Lecture 7 Cogsci 109

Thurs. Oct. 12, 2006

Announcements

- Homework 2 is posted.
- Office hours
- Midterm is coming up somewhere in the next couple of weeks
  - Anything lectured on, presented in section, in the reading and on homework is fair game
  - Will be part multiple choice, part short answer
  - Any calculations will be simple
  - Programming questions will be fairly qualitative

A few notes about last time

- Demos, transparency maps

Fourier transforms

Frequency analysis and Fourier transforms

Super/subsampling

- Super sampling (up-sampling)

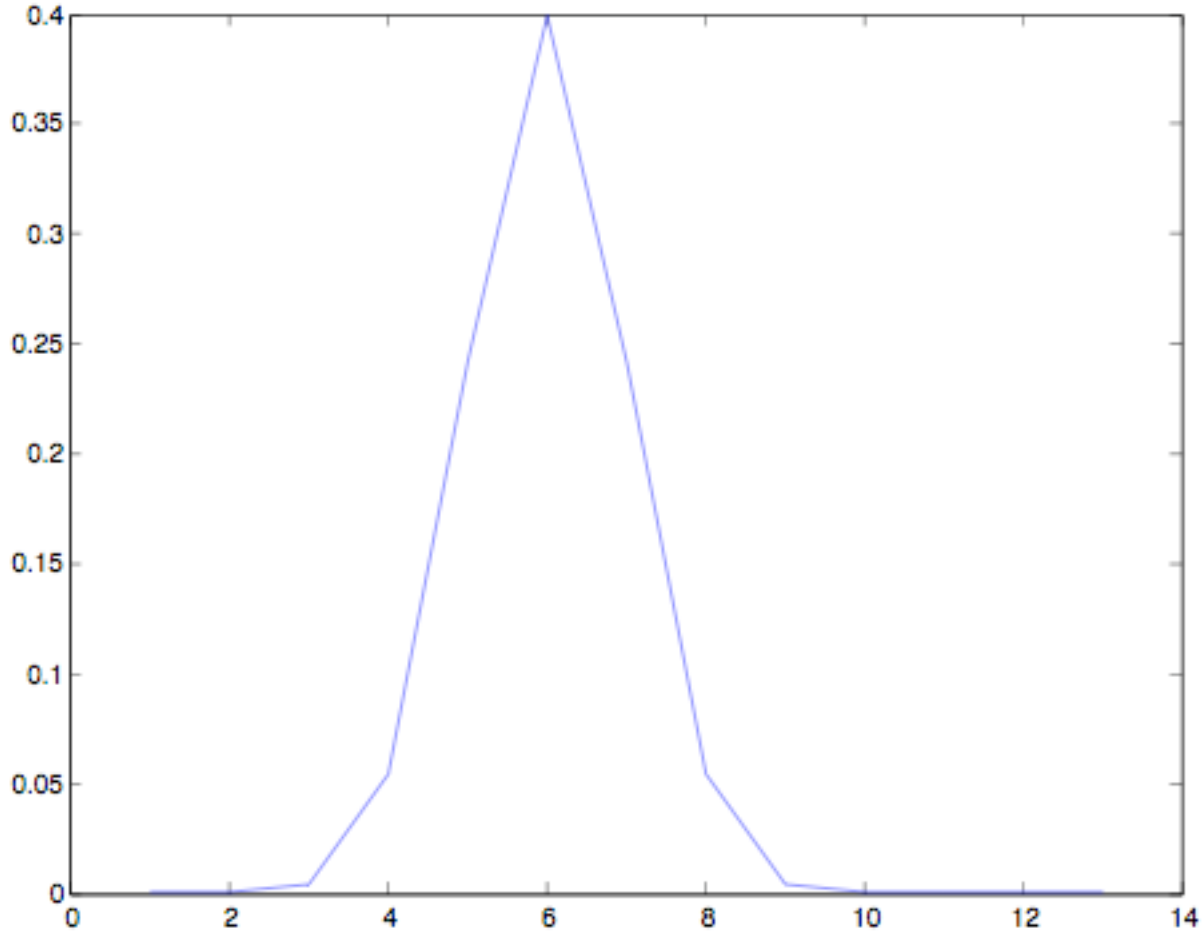
$$y(n) = \begin{cases} x\left(\frac{n}{N}\right), n = 0, \pm N, \pm 2N \\ 0, \quad otherwise \end{cases}$$

- Sub-sampling (down-sampling)
  - Take every M-th sample

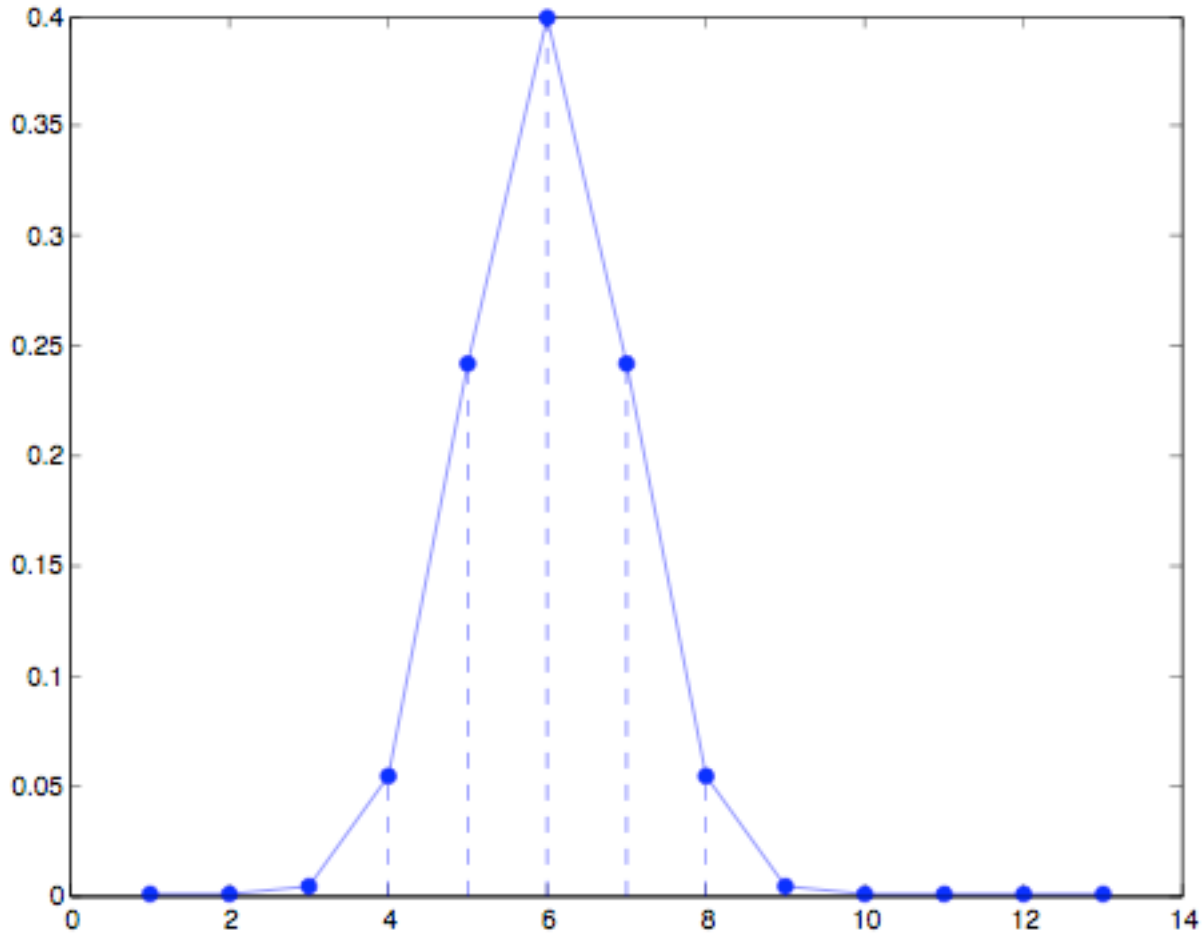
$$y(n) = x(Mn)$$

  - 1000Hz sample rate becomes 100Hz sample rate if we down sample by a factor of 10

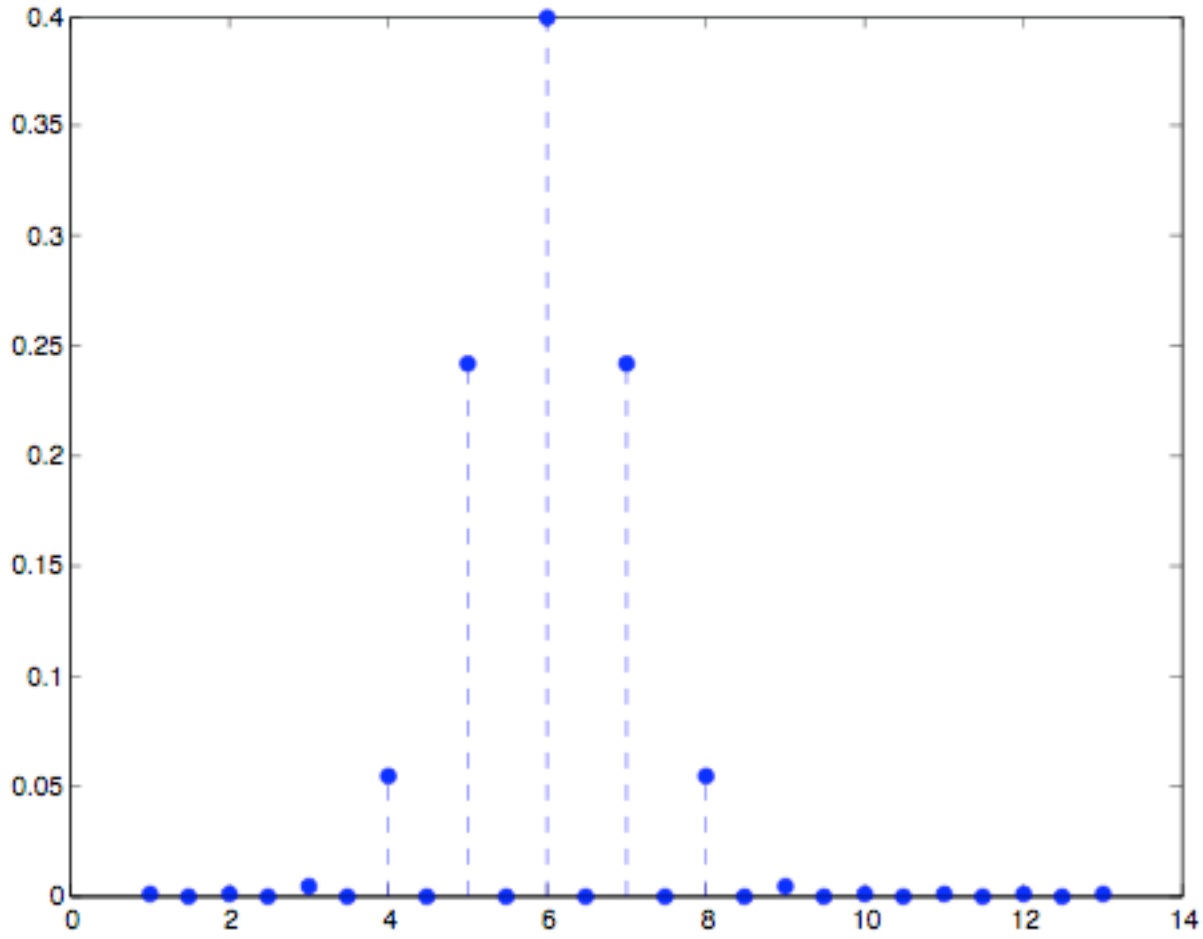
What does super-sampling look like?



What does super-sampling look like?



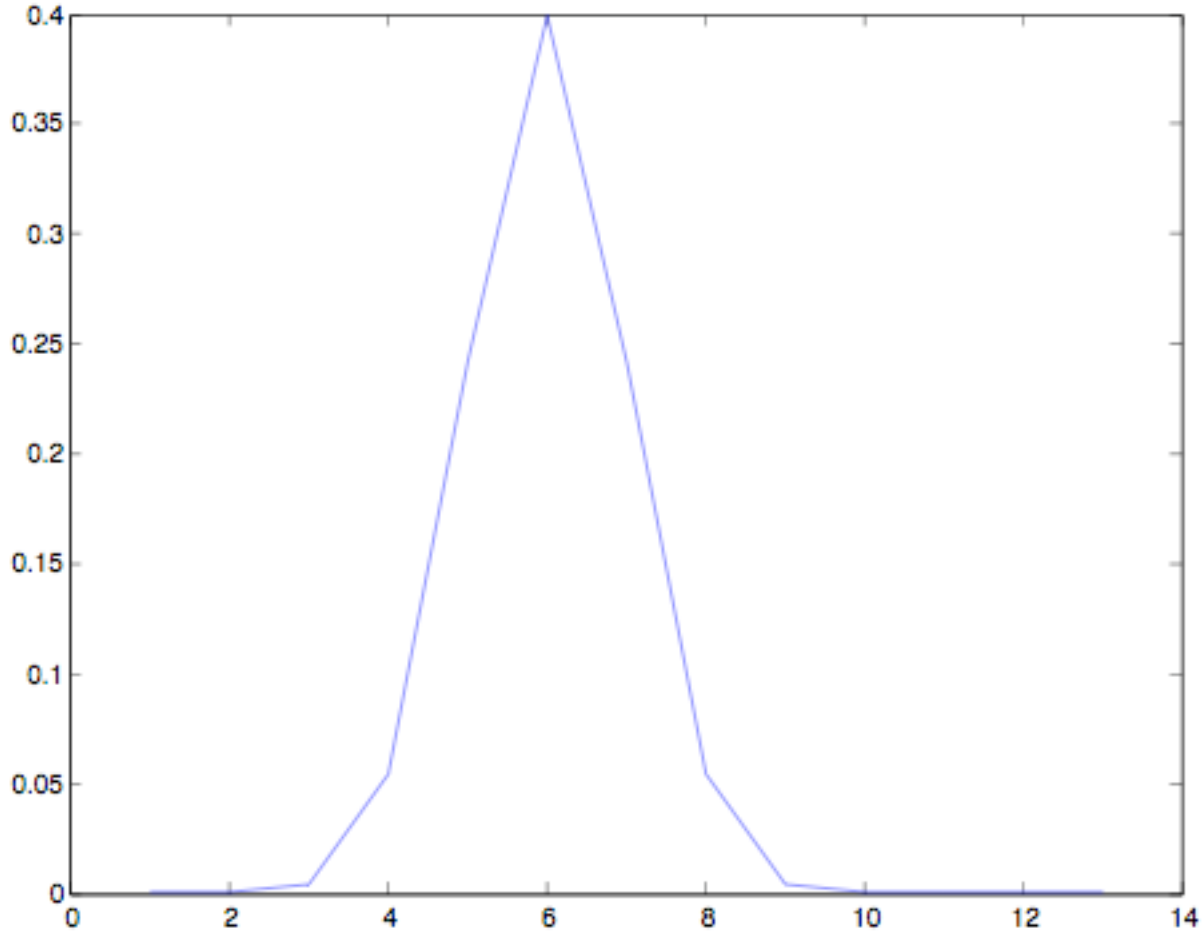
Example: Up-sampling



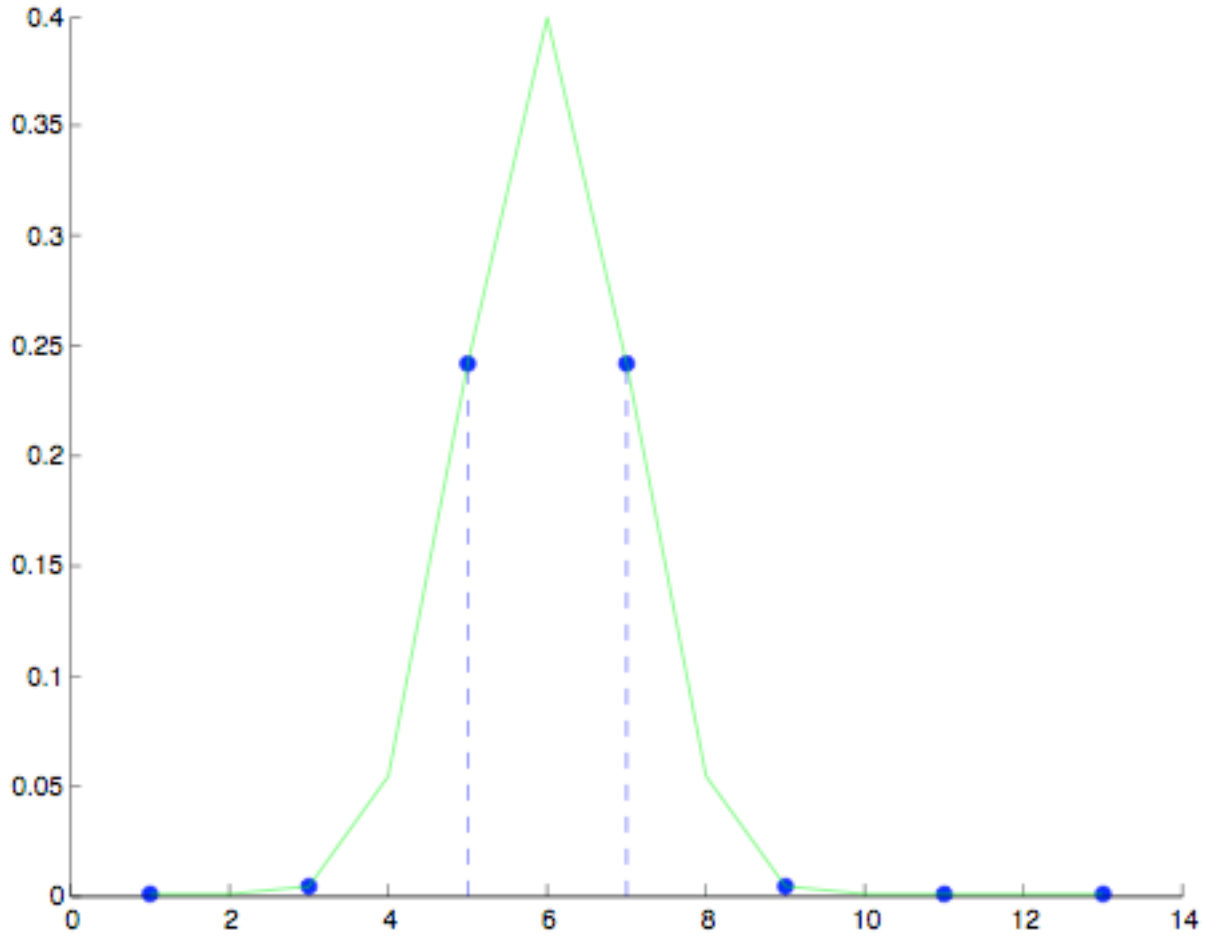
Are there other ways to do this?

- We could also interpolate the points instead of adding zeros…but that involves data fitting which we'll discuss soon

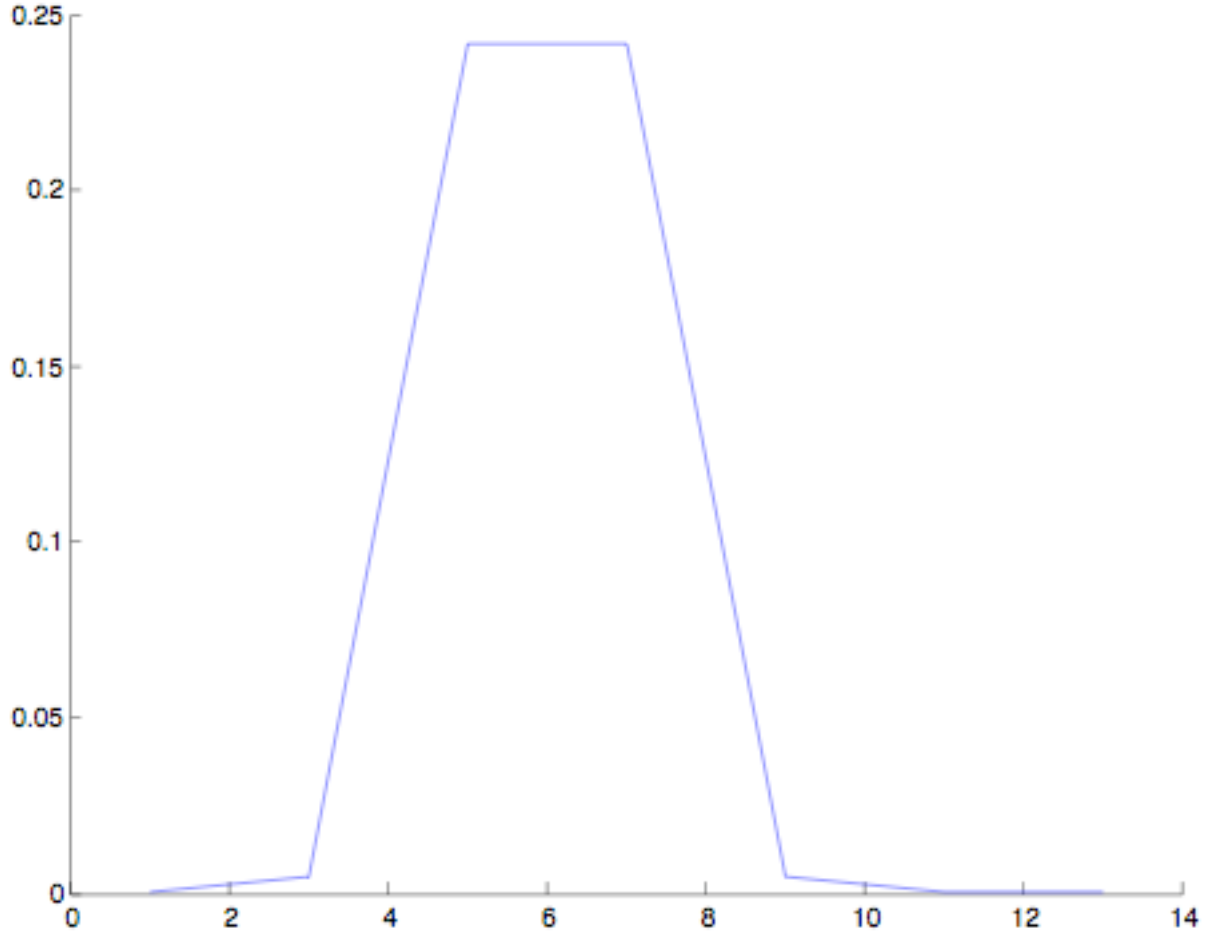
What does sub-sampling look like?



What does sub-sampling look like?



Then the sub-sampled data looks like



Quantitatively: Another way to ‘look’ at data

- How do we look at data quantitatively and extract meaningful information?
- Basic statistics is a start
  - Mode
  - Mean
  - Median
  - Standard Deviation
  - Variance
  - Covariance
  - Correlation

Central Tendency - *Mode*

- Most common number of a distribution
- Tells you which value has the highest frequency
- What if there are ties?
  - More than one mode!
  - Which of the following is the mode?

$$\{1,2,2,2,2,2,2,3,4,5,6,7,8,8,8,9,9\}$$

  - 2
- Matlab help: *help mode*

Central Tendency - *Mean*

- Think of it as similar to a balance point
- 'Expected value'
- Computed by the following
  - Sum all scores
  - Divide that sum by the number of scores
- Here's the formula:

$$M = \left(\sum_{i=1}^{N} x_i \right) \Big/ N$$

- And an example:

$$\{1.0, 1.0, 2.0, 3.0, 4.0, 4.0, 4.0, 4.0, 8.0, 8.0, 8.0, 8.0, 8.0, 8.0, 9.0, 0.0, 0.0, 0.0\}$$

$$N = 18$$

$$\sum x_i = 80.0$$

$$M = \left(\sum x_i \right) \Big/ N = 80.0 / 18 = 4.4$$

How to compute the mean in Matlab

- Function : *mean()*
- *Side note about matlab help and capitals*
- Example computation…
- Matlab help : *help mean*

Median

- The middle number of a distribution when the numbers have been ordered (sorted)
- Each score is counted separately, so if you have repeating scores such as 50 and 50, each one becomes part of the count
- Order the scores from low to high or high to low
- Count from both ends to the middle position

Median

- If **odd** number of scores, there will be one median
  - Example: *Find the median*

$$\{1,2,3,10,50\}$$
$$Median = 3$$

- If an **even** number of scores, count to the two closest to the middle (ie count from low towards high, high towards low) and take their average (add them up and divide by two)
  - Example: *Find the median*

$$\{1,2,2,3,3,4\}$$
$$2,3$$
$$Median = (2+3)/2 = 2.5$$

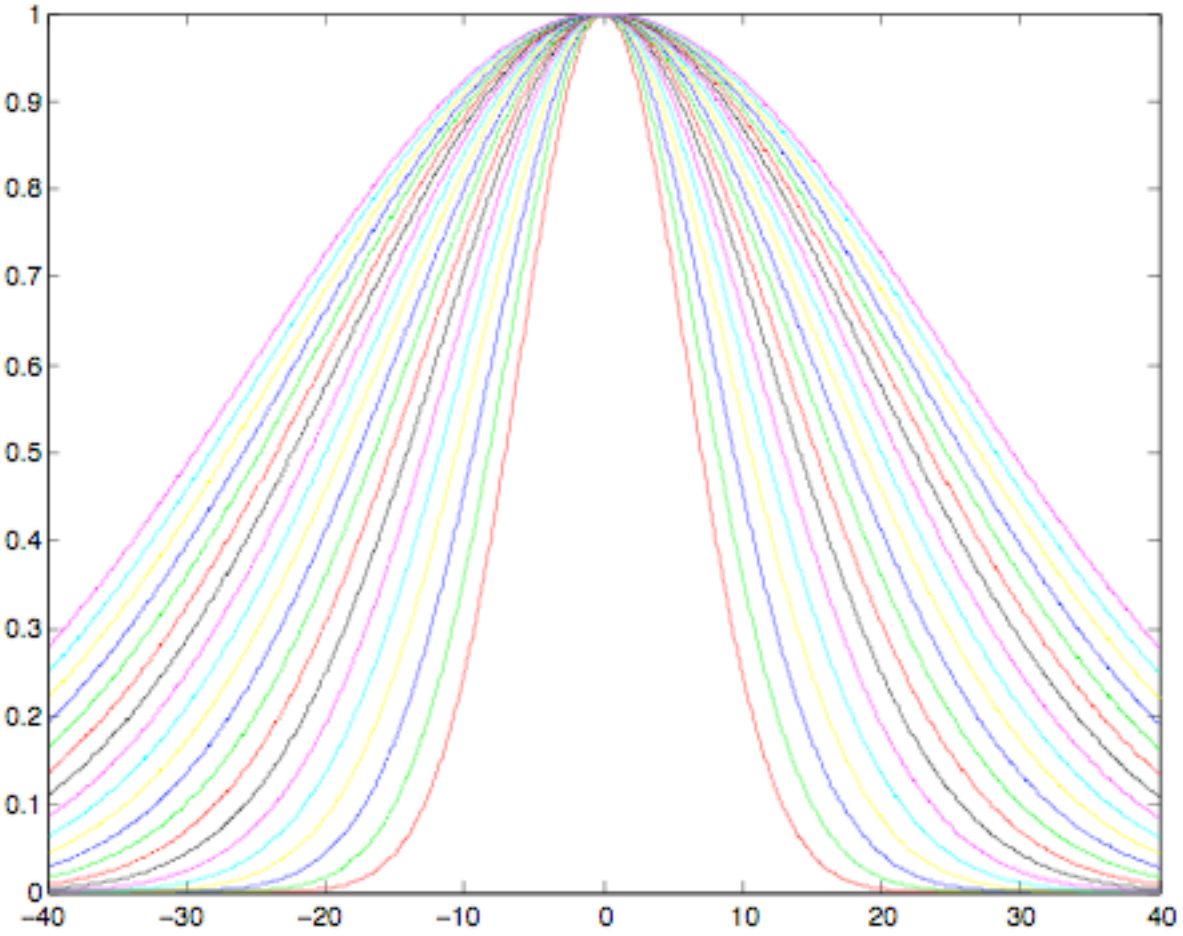
How to compute the median in matlab

- Function *median()*
- Example
- Matlab help: *help median*

How are they related?

- If you have a…
  - Normal distribution,
    - Mean=Median=Mode
  - Symmetric distribution
    - Median = Mean
  - Skew distribution
    - Median towards the body, mean towards the tail
      - +skew: mean>median
      - -skew: mean<median
- But this doesn't seem to be saying everything…

The mean isn't everything!
These all have the same mean



We need a measure of Variability

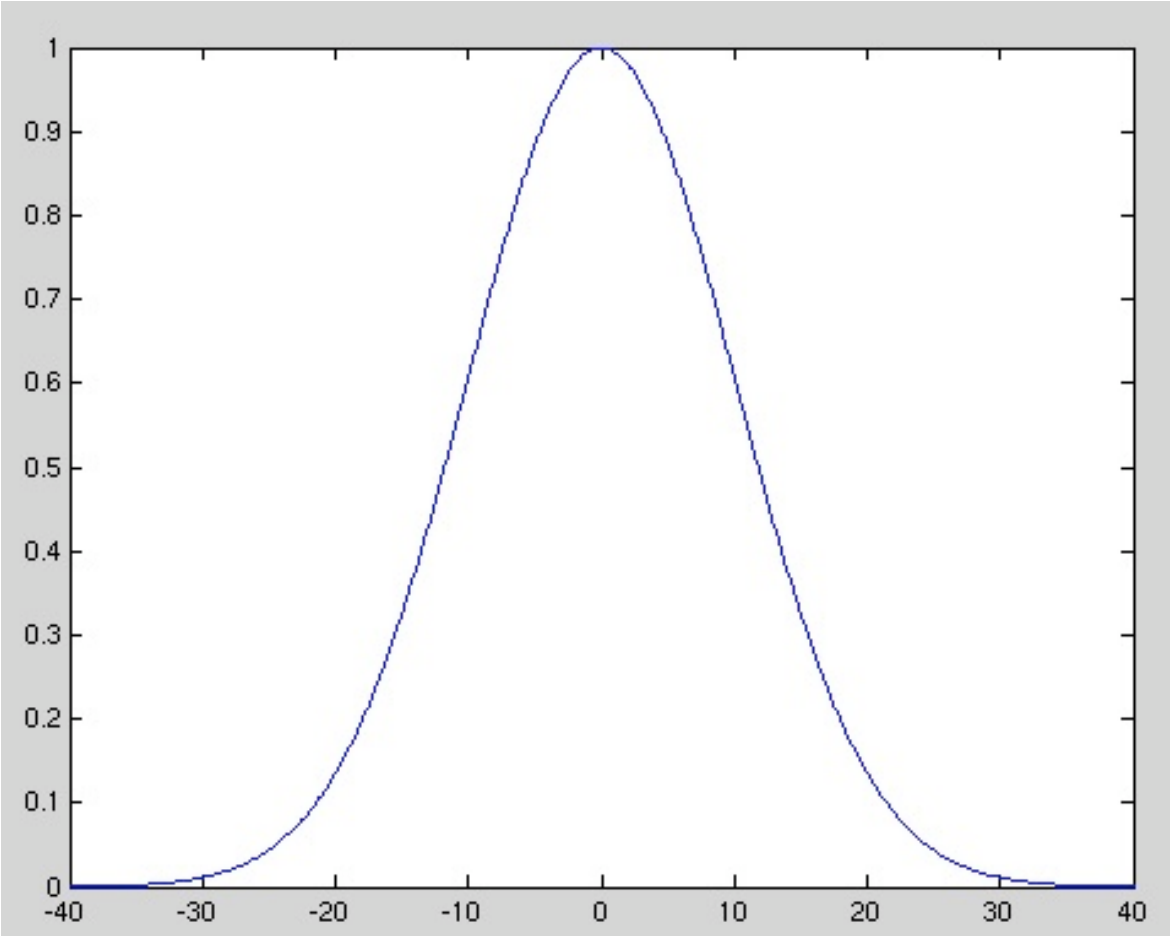
- **Range**
  - From math review, difference between max and min values of the data

$$\boxed{Range(x) = Max(x) - Min(x)}$$

- **Variance**
  - Mean of squared deviations from the mean

- **Standard deviation**
  - Square root of variance

Returning to the normal distribution…



Variance

- Computing the variance
  - Compute the deviations from the mean for all the data

$$d_i = \left(x_i - \bar{x}\right)$$

  - Compute the square of each of the deviations

$$sd_i = \left(d_i\right)^2$$

  - Sum up all these squared deviations

$$ssqd = \sum_{i=1}^{N} \left(sd_i\right)$$

  - Divide the mean squared deviations by N, the number of observations

$$Var = \frac{ssqd}{N}$$

How to compute the variance in matlab

- Function *var()*
- Example
- Matlab help: *help var*

Standard Deviation

- Typical ‘deviation’ from the mean
- Ie how far on average scores depart on either side from the mean
- Easy to compute after the variance - just take the square root of the variance

$$SD = \sqrt{Var} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

$$\bar{x} = \frac{\sum x_i}{N}$$

How to compute the standard deviation in matlab

- Function *std()*
- Example
- Matlab help: *help std*

Z scores

- A Z score is simply a measure of how many standard deviations away from the mean a score is
- Units are standard deviations

$$Z_i = \frac{X_i - \mu}{SD}$$

Covariance

- Covariance is very commonly used in statistical analysis as the basis for advanced statistics
- Gives a quantitative measure of the relationship between two variables

$$Cov(X,Y) = E\left[\left(X - \mu_x\right)\left(Y - \mu_y\right)\right]$$

$E = \exp ectation$

$\mu = mean$

More Covariance

- If the two variables are independent, the covariance is 0
  - (BUT IF COVARIANCE IS 0 THAT DOESN'T MEAN THE VARIABLES ARE INDEPENDENT!!!)
- If they are totally dependent the covariance of data, can be arbitrarily large
  - (AGAIN THE CONVERSE IS NOT NECESSARILY TRUE)
- The diagonals are the variance of each variable
- If each row is an observation, and each column a variable…

$$\operatorname{cov}(X) = \left(\frac{1}{N-1}\right)\left(X - mean(X)\right)^T\left(X - mean(X)\right)$$

Matlab does it easily with

- Function: cov(X) where X is a matrix with rows being observations, columns being variables
- Cov(X) where X is a vector yields the variance (a single scalar number)

As an aside: be careful about ‘sample’ vs. ‘population’ measures

- You can’t usually measure every possible subject or situation
  - Can you measure the height of every SINGLE individual in the United States?
    - Theoretically yes but it would take too long and too many resources
  - Measure a representative group which is large enough to minimize the bias due to the fact that it is only a portion of the total possible measurements you could make
  - Can make some mathematical adjustments
    - We won’t deal with this too much, since you learned about this in statistics, but you should know about the implications of each type of measure
- Matlab uses different equations to compute these statistics depending on you, but it has defaults of typically estimating populations

Trace

- Sum of the variances (the sum of the elements of the diagonal of the covariance matrix)