# Lecture 25, October 24

## Center and Spread

Slides created by Ani Adhikari and John DeNero

# Announcements

- Project 2 will be released this week!
- Homework due as usual
- I've posted on Piazza about courses to consider if you are interested in data science. I have no further info yet. I'll post on Piazza as soon as I do.

# The Average

Data: 2, 3, 3, 9     **Average = (2+3+3+9)/4 = 4.25**

- Not a value in the collection
- Need not be an integer even if the data are integers
- Somewhere between min and max, but not necessarily halfway in between
- Same units as the data
- Smoothing operator: collect all the contributions in one big pot, then split evenly

# Weights

Data: 2, 3, 3, 9

$4.25 = 2*(¼) + 3*(¼) + 3*(¼) + 9*(¼)$

$= 2*(¼) + 3*(2/4) + 9*(¼)$

$= 2*0.25 + 3*0.5 + 9*0.25$

(Demo)

# Average, Histogram, Median

- The average is the center of gravity of the histogram

- If the distribution is symmetric about a point, then that point is both the average and the median

- If the histogram is skewed (has a tail) in one direction, then the average is pulled away from the median in the direction of the tail

(Demo)

# How Far from the Average?

- Standard deviation (SD) measures roughly how far the data are from their average

- SD = root mean square of deviations from average

  5    4      3         2             1

- SD has the same units as the data; hence OK to say "average plus or minus a few SDs"

  (Demo)

# Chebychev's Bounds

For **all distributions**, no matter what their shape:

- % in the range "average ± **2** SD": **at least 75%**

- % in the range "average ± **3** SD": **at least 88.8888…%**

- % in the range "average ± *z* SD": **at least 1 - 1/$z^2$**

(Demo)

# Standard Units

- How many SDs above average?

- **$z = $ (value - mean)/SD**
  - Negative z: value below average
  - Positive z: value above average
  - z=0: value equal to average

- List in standard units: average = 0, SD = 1

- By Chebychev, most values of *z* are in (-5, 5)