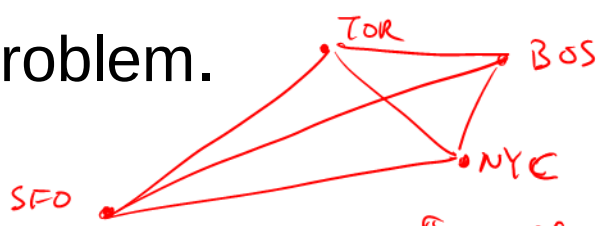
Permutation Variables and Traveling Salesman Problem

- Permutation– an ordered list of the numbers 1 to N. Hence a different order is a different value of the variable (e.g. (1 2 3) is different from (2 1 3)
- The classical permutation problem is the “traveling salesman” problem which tries to determine the least costly way to visit N cities (with each city having a number between 1 and N) given the cost to travel between any two cities
- You will see in the literature numerous reference to the “traveling salesman” problem and most permutation problems can be converted into a traveling salesman problem.



Handout 9-26-11

different costs for each inter-city hop.
find cheapest path

How many possible permutations are possible for N-length permutations?

- Imagine you have a permutation of length 4. for clarity assume we start with (1234) So the permutations would be:
- 1342, 1243, etc.

each digit here is a "position" which we can permute on

- (This is related to the third question on the homework due Friday.)

Computing number of Permutations Possible for length 4 vector

- A systematic way to look at this is to say you choose
- A) the first value is one of 4 numbers
- B) the second value is one of 4-1 values (since you can't use the value in A)
- C) the third value is one of 4-2 values (since you cannot use the value in A) or B)
- D) the last value is what ever is left.
- Hence the number of permutations in a string of length 4 is

4*3*2*1=24= 4! ≡ 4 factorial ??

What about for a vector of length N?

- First element is a choice of N numbers
- Second element is a choice of $N-1$ numbers
- Third element is a choice of $N-2$ numbers
- Etc.
- So a vector of length N has how many possible values $\prod_{i=0}^{N-1}(N-i) = \prod_{i=1}^{N}(N-i+1) = N!$??

$\Rightarrow N! = N(N-1)!$

$\Rightarrow 0! = 1$

- Is this a large number for N=3 or for N=10?

$3! = 6$ $\quad$ $10! = 3628800$

$\Rightarrow$ for a traveling salesman visiting 10 cities, he must consider $10! = 3628800$ permutations to find the optimum!

Pairwise Swapping

- The typical approach for creating neighborhoods with permutation variables is with pairwise swaps.
- Hence if you permutation is (1234), you pick two of the positions and swap the numbers in those locations.
- The pariwise swaps of (1234) include
  - (2134), 1324, ~~4123~~ 4231, etc.
  - Each of these picks two positions and swaps the numbers, e.g. ~~4123~~ 4231 involves picking positions one and 4 and swapping the numbers in those positions.

Possible number of Pairwise swaps

- For the permutation 1234 (where N=4) , how many pariwise swaps are there?
- You can pick from any of N=4 positions for the first member of the pair and you can pick from any one of the remaining locations (=4-1=N-1) for the second member of the pair
- This gives you a total of N*(N-1) ways to pick a first and second position for the swap.
- However the swap of (for example) the numbers in position 1 and 4 is the same as the swap of 4 and 1, so we need to divide by two so we don't double count.
- Hence the number of pairwise swaps is N choose 2 $\equiv \binom{N}{2}$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \Rightarrow \binom{N}{2} = \frac{N!}{2!(N-2)!} = \frac{N(N-1)}{2}$$

order does not matter!

Statistical Background

- As stated in the course description, students with no prior statistical background will need to do some reading in very basic (and very practical) statistics.
- A reading on basic statistics is available on Blackboard for our course.
- The following slides will review basic probability and statistics.
  - This is a review for students with prior background.
  - This is an introduction for students with no prior background and they will need to read the additional material.

Motivation: Statistical Comparison of Algorithms

- Consider the following table which shows the objective function values for the best solution in each trial for two algorithms applied to the same problem (minimization)

Algorithm1	45.3	77.2	68.5	42.4	42.4	47.7	26.9	46.6	97.9	122.6	**61.8**	**29.7**
Algorithm2	11.1	43.6	24.8	24.3	24.3	37.3	86.2	82.2	89.2	113.6	**53.7**	**35.7**

- Which algorithm is better? How do you measure this?

Statistical Tests (Hypothesis testing)

- Hypothesis tests are a formal statistical way of making a decision about data that exhibit variability.
- In our case, the performance criterion (the best Objective Function or the number of iterations required by an algorithm to reach a pre-specified low value) is usually variable between trials and between algorithms.

Statistical Assumptions

- The following assumptions are usually made in applying statistical tests:
  - The random variables (the performance criterion) are all identically distributed with the same shape and spread .
  - The random samples obtained from each trial of the algorithm are independent

- I will give some introduction so you understand the meaning of these Assumptions

Introduction To Random Variables

- *A random variable* $X(s)$ is a real-valued function which assigns a real number $X(s) = x$ to every sample point $s \in S$

- In our examples the random variable can be :
  - The best objective function found in a trial
  - The number of objective function evaluations to come within some percentage of the optimal value

Basic Terms in Probability

- **Experiment:** a procedure that generates a sample point *x* in the sample space according to some probabilistic law.

- **Examples:**

1. Experiment rolling a die once. fair die $\Rightarrow P(1, \ldots, n) = 1/n$ (# of sides)
2. Experiment counting the number of students in a single row, 5 minutes after class starts.

$\Rightarrow$ likely correlated to other factors such as homework due dates

Basic Terms in Probability

- **Event, E** : a subset of $\mathcal{S}$ – any collection of outcomes of an
- experiment

- **Examples:** six-sided die (1d6)

1. Experiment rolling a die once:
    1. Event A = 'score < 4' = {1, 2, 3} → $\frac{1}{6} \times 3 = \frac{1}{2}$
    2. Event B = 'score is even' = {2, 4, 6} $\frac{1}{6} \times 3 = \frac{1}{2}$
    3. Event C = 'score = 5' = {5} $\frac{1}{6}$
2. Experiment counting the number of students in a
- single row with 12 seats per row:
    - Event A = 'all seats are taken' = {12}
    - Event B = 'no seats are taken' = {0}
    - Event C = '< than 6 seats are taken' = {0, 1, 2, 3, 4, 5}

Probabilistic Independence

- **Independence:** A and B are independent if one event *A* occurring has no impact on the probability that another event *B* will or will not occur.
- If events *A* and *B* are independent then the probability they both occur is the product of the probability that each occurs:

intersection product

- $$P(A \cap B) = P(A)\,P(B)$$

Probabilistic Independence (continued)

- **Example:** (Let Hi be event that you get head in i^{th} flip)

1. The probability of flipping a fair coin and getting heads twice in sequence with independent tosses is:

- $P(H1 \cap H2) = P(H1)\,P(H2) = (0.5)(0.5) = 0.25$

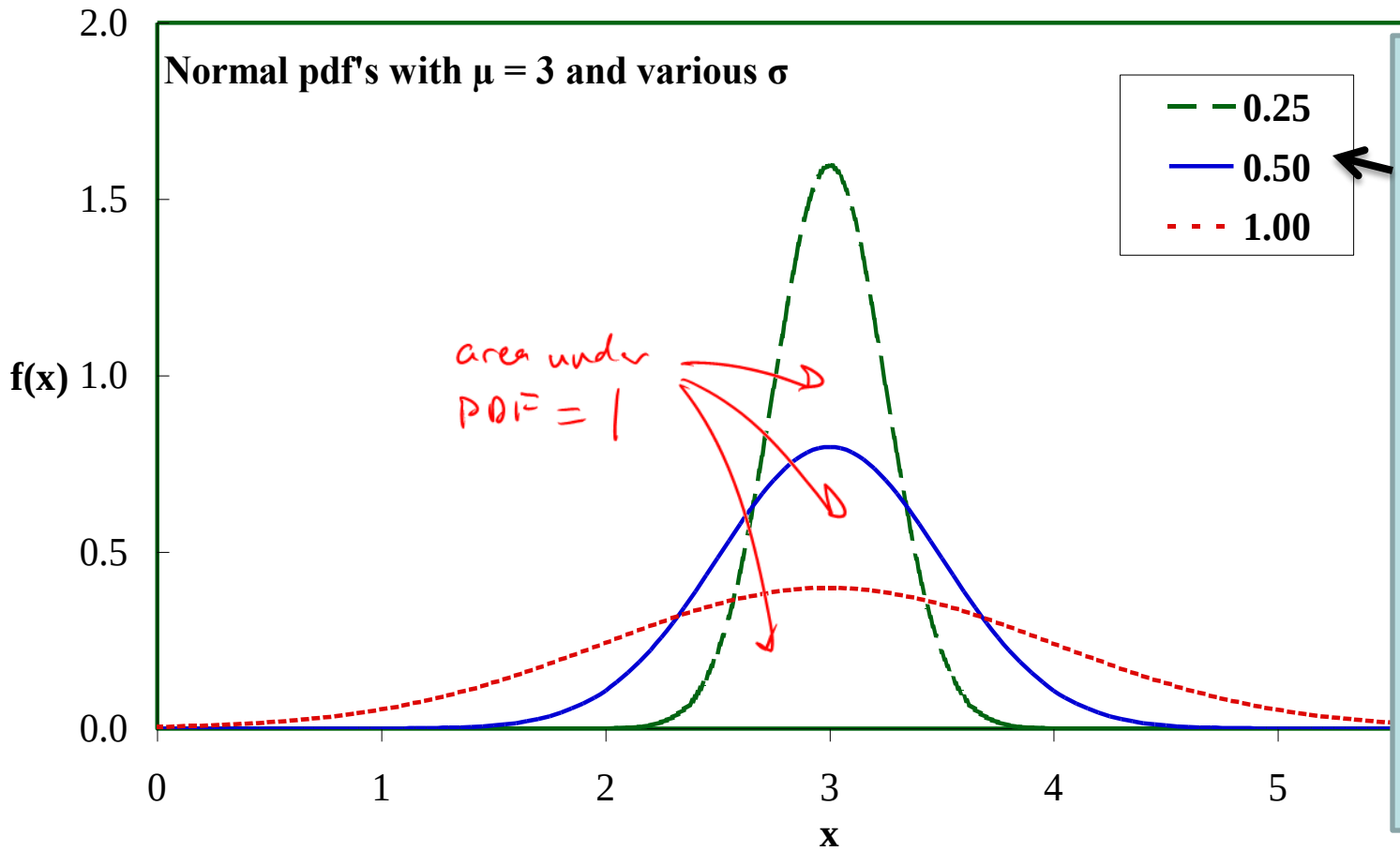
- 2. The probability of flipping a fair coin and getting heads three times in sequence with independent tosses is:
  - $P(H1 \cap H2 \cap H3) = P(H1 \cap H2)\,P(H3) = (0.25)(0.5) = 0.125$

Probability Density Function ⇒

aka Gaussian or "Bell" Curve

PDF of Normal Distribution

Here are three examples of normal distribution with different variances

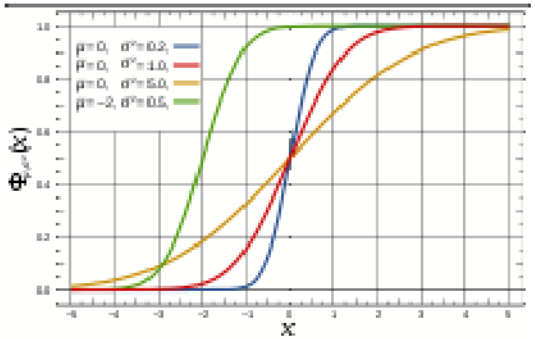


Difference among these curves is the value of sigma, which is the "standard deviation"

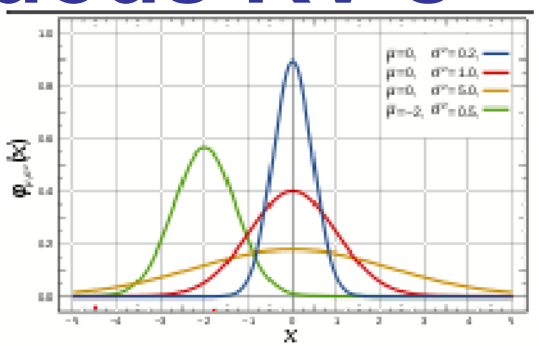
Sigma squared=variance; area under curve is 1 for all pdfs

Describing Continuous RV's

from wikipedia



from wikipedia



CDF=

Cumulative Distribution Function

$$F(x) = \int_{s=-\infty}^{x} f(s)\,ds$$

PDF

monotonically increases to 1

The CDF F(x) is the probability that the value of the Random Variable s is less than x.

Describing the Average

- **Expected value of a random variable:**
- mean =

value of X

$$\mu = E[X] = \int_{-\infty}^{+\infty} s\, f(s)\, ds$$

probability of that value

- More generally for any function *h(X)*, one can compute its expected value equal to its average value in a large number of trials as:

expected value of function h with input X.

$$E[h(X)] = \int_{-\infty}^{+\infty} h(s)\, f(s)\, ds$$

- **Useful property of expectations:**

- $E[a + bX] = a + bE[X]$

Describing Variability

- Assume one set of best objective functions (samples) is: {1,4,8,3,7,1}
- And a second set is: {3,4,5,3,4,5}

- Both sets have a mean of 4.

- Which set of samples is more variable? How do we measure variability?

Describing Variability

- **Mean, E[X]:** Measure of central tendency; center of mass.
- **Variance**: Measure of dispersion, variability, uncertainty, or imprecision = Var[X] = σ^2

$$\sigma^2 = E\left[\left(X-\mu\right)^2\right] = \cancel{\int_{-\infty}^{+\infty} (s-\mu)^2 f(s)\,ds} = \int_{-\infty}^{\infty} (s-\mu)^2 f(s)\,ds$$

- Another definition: σ = **Standard Deviation**

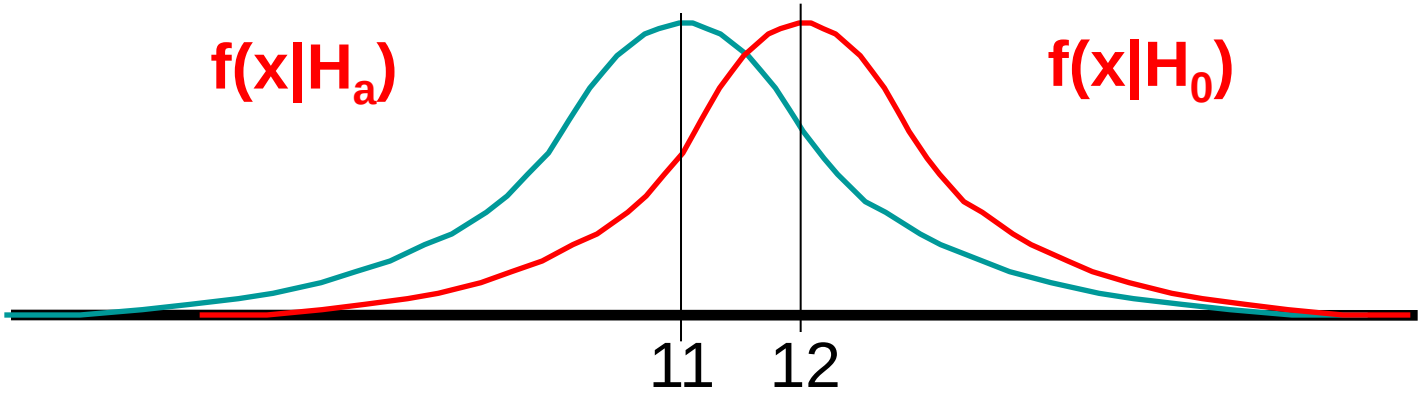
- **Computation of the Variance** (a useful "shortcut" formula):
- $\sigma^2 = E\{ [X-\mu]^2 \} = E\{ X^2 \} - \mu^2$

Hypothesis Testing: Introduction

- **How to make a decision with data that exhibit variability.**

- **Examples**
  - We have a robot that is supposed to pick up an object and move it to a spot 12 meters away.
  - We have done some trials and distances vary from trial to trial.
  - In the trials the placement is only 11 meters away on average.

Hypothesis Testing: Example

- The distances in different trials can come from two possible distributions:
  - Target, or Null Hypothesis: $X \sim N[12.0, 1]$
  - Alternative Hypothesis: $X \sim N[11.0, 1]$
- Or, we may say that $X \sim N[\mu, \sigma^2]$ where either
  - State #1, H_0: $\mu = 12$ ← robot is performing correctly, just sloppy
  - State #2, H_a: $\mu = 11$ ← robot is miscalibrated to 11 meters
- 

Note: $f(x|A)$ means the probability of x occurring if A is true

Hypothesis Testing: Example

- Run n trials and then decide which is true.

- Accept H_0: $\mu = 12$ if $> c_x$
- Accept H_a: $\mu = 11$ if $\leq c_x$ (This is the **rejection region** for H_0)

- c_x = critical x-value for test, a cut-off value chosen with the aim of making both α and β unlikely.

Type I error:

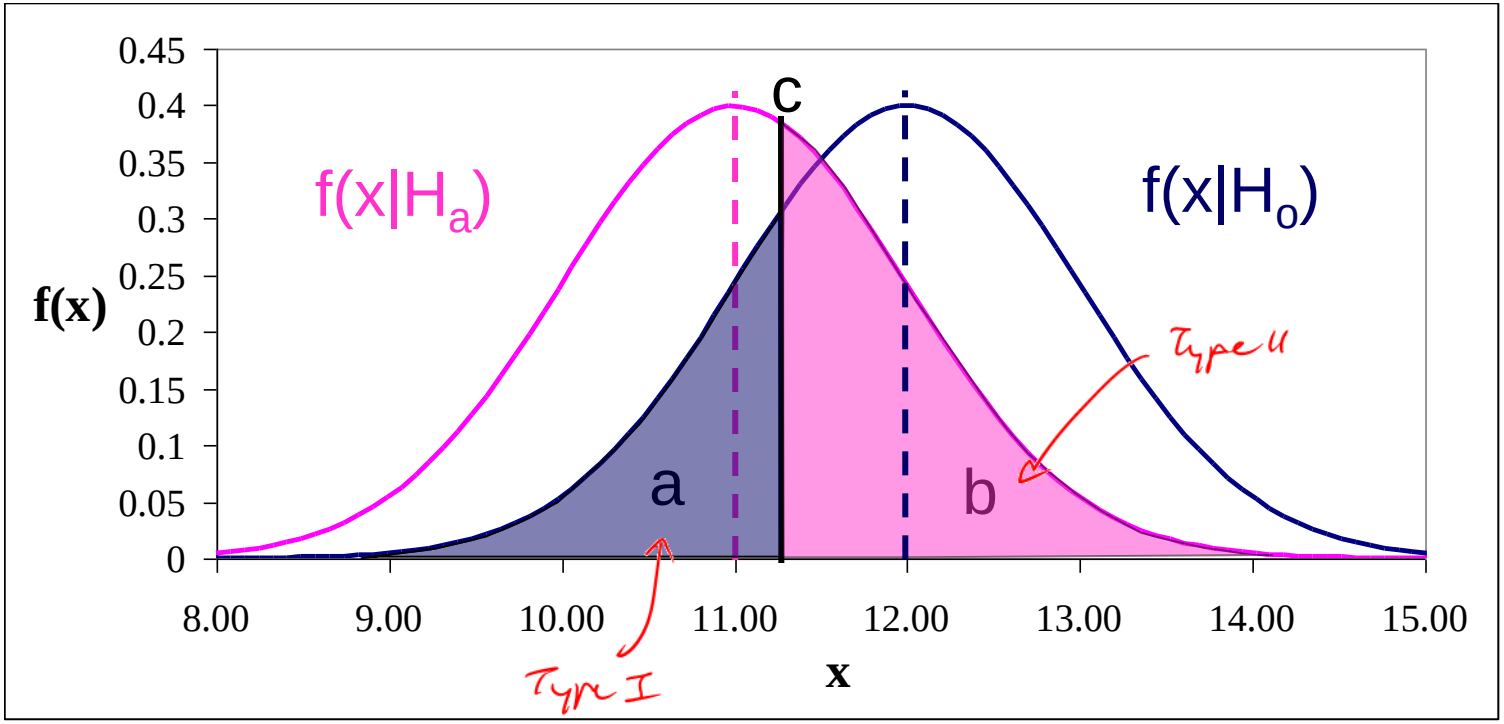
- α = P[Reject H_0 | H_0 true] $= \int_{-\infty}^{c_x} f(x \mid H_0)dx = \Phi\left[\sqrt{n}(c_x - 12)/1\right]$

Type II error:

β = P[Accept H_0 | H_0 false] $= \int_{c_x}^{\infty} f(x \mid H_a)dx = 1 - \Phi\left[\sqrt{n}(c_x - 11)/1\right]$

End of handout 9-23-11

Type I & II Errors



Type I error:

α = P[Reject H_0 | H_0 true]

Type II error:

β = P[Accept H_0 | H_0 false]

	H_o true	H_a true
Accept H_o	$1-\alpha$ Correct True Negative	β Type II error False Negative
Accept H_a	α Type I error False Positive	$1-\beta$ Correct True Positive

Hypothesis Testing: Details

- **Formal testing procedure**
- Test statistic: $T_{n-1} = \dfrac{\bar{X} - \mu}{S / \sqrt{n}}$ appropriate when σ is unknown
- T is dimensionless and allows many problems to be formulated in a common framework.
- If Ho is true, then T_{n-1} ~ (Student) *t*-distribution with v= n-1
- **Choice of Hypothesis**
- "Statistical tests are predisposed to accept Ho. A test is only effective if
- one collects sufficient data to **reject the null hypothesis**."
- Upon which hypothesis should the burden of proof be placed?

Hypothesis Testing: Details (1 of 2)

- **Decision Rules**
- The null hypothesis is Ho: $\mu = \mu_o$
- The test statistic value is $$t = \frac{\bar{x} - \mu_o}{s/\sqrt{n}} = \sqrt{n}(\bar{x} - \mu_o)/s$$
- We construct a rejection region such that the Type I error probability is
- controlled to a desired level, i.e., we select an α.

If the **alternative hypothesis** is:	Then the **rejection region** for a level a test is:
$H_a = \mu > \mu_0$	$t \geq t_{\alpha,\nu}$
$H_a = \mu < \mu_0$	$t \leq t_{\alpha,\nu}$
$H_a = \mu \neq \mu_0$	$\lvert t \rvert \geq t_{\alpha/2,\nu}$

- If H_a is true, then the type II error β can be computed using the type I error
- α, the degrees of freedom ν, and the standardized distance

Hypothesis Testing: Example

- On a national test the average is 75. I think Cornell students are smarter, so we randomly select 7 Cornell students and they take the test.

- **Results**:
- $\bar{x} = 81.3$ $s_x = 6.83$ $n = 7$
- Null Hypothesis: Ho: $\mu = 75$
- Alternative Hypothesis: Ha: $\mu > 75$
- Compute:

$$t = \sqrt{n}(\bar{x} - \mu_o)/s = \sqrt{7}\,(81.3 - 75)/6.83 = 2.45$$

-
- Use $\alpha = 1\%$ => $t_{0.01,6} = 3.143$
- Because $t < t_{\alpha,\nu}$, we should not reject the Null Hypothesis.
- What if one used $\alpha = 2.5\%$, 5% or 10% ?
- What if we took a larger sample?