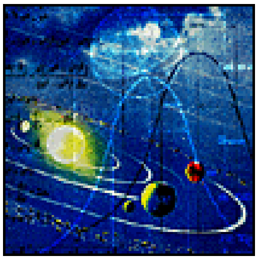
Great Theoretical Ideas In Computer Science

Victor Adamchik

CS 15-251

Carnegie Mellon University

# Probability - II



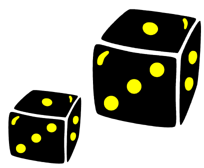
## Exam

Tuesday at 3 pm, same room
All topics
3x5 index card
Practice exam
Review: Sunday, 6pm at DH 2210
Format:
- short answers (5-6 prblms)
- variant of HW question
- long problems with proofs (3-4 prblms)

## Outline

Distributions
Expectation
Conditional Expectation
Tail bounds

Review:
some useful
sample spaces…



1) A fair coin

sample space S = {H, T}

Pr(H) = $\frac{1}{2}$ , Pr(T) = $\frac{1}{2}$.

2) A "bias-p" coin

sample space S = {H, T}

Pr(H) = p, Pr(T) = 1-p.

## Binomial Distribution B(n,p)

3) We flip a bias-p coin n times

sample space S = $\{H,T\}^n$

if outcome x in S has k heads and n-k tails

Event $E_k$ = {x in S | x has k heads}

## Example

Teams A is better than team B

The odds of A winning are 6:5

i.e., in any game, A wins with probability 6/11

What is the chance that A will beat B in the "best of 7" world series?

## Example

Team A beats B with probability p=6/11 in each game

Sample space $S = \{W, L\}^7$

$Pr[x] = p^k (1-p)^{7-k}$ if there are k W's in x

Want event E = "team A wins at least 4 games"

$E = E_4 \cup E_5 \cup E_6 \cup E_7$ where $E_k = \{x \text{ in } S \mid x \text{ has } k \text{ W's}\}$

## Geometric Distribution

A biased coin is tossed until the first time that a head turns up.

sample space S = {H, TH, TTH, TTTH, ...}

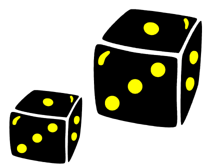
(shorthand S = {1, 2, 3, 4, ...})

$Pr[k] = (1-p)^{k-1}p$

sanity check:

$$\sum_{k\geq 1} Pr[k] = \sum_{k\geq 1} p\,(1-p)^{k-1} = p\sum_{k\geq 0}(1-p)^k = p\frac{1}{1-(1-p)} = 1$$

## Expectation
## a.k.a. Expected Value
## a.k.a. Mean



## Expectation

Intuitively, expectation of X is what its average value would be if you ran the experiment millions and millions of times.

Definition:

Let X be a random variable in experiment with sample space $\Omega$. Its expectation is defined by:

$$E[X] = \sum_{k\in\Omega} Pr[k]\,X(k)$$

## Expectation

E[X] can be viewed as a sum of possible outcomes, each weighted by its probability

$$E[X] = \sum_{k\in\Omega} Pr[k]\,X(k)$$

$$E[X] = \sum_{k\in\Omega} Pr[X=k]\;k$$

Here a discrete r.v. X takes values X(k) with corresponding probabilities Pr[k]

## Example

Let R be the roll of a standard die.

$$E[R] = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6$$

$$= 3.5$$

Question: What is Pr[R = 3.5]?

Answer: 0. Don't always expect the expected!

## Example

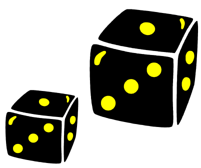
Suppose you win $30 on a roll of double-6, and you lose $1 otherwise. Let W be the random variable representing your winnings.

$$E[R] = \frac{1}{36}\cdot(-1) + \frac{1}{36}\cdot(-1) + \ldots + \frac{1}{36}\cdot(-1) + \frac{1}{36}\cdot 30$$

$$= -5/36 \approx -13.9¢$$

## One of the top tricks in probability...

Linearity of Expectation



## Linearity of Expectation

Given an experiment, let X and Y be random variables.

Then $E[X+Y] = E[X] + E[Y]$

X and Y do *not* have to be independent!!

## Linearity of Expectation

$E[X+Y] = E[X] + E[Y]$

Proof: Let Z = X+Y (another random variable).

Then

$$E[Z] = \sum_{k\in\Omega} \Pr[k]\, Z(k) = \sum_{k\in\Omega} \Pr[k]\, (X(k) + Y(k))$$

$$= \sum_{k\in\Omega} \Pr[k]\, X(k) + \sum_{k\in\Omega} \Pr[k]\, Y(k) = E[X] + E[Y]$$

## Linearity of Expectation

$E[X+Y] = E[X] + E[Y]$

Also: $E[aX + b] = a\,E[X] + b$ for any $a,b \in \mathbb{R}$,

By Induction

$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$

## Indicator Random Variables

Definition:

Let A be an event. The indicator of A is the random variable X which is 1 when A occurs and 0 when A doesn't occur.

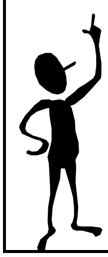
$X : \Omega \to \mathbb{R}$ $\qquad X(k) = \begin{cases} 1, \text{if } k \in A \\ 0, \text{if } k \notin A \end{cases}$

## Expectation of an Indicator

Let X be an indicator r.v. for the event A.

The expectation of this indicator is

$E[X] = 1 * \Pr[A] + 0 * \Pr[A^c] = \Pr[A]$



Linearity of Expectation

+

Indicators

= best friends forever

## Exercise

There are 251 students in a class.

The TAs randomly permute their midterms before handing them back.

Let X be the number of students getting their own midterm back.

What is E[X]?

## Solution

Let $A_i$ be the event that $i^{th}$ students gets own midterm.

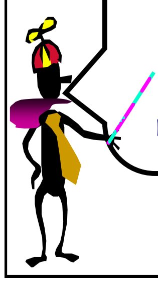
Let $X_i$ be the indicator of $A_i$.

Then $X = X_1 + X_2 + \cdots + X_n$

Thus $E[X] = E[X_1] + E[X_2] + \cdots + E[X_n]$ by linearity of expectation

$E[X_i] = \Pr[A_i]$, and $\Pr[A_i] = 1/251$ for each i.

It follows $E[X] = 251 \cdot (1/251) = 1$



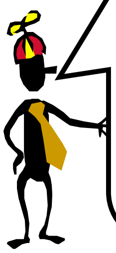
So, in expectation, 1 student will receive his/her midterm.

Pretty neat: it doesn't depend on how many students!

Question: were the $X_i$ independent?

No! E.g., think of n=2

## Type Checking



Pr[B] — B must be an event

E[X] — X must be a R.V.

cannot do Pr( R.V. ) or E[event ]

## Operations on R.V.s

You can sum them, take differences, or do most other math operations (they are just functions!)

E.g., $(X + Y)(t) = X(t) + Y(t)$

$(X*Y)(t) = X(t) * Y(t)$

$(X^Y)(t) = X(t)^{Y(t)}$

## Expectation of a Sum of r.v.s = Sum of their Expectations

even when r.v.s are not independent!

---

## Expectation of a Product of r.v.s vs. Product of their Expectations ?

## Multiplication of Expectations

A coin is tossed twice.
$X_i$ = 1 if the $i^{th}$ toss is heads and 0 otherwise.

$E[ X_i ] = 1/2$

$E[ X_1 X_2 ] = 1/4$ $\quad$ $E[ X_1 ] E[ X_2 ] = 1/4$

Lemma:
$E[XY] = E[X] E[Y]$
if X and Y are independent random variables.

Proof left as exercise.

## Multiplication of Expectations

Consider a single toss of a coin.
X = 1 if heads turns up and 0 otherwise.

Set $Y = 1 - X$

$E[X] = E[Y] = 1/2$

$E[X Y] \neq E[X] E[Y]$

since $X Y = 0$

X and Y are not independent

## More examples of Computing Expectations

## Geometric Random Variables

X ~ Geometric(p)

What is E[X]?

Average number of p-biased coin flips until you get Heads: you might guess 1/p.

$$E[X]=\sum_{k\geq 1} k\cdot \Pr[k] = \sum_{k\geq 1} k\, p\,(1-p)^{k-1} = p\sum_{k\geq 0} k\, q^{k-1}$$

q = 1-p

$$= p\frac{d}{dq}\left(\sum_{k\geq 0} q^k\right) = p\frac{d}{dq}\left(\frac{1}{1-q}\right) = \frac{p}{(1-q)^2} = \frac{1}{p}$$

## The Coupon Collector

There are n different kinds of coupons.

On each day, you get a random coupon. (You may get duplicates.)

Let X be the # of days till you have them all.

What is E[X]?

## The Coupon Collector

Let X be the # of days till you have them all.

What is E[X]?

Key idea: Let $X_i$ be # of days it took you to go from i–1 to i coupons.

Key idea: $X = X_1 + X_2 + \cdots + X_n$

$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n]$

So we need to figure out $E[X_i]$.

## The Coupon Collector

Key idea: Let $X_i$ be # of days it took you to go from i–1 to i coupons.

When sitting on i–1 distinct coupons, each day you have probability $\frac{n-(i-1)}{n}$ of getting a new one.

$$X_i \sim \text{Geometric}\left(\frac{n-(i-1)}{n}\right)$$

$$E[X_i] = \frac{n}{n-(i-1)}$$

## The Coupon Collector

$$E[X]=\sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n}\frac{n}{n-(i-1)} = n\sum_{i=1}^{n}\frac{1}{i} = n\,H_n = O(n\ln n)$$

where $H_n$ = "the nth harmonic number"

## Using linearity of expectations in unexpected places…

10% of the surface of a sphere is colored green, and the rest is colored blue. Show that now matter how the colors are arranged, it is possible to inscribe a cube in the sphere so that all of its vertices are blue.

## Solution

Pick a random cube. (Note: any particular vertex is uniformly distributed over surface of sphere).

Let $X_i$ = 1 if $i^{th}$ vertex is blue, 0 otherwise

Let $X = X_1 + X_2 + \ldots + X_8$

$E[X_i] = P(X_i=1) = 9/10$

$E[X] = 8 * 9/10 > 7$

So, must have some cubes where X = 8.

## The general principle we used in this example:

Show the expected value of some random variable is "high"

Hence, there must be an outcome in the sample space where the random variable takes on a "high" value.

(Not everyone can be below the average.)

This is called "the probabilistic method"

## The Probabilistic Method

It was developed by Paul Erdos as a technique for proving that something exists by setting up some probability distribution and showing that what we want happens with probability > 0.

The basic technique is based on two observations:

1) If $E[X]=\mu$, then $\exists x>\mu$ s.t. $Pr[X=x] > 0$

2) If Pr[X] > 0, then X exists

## Conditional expectations

Just like probabilities, we can also talk about expectations *conditioned on some event.*

E[X | A] = expectation of X conditioned on event A

$$E[X|A] = \sum_k k\,Pr[X=k|A]$$

Law of total expectation:

$E[X] = Pr(A)\,E[X \mid A] + Pr(A^c)\,E[X \mid A^c]$

## Example

Two discrete r.v. X and Y have probabilities defined by the table below. Find E[X|Y=2].

| | | | |
|---|---|---|---|
| Y=2 | 0 | 1/6 | 1/8 |
| Y=1 | 1/8 | 1/6 | 1/8 |
| Y=0 | 1/6 | 1/8 | 0 |
| | X=0 | X=1 | X=2 |

$$E[X|Y=2] = \sum_k k\,Pr[X=k|Y=2] =$$

$$= 0\cdot\frac{Pr[X=0 \& Y=2]}{Pr[Y=2]} + 1\cdot\frac{Pr[X=1 \& Y=2]}{Pr[Y=2]} + 2\cdot\frac{Pr[X=2 \& Y=2]}{Pr[Y=2]}$$

$$= 1\cdot\frac{1/6}{1/6+1/8} + 2\cdot\frac{1/8}{1/6+1/8} = \frac{10}{7}$$

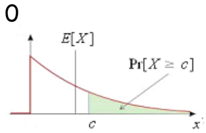
## Markov's inequality

Not too many people can be well above the average.

Suppose X is a non-negative r.v. with E[X] = 10
How often can X be 20 or higher?
i.e., How high can $Pr[X \geq 20]$ be?

Markov's inequality:
For a non-negative r.v. X, and c > 0

$Pr[X \geq c] \leq E[X]/c$



## Markov's inequality

For a non-negative r.v. X,

Pr[X ≥ a] ≤ E[X]/c

for every c > 0.

Proof.

$$E[X] = \sum_{x} x\,\Pr[X=x] = \sum_{0\le x<c} x\,\Pr[X=x] + \sum_{x\ge c} x\,\Pr[X=x]$$

Drop the first sum and replace x by c

$$E[X] \ge \sum_{x\ge c} x\,\Pr[X=x] \ge c\sum_{x\ge c} \Pr[X=x] = c\,\Pr[X\ge c]$$

QED.

Here's What You Need to Know…

Geometric and Binomial Distributions
Expected Value
Linearity of Expectation
Conditional Expectation
Law of Total Expectation
Probabilistic Method
Markov's Inequality