 1000 loans make up a bond, and the fail rate is 5% in a year. Assuming mutual independence, give an upper bound for the probability that there are one or more failures in the second-worst tranche. What is the probability that there are failures in the best Tranche? $Ex = 50.$ $Pr(>100) =$ $c=2.$ $\le e^{-(2\cdot\ln 2+c-1)\cdot 50}$
   几乎为0.
2. Now, do not assume that the loans are independent. Give an upper bound for the probability that there are one or more failures in the second tranche. What is an upper bound for the probability that the entire bond defaults? Show that it is a tight bound. (Hint: Use Markov's theorem). $Pr(>100) = Pr(X > 2\cdot Ex) \le Pr(X \ge 2\cdot Ex) \le \frac{1}{2}.$
   $Pr(>1000) = \frac{1}{20}.$
3. Given this setup (and assuming mutual independence between the loans), what is the expected failure rate in the mezzanine tranche? 50%
4. We take the mezzanine tranches from 100 bonds and create a CDO. What is the expected number of underlying failures to hit the CDO? 100×100 50% $= 5000$
5. We divide this CDO into 10 tranches of 1000 bonds each. Assuming mutual independence, give an upper bound on the probability of one or more failures in the best tranche. The third tranche? $Pr(X > \underbrace{9000}_{1.8\cdot Ex}) < e^{-(1.8\cdot\ln 1.8+1.8-1)5000}$
6. Repeat the previous question without the assumption of mutual independence.
   $\frac{5}{9}$. $\frac{5}{7}$

Problems for Recitation 22

1 Properties of Variance

In this problem we will study some properties of the variance and the standard deviation of random variables. $Var[R] = \sum_x ((R - E(x))^2) = Ex(R^2 - 2R \cdot Ex R + E^2(R))$

a. Show that for any random variable R, $\text{Var}[R] = \text{E}[R^2] - \text{E}^2[R]$. $= Ex[R^2] - 2Ex[R] \cdot Ex[R] + Ex^2 =$

b. Show that for any random variable R and constants a and b, $\text{Var}[aR+b] = a^2 \text{Var}[R]$. Conclude that the standard deviation of $aR+b$ is a times the standard deviation of R.

$Var[aR+b]$
$= Ex[(aR+b)^2] - Ex^2(aR+b)$
$= Ex(a^2R^2 + 2abR + b^2) - (a \cdot Ex(R) + b)^2$
$= a^2 Ex(R^2) - a^2 Ex^2(R)$
$= a^2 \cdot Var[R]$

c. Show that if R_1 and R_2 are independent random variables, then

$$\text{Var}[R_1 + R_2] = \text{Var}[R_1] + \text{Var}[R_2].$$

$Var[R_1+R_2] = Ex(R_1^2 + 2R_1R_2 + R_2^2) - [Ex(R_1) + Ex(R_2)]^2$
$= Var[R_1] + Var[R_2] + Ex(2R_1 \cdot R_2) - 2ExR_1 \cdot ExR_2$

independ.
$Ex(R_1 \cdot R_2) = Ex(R_1) \cdot Ex(R_2)$
so 0.

d. Give an example of random variables R_1 and R_2 for which

$$\text{Var}[R_1 + R_2] \neq \text{Var}[R_1] + \text{Var}[R_2].$$

投一个硬币, $R_1 = T$, $R_2 = H$.

e. Compute the variance and standard deviation of the Binomial distribution $H_{n,p}$ with parameters n and p.

f. Let's say we have a random variable T such that $T = \sum_{j=1}^{n} T_j$, where all of the T_j's are mutually independent and take values in the range $[0, 1]$. Prove that Var(T)≤Ex(T). We'll use this result in lecture tomorrow. *Hint: Upper bound* $\text{Var}[T_j]$ *with* $\text{E}[T_j]$ *using the definition of variance in part (a) and the rule for computing the expectation of a function of a random variable.*

e. $Ex = p$.

p

indep $Var = Var_1 + Var_2 + \cdots + Var_n$
$= n \cdot (Var_1)$
$= n \cdot [p - p^2]$

f. $Var(T) = Var(T_1) + Var(T_2) + \cdots + Var(T_n)$
对任意 T_j. $Var(T_j) = Ex(T_j^2) - Ex^2(T_j)$

$T_j \in [0,1]$

$\leq Ex(T_j) - E^2_{(x)}(T_j)$

$\leq Ex(T_j)$.

$\therefore Var(T) \leq Ex(T_1) + Ex(T_2) + \cdots Ex(T_n)$
$= Ex(T)$