AST 418/518
Instrumentation and Statistics

Class Website:

http://ircamera.as.arizona.edu/Astr_518

Class Texts:

Practical Statistics for Astronomers, J.V. Wall, and C.R. Jenkins, Second Edition.

Measuring the Universe, G.H. Rieke

Statistics and Instrumentation

Statistical Methods for Astronomy

Probability (Lecture 1)

Read: W&J Ch. 1-2.5

- Why do we need statistics?
- Definitions
- Probability distributions
    - Binomial Distribution
    - Poisson Distribution
    - Gaussian Distribution
- Bayes' Theorem
- Central Limit theorem

Statistics (Lecture 2)

Read: W&J Ch. 3

- Useful Statistics
- Error Analysis
    - Error Propagation
- Least Squares
    - chi-squared
    - Significance
- Comparison Statistics

Practical Statistics

- Lecture 3 (Aug. 30)

  Read: W&J Ch. 4-5

  - Correlation

  - Hypothesis Testing

- Lecture 4 (Sep. 1)

  - Principle Component Analysis

- Lecture 5 (Sep. 6):

  Read: W&J Ch. 6

  - Parameter Estimation

  - Bayesian Analysis

  - Rejecting Outliers

  - Bootstrap + Jack-knife

- Lecture 6 (Sep. 8)

  Read: W&J Ch. 7

  - Random Numbers

  - Monte Carlo Modeling

- Lecture 7 (Sep. 13):

  - Markov Chain MC

- Lecture 8 (Sep. 15):

  Read: W&J Ch. 9

  - Fourier Techniques

  - Filtering

  - Unevenly Sampled Data

What use are statistical methods?

- Help you make a decision!

    Is a signal in a set of observations meaningful?

    Do the data fit our model of the phenomenon under study?

- Simulate Observations

    Plan size of sample, etc.

    What would happen if we repeated the observations?

- Compare different observations

    Are two sets of data consistent with each other?

    Are the observations truly independent?

Useful References

- *Data Reduction and Error Analysis*, Bevington and Robinson

  - Good basic introduction to statistics

- *Practical Statistics for Astronomers*, Wall and Jenkins

- *Numerical Recipes*, Press et al.

  - The bible for numerical computation.

- "*Understanding Data Better with Bayesian and Global Statistical Methods*", Press, 1996 (on astro-ph)

Process of Decision Making

Ask a Question

Take Data

Reduce Data

Derive Statistics describing data

Reflect on what is needed

Probability Distribution

Error Analysis

Hypothesis Testing

Simulation

Does the Statistic answer your question?

No

Yes

Publish!

7
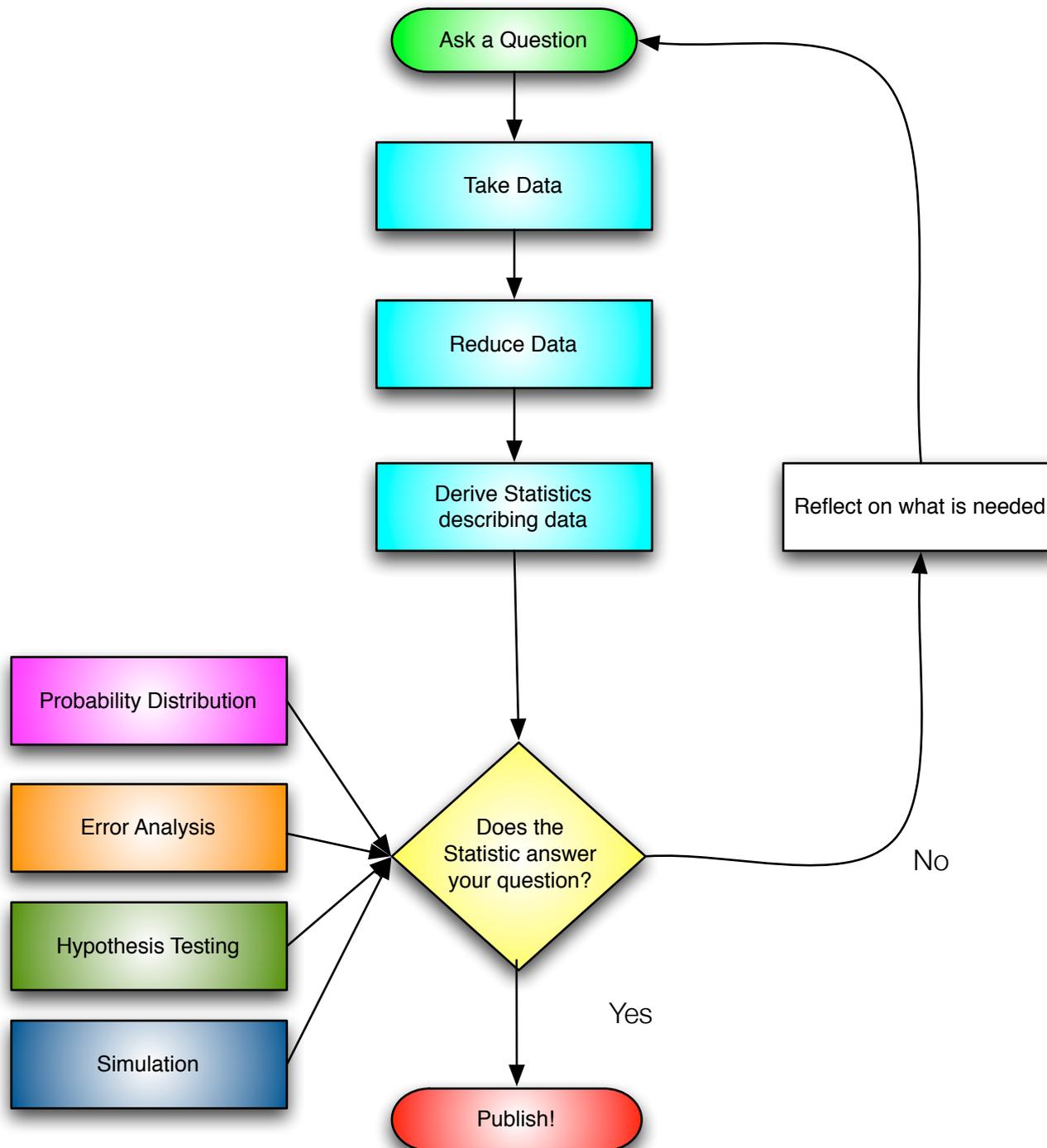
Some definitions

Statistic – a number or set of numbers that describe a set of data.

Probability distribution – the relative chances for different outcomes for your set of data.

Sample distribution – Set of data that allow us to estimate useful values of the object under study.

Parent distribution – Presumed probability distribution of the data that one would measure if an infinite data set were acquired.

Mean - the 1st moment of a distribution, which gives information about the most likely value one will observe.

Variance – the 2nd moment of a distribution, which gives information about the range of values one will observe.

Typical Statistics

If we have n data points, we typically want to know things like the value of a data point, or how much the variation in the data is:

Mean

$$m = \sum_j \frac{x_j}{n}$$

Variance

$$s^2 = \sum_j \frac{(x_j - m)^2}{n - 1}$$

More on this next lecture.

Probability Distributions

Assumptions

Principle of Indifference:

> The system under study has known relative chances of arriving at a particular state. The state of the system is arrived at by random chance.

Independence:

> For two observations, the result of one outcome is not influenced by the other ->
>
> > P(A and B) = P(A)P(B)

Probability Distributions

- Probability distributions (P(x)) can be any strange function (or non-analytical curve) that can be imagined as long as

$$prob(a < x < b) = \int_a^b P(x)dx$$

$$\int_{-\infty}^{\infty} P(x)dx = 1$$

P(x) is a single, non-negative value for all real x.

Mean and Variance of Probability Distributions

Mean	Variance

Discrete: $\quad \mu = \displaystyle\sum_j n_j P(n_j) \qquad\qquad \sigma^2 = \displaystyle\sum_j (n_j - \mu)^2 P(n_j) = \sum_j (n_j)^2 P(n_j) - \mu^2$

Continuous: $\quad \mu = \displaystyle\int x P(x)\,dx \qquad\qquad \sigma^2 = \displaystyle\int (x-\mu)^2 P(x)\,dx = \int x^2 P(x)\,dx - \mu^2$

$$\sigma^2 = \langle \overline{x^2} \rangle - \overline{\langle x \rangle^2}$$

The Binomial distribution

- You are observing something that has a probability, p, of occurring in a single observation.

- You observe it M times.

- Want chance of obtaining n successes. For one, particular sequence of observations the probability is:

$$P_1(n) = p^n(1-p)^{M-n}$$

- There are many sequences which yield n successes:

$$P(n) = \frac{M!}{n!(M-n)!} p^n(1-p)^{M-n} = \begin{pmatrix} M \\ n \end{pmatrix} p^n(1-p)^{M-n}$$

Mean	Variance
Mp	Mp(1-p)

Often said
"M choose n"

14

The Poisson Distribution

- Consider the binomial case where p → 0, but Mp → μ. The binomial distribution, then becomes:

$$P(n) = \mu^n \frac{e^{-\mu}}{n!}$$

Mean	Variance
Mp=μ	Mp(1-p)~Mp=μ

Gaussian Distribution

- The limiting case of the Poisson distribution for large μ is the Gaussian, or normal distribution

$$P(x)dx = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx$$

Mean	Variance
μ	σ^2

- Large μ Poisson distributions are Gaussian with $\sigma^2 = \mu$.

- In general, Gaussian distributions can have unrelated mean and variance values.

Gaussian Distribution

The Gaussian distribution is often used (sometimes incorrectly) to express confidence.

$$P(|x < \mu + \sigma|) > 0.68$$

$$P(|x < \mu + 2\sigma|) > 0.95$$

$$P(|x < \mu + 3\sigma|) > 0.997$$
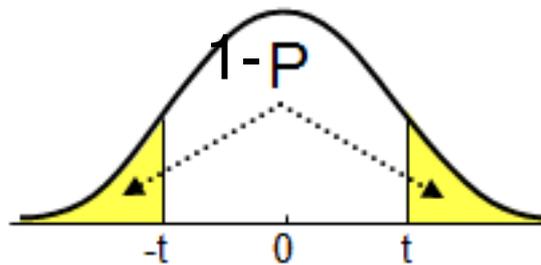
Mean and Variance of Distributions

Distribution	Mean	Variance
Binomial	Mp	Mp(1-p)
Poisson	μ	μ
Gaussian	μ	σ^2
Uniform [a,b)	(a+b)/2	(b-a)/12

Frequentist vs. Bayesian Statistics

Two approaches to discussing the problem:

- Knowing the distribution allows us to predict what we will observe.
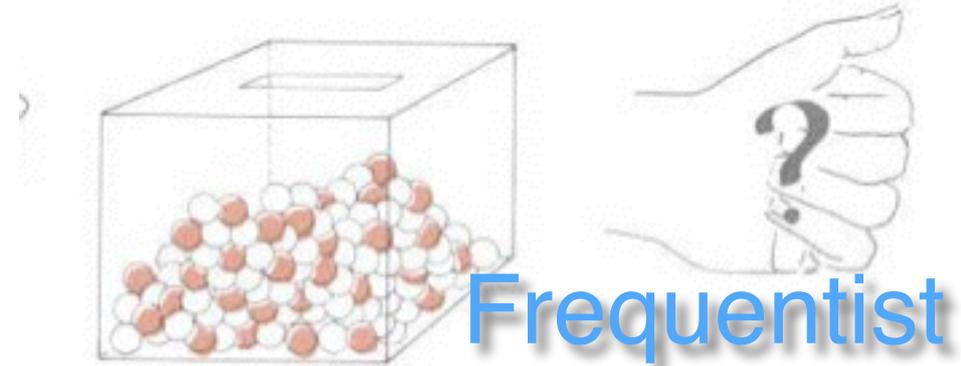
Frequentist

- We often know what we have observed and want to determine what that tells us about the distribution.

Bayesian
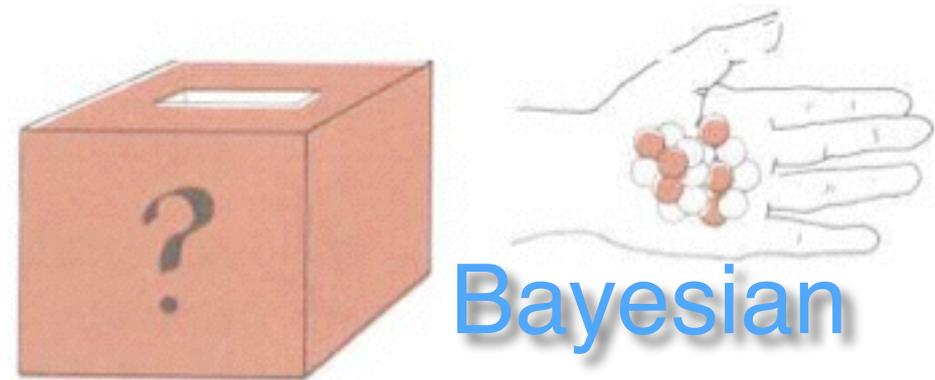
Frequentist Approach
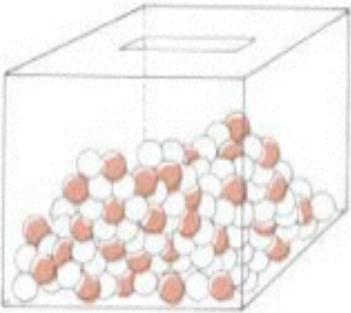
"I hypothesize that there are an equal number of red and white balls in a box."
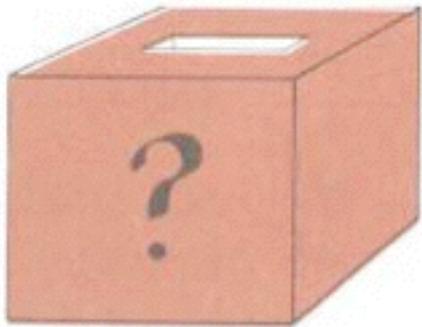
"I see I have drawn 6 red balls out of 10 total trials."

Prediction
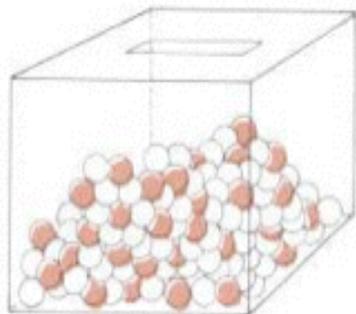
"A box with equal number of balls will have a mean of 5 red balls with a standard deviation of 1.6."

"Based on this I cannot reject my original hypothesis."
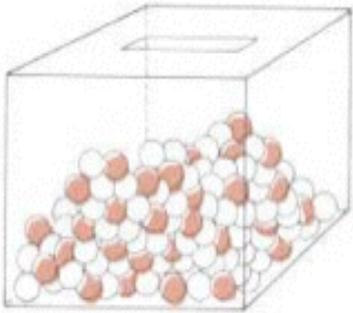
21

Bayesian Approach

"I hypothesize that there are an equal number of red and white balls in a box."
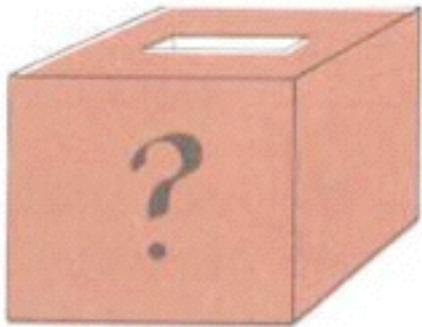
"I see I have drawn 6 red balls out of 10 total trials."

"Odds" on what is in the box.

"There is a 24% chance that my hypothesis is correct."
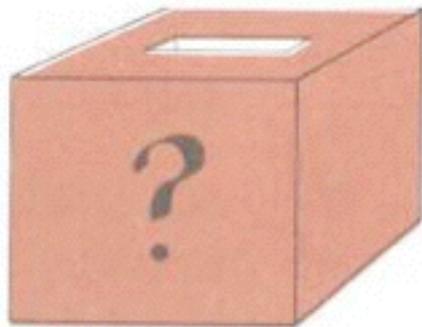
Approaches to Statistics

- "Frequentist" approaches will calculate statistics that a given distribution would have produced, and confirms or rejects a hypothesis.

  - These are computationally easy, but often solve the inverse of the problem we want.

  - Locked into a distribution (typically Gaussian)

- Bayesian approaches use both the data and any "prior" information to develop a "posterior" distribution.

  Allows calculation of parameter uncertainty more directly.

  More easily incorporates outside information.

Conditional Probability

If two events, A and B, are related, then if we know B the probability of A happening is:

$$P(A|B) = \frac{P(A \, and \, B)}{P(B)}$$

Reversing the events, we get:

$$P(B|A) = \frac{P(B \, and \, A)}{P(A)}$$

P(B | A) should be read as "probability of B given A"

Now, P(A and B) = P (B and A) which gives us the important equality:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

This is Bayes' Formula.

Bayes' Theorem

- Bayes' formula is used to merge data with prior information.

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

- A is typically the data, B the statistic we want to know.

- P(B) is the "prior" information we may know about the experiment.

- P(data) is just a normalization constant

$$P(B|data) \propto P(data|B) \cdot P(B)$$

--- prior
— likelihood
····· posterior

0   5   10   15   20
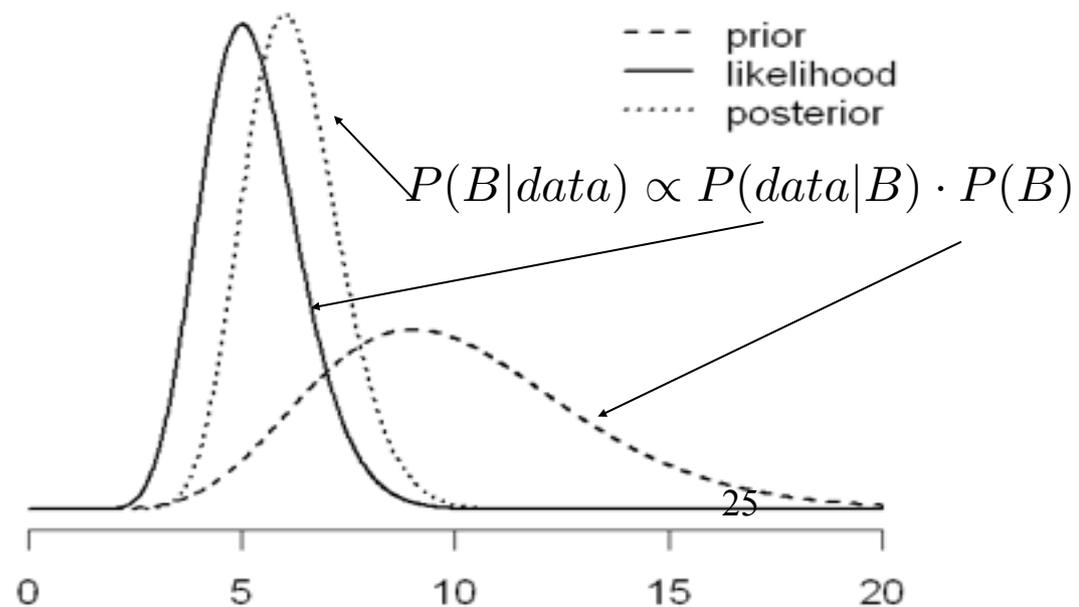
Example of Bayes' Theorem

A game show host invites you to choose one of three doors for a chance to win a car (behind one) or a goat (behind the other two).

After you choose a door (say, door 1), the host opens another door (say, door 3) to reveal a goat. Should you switch your choice?
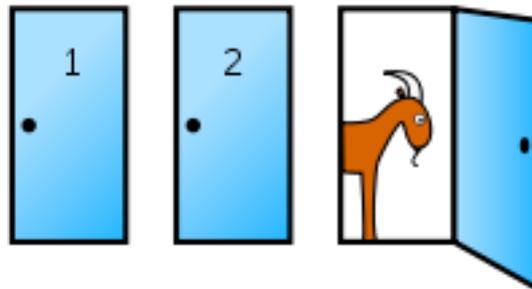
Using Bayes' theorem

- Assume we are looking for faint companions, and expect them to be around 1% of the stars we observe.

$$P(planet) = 0.01$$
$$P(noplanet) = 0.99$$

- From putting in fake companions we know that we can detect objects in the data 90% of the time.

$$P(+det.|planet) = 0.9$$
$$P(-det.|planet) = 0.1$$

- From the same tests, we know that we see "false" planets 3% of the observations.

$$P(+det.|noplanet) = 0.03$$

- What is the probability that an object we see is actually a planet?

$$P(planet| + det.) = \frac{P(+det|planet)P(planet)}{P(+det)}$$

$$P(+det.) = P(+det|planet)P(planet) + P(+det|noplanet)P(noplanet)$$

$$P(planet| + det.) = \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.03 \cdot 0.99} = 0.23$$

General Bayesian Guidance

- Focuses on probability rather than accept/reject.

- Bayesian approaches allow you to calculate probabilities the parameters have a range of values in a more straightforward way.

- A common concern about Bayesian statistics is that it is subjective. This is not necessarily a problem.

- Bayesian techniques are generally more computationally intensive, but this is rarely a drawback for modern computers.

Why are Gaussian statistics so pervasive?

- Even an unusual probability distribution will converge to a Gaussian distribution, for a large enough number, N, of samplings.

- Referred to as the "Central Limit Theorem"

NonNormal Distribution of X

From statisticalengineering.com
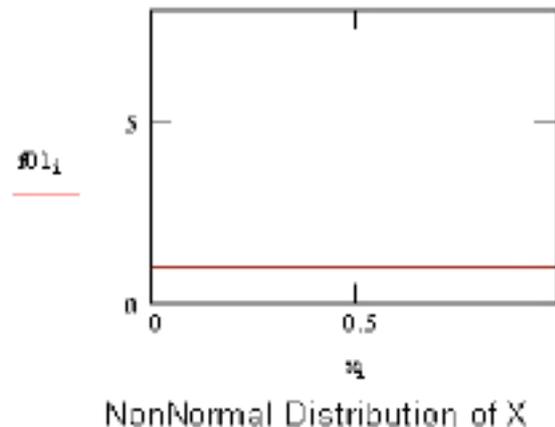
Odd Distributions

- This works for any unusual distributions that an individual random number may be drawn from:
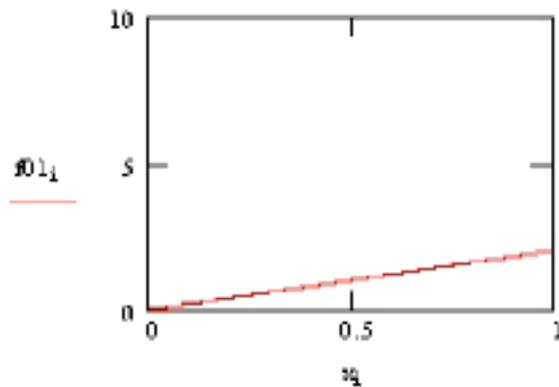
NonNormal Distribution of X
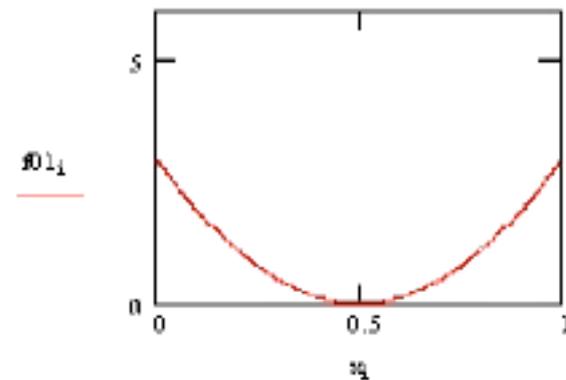
NonNormal Distribution of X

Thursday:

Read Wall and Jenkins Ch. 2-3.