Mathematics for Computer Science

MIT 6.042J/18.062J

Deviation from the Mean

Bernoulli Trials

R ::= number of heads in n tosses of a fair coin

$$\Pr\{R=k\} = \binom{n}{k} 2^{-n}$$

Binomial Distribution (fair)

$$E[R] = \frac{n}{2}$$

Exactly the Mean?

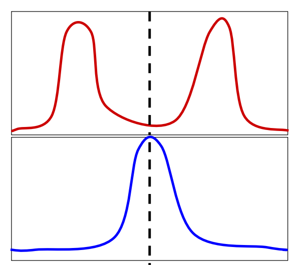
What is Pr{*exactly n*/2 heads}?

$$\Pr\{R=n/2\} = \binom{n}{n/2} 2^{-n}$$

If $n = 100$, $\Pr\{R=50\} = 8\%$!!!

Two Distributions, Same Mean

$\Pr\{R=x\}$

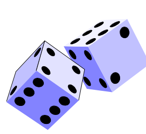


$x \longrightarrow$

Two Dice with the Same Mean

Fair Dice

- E[R] = 3.5



Loaded Dice: (throws either 1 or 6)

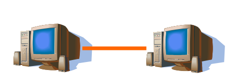
- E[R] = (1+6)/2 = 3.5!!!

Noisy Transmission Line

Total number of bits transmitted = 10,000

R = number of errors in 10,000 bits

$E[R]$ = 100 bits



- But what does the error look like on any specific transmission? ±100? ±1000?

What to Expect from “Expected Value”?

- If I toss *n* coins today, what should I expect to see?
- *Power of Probability:* the ability to predict the outcome of a specific event.

Back to Bernoulli Trials

EXPERIMENT:
Toss 100 coins and count # of heads (*R*)

Run this experiment many times

- Average number of heads = ?
- Exactly 50 heads = ?
- Between 40-60 heads = ?

Scheme Buffer (of course!)

```
; THROW A SEQUENCE OF n TOSSES of an UNBIASED COIN
(define (ntoss n)
  (if (= n 0)
      '()
      (cons (random 2)
            (ntoss (- n 1)))))

; RUN the EXPERIMENT t TIMES, and return a list with # HEADS
(define (experiment n t)
  (if (= t 0)
      '()
      (cons (reduce + 0 (ntoss n)) ; count # of heads (ones)
            (experiment n (- t 1)))))


; Functions to compute average, exactly V heads, heads in RANGE
(define (average-heads foo) (exact->inexact (/ (reduce + 0 foo) (length foo))))
(define (total-heads-v foo v)(reduce + 0 (map (lambda (heads)(if (= heads v) 1 0)) foo)))
(define (total-heads-range foo low high) (reduce + 0 (map (lambda (heads)
          (if (and (<= heads high) (>= heads low)) 1 0)) foo)))
```


Results

EXPERIMENT:
Toss 100 coins and count # of heads (*R*)

Ran this experiment 100 times

- Average number of heads = 49.5
- Exactly 50 heads = 4 out of 100
- Between 40-60 heads = 97 out of a 100!

Giving *Meaning* to the “Mean”

Let μ ::= E[*R*]

- What is Pr{R is within μ ±10}?
- What is Pr{R > 2% error}?
- On average, how much does R deviate from the expected value?
  – What is E[|*R* − μ|]?

Deviation from the Mean

Using the expected value to infer the probability of occurrence

- Markov Bound
- Chebyshev Bound (Variance)
- Chernoff Bound
- Binomial Approximation
- Actual Distribution…

↓ More information about the distribution

Sampling and Deviation from Mean

Using *event occurrence* to infer the expected value.

- Weak Law of Large Numbers
- Pairwise Independent Sampling Theorem
- Applications to Polling

Intuition about Averages

- Lets say we test n people and find out that the average IQ = 100
- *EXERCISE:* What fraction of the people can have an IQ ≥ 200?

IQ Higher than 200

Let f be the fraction of people who have an IQ of at least 200

- Then these people contribute at least $200f$ to the average IQ.
- So, $200f \leq 100$ or $f \leq 100/200 = ½$
- No more than ½ people have IQ ≥200.

Pr {IQ ≥ value} ≤ E[IQ] / value

Markov Bound

If R is non-negative, then

$$\Pr\{R \geq x\} \leq \frac{E[R]}{x}$$

for $x \geq 0$.

Another form of Markov Bound

Let $x = c \cdot E[R]$

$\Pr\{R \geq c \cdot E[R]\} \leq 1/c$

- Probability that deviate more than 2 times the expected value, is less than ½

Formal Proof of Markov Bound

$$E[R] ::= \sum_{r \in range(R)} r \cdot \Pr\{R = r\}$$

$$= \sum_{r \geq x} r \cdot \Pr\{R = r\} + \sum_{0 \leq r < x} r \cdot \Pr\{R = r\}$$

$$\geq \sum_{r \geq x} r \cdot \Pr\{R = r\} + 0$$

$$\geq \sum_{r \geq x} x \cdot \Pr\{R = r\}$$

$$= x \cdot \sum_{r \geq x} \Pr\{R = r\} = x \cdot \Pr\{R \geq x\}$$

Markov Bound

Markov's bound is

- Weak
- Obvious

Turns out to be useful anyway.

Better Bound

- Let $\mu ::= E[R]$
- What is Pr $\{|R - \mu| \geq x\}$?
  - what is the probability that R deviates more than x from the expected value

Chebyshev Bound

Variance

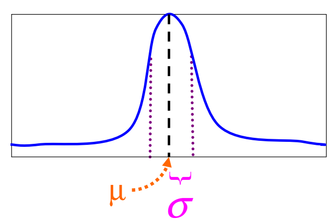
$$\Pr\{|R-\mu| \geq x\} \leq \frac{E[(R-\mu)^2]}{x^2}$$

Probability that I see a value more than x away from the expected value

Variance and Standard Deviation

$$\text{Var}[R] ::= E[(R-\mu)^2]$$

$$\sigma ::= \sqrt{\text{Var}[R]}$$



Why not just use E[|R−μ|] ?

Problem 1

Write solutions on the board

Interpreting Variance

Gambling Game: Same E[*Win*]
Different Var[*Win*]
(σ^2)

From Chebyshev:

$\Pr\{|W - E[W]| \geq c\} \leq \text{Var}[W]/c^2$

so, $\Pr\{|W - E[W]| \geq 2\sigma\} \leq$ ¼ = 25%

Pr $\{W$ in within $\pm 2\sigma\} \geq 75\%$

Conventional Wisdom

Probability that you are

within $\mu \pm 2\sigma$ > 75%

within $\mu \pm 3\sigma$ > 90%

Bernoulli Trials

$E[R] = \mu = 50$ heads

- $\Pr\{R \geq 60 \text{ heads}\} \leq 50/60$ (*Markov*)

$\text{Var}[R] = 25, \sigma = 5$

- $\Pr\{40 \leq R \leq 60\} \geq 1 - 25/100$ (*Chebyshev*)
  $= 75\%$

Actual Data = 97/100 times

In Class Problems

Problems 2-4