Applied Computational & Numerical Methods

Xinchou Lou

Chinese Academy of Sciences School of Graduate Studies

&

the Institute of High Energy Physics

Summer, 2011

Applied Computational & Numerical Methods

Lecture I

- ❑ Course Syllabus & Requirement
- ❑ C++ tutorials
- ❑ Weekly Computing Projects
- ❑ Motivations for this course
- ❑ Probability: concepts, rules, distributions
- ❑ Limitations of the Gaussian approximation

SYLLABUS and REQUIREMENTS

Applied Computational & Numerical Methods

Professor Xinchou Lou

<u>Objectives</u>

To learn and apply computational techniques to analyze data and to solve scientific problems numerically in most computing environments by using the ROOT program, or other programming and visualization tools.

SYLLABUS and REQUIREMENTS

Applied Computational & Numerical Methods

Course Details

(1) Lecture will be in English. Students can use English or Chinese in the class.

(2) The programming language for weekly labs/projects is C++. Familiarity with C++ is very useful, but not required if you are willing to learn the basics of C++.

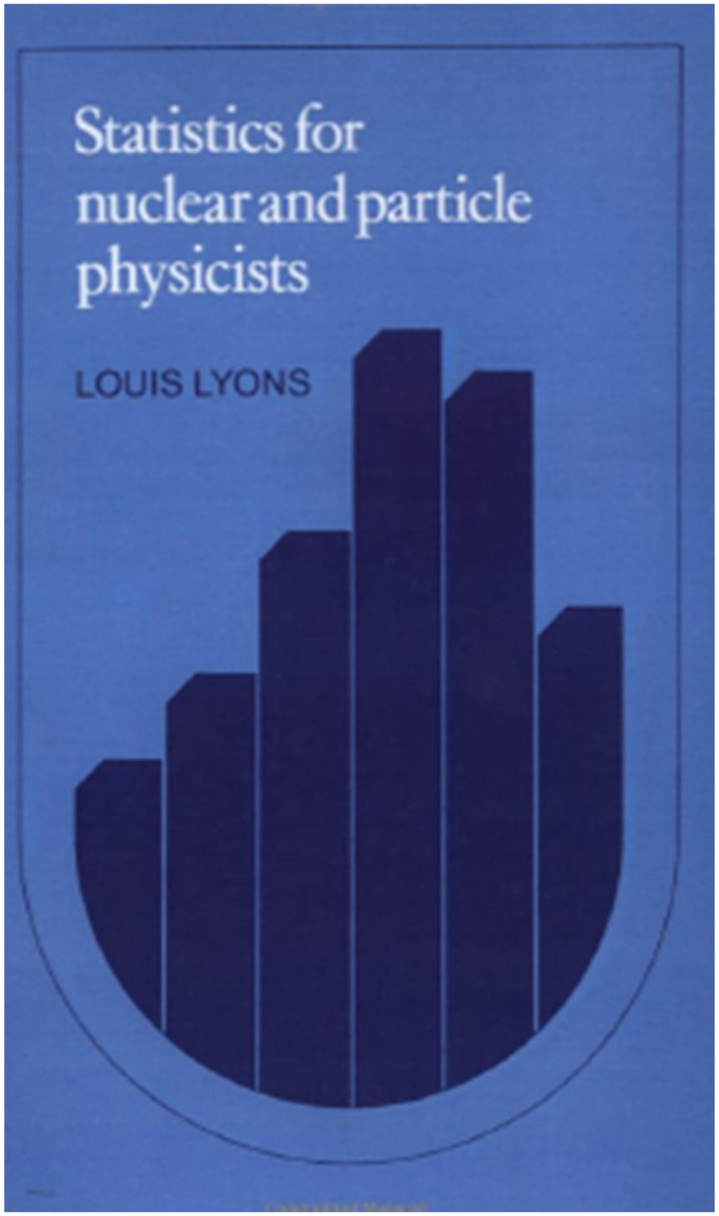
(3) Each of the weekly computing projects is expected to be completed in one week for best effect.

SYLLABUS and REQUIREMENTS

Applied Computational & Numerical Methods

Reference Books & Material (Not required. Available in my office for browsing)

Numerical Methods for Physics
A. L. Garcia, ISBN 0-13-906744-2, Prentice Hall, Inc.
Numerical Methods for Scientists and Engineers, R.W. Hamming
Statistics for Nuclear and Particle Physicists
by Louis Lyons, Cambridge University Press, ISBN 0 521 37934 2
A Course in Probability and Statistics, Charles J. Stone
Numerical Recipes in C
William H. Press *et al*., Cambridge Univ. Press
(available online at http://www.nrbook.com/a/bookcpdf.php)
Online C++ tutorial: http://www.cplusplus.com/doc/tutorial/



Experimental Errors

- meaning of errors
- combinations of errors
- random & systematic errors

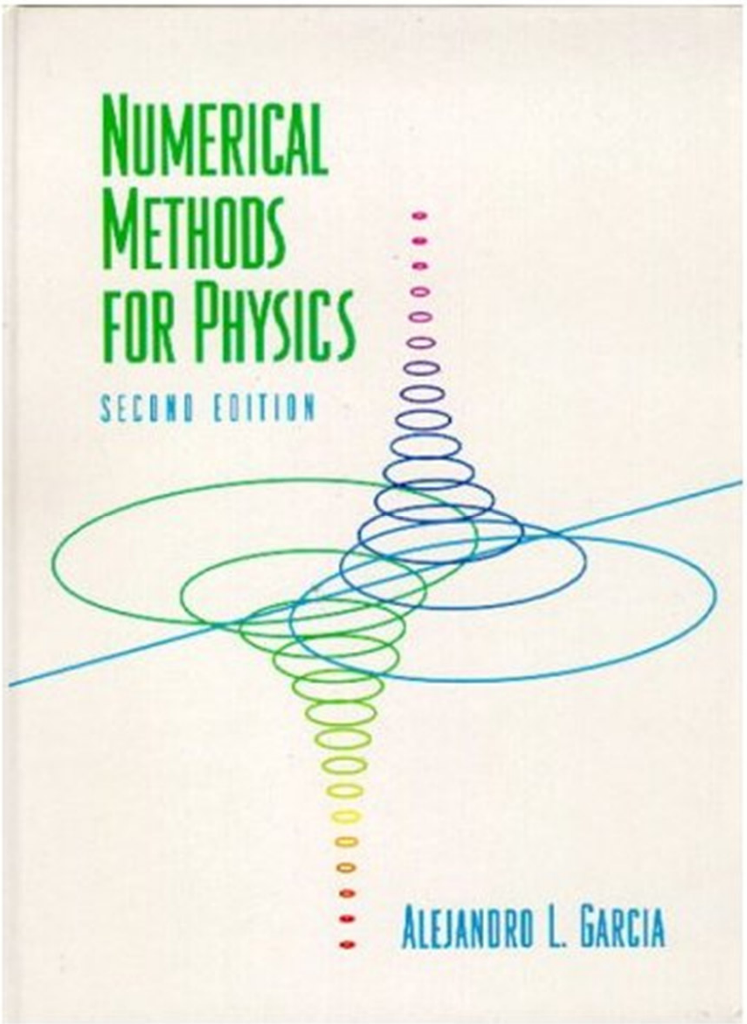
Probability and Statistics

- Rules of Probability
- P&S
- Binomial, Poisson, Gaussian
- Error Matrix
- Correlations

Parameter fitting & Hypothesis testing

- normalization
- error estimate
- interpretation of error
- upper & lower limits
- Maximum likelihood
- Least squares
- Minimization

ISBN: 0521379342

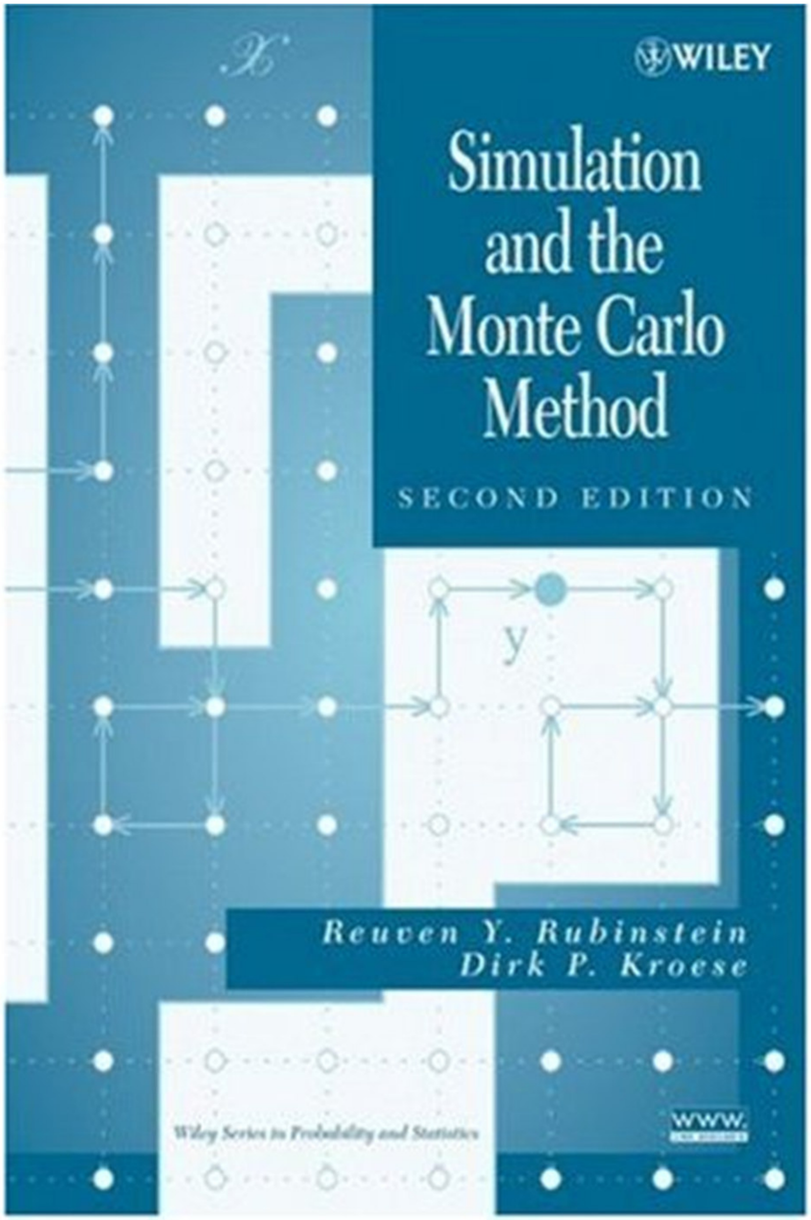


Garcia, Alejandro,
Numerical Methods for Physics
Second Edition

Partial Differential Equations

- Diffusion Equations
- Advection Equations
- Stability Analysis
- Examples

ISBN-10: 0139067442



Random Number Generation

- Chi-square Goodness-fit Test
- Kolmogorov-Smirnov Test
- Serial Test
- Gap Test, Maximum Test

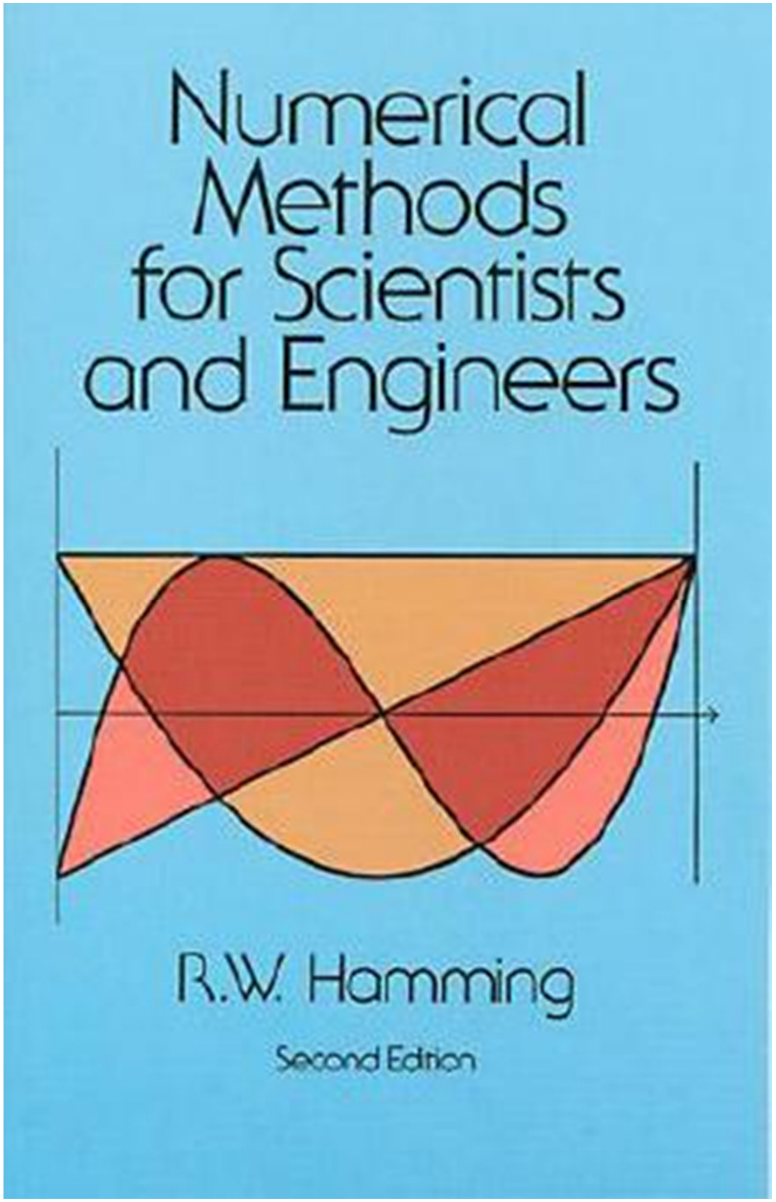
Random Variate Generation

- Inverse Transform Method
- Composition Method
- Acceptance-Rejection Method
- Examples of RV Generations

Monte Carlo Simulation

- Applications and Examples

ISBN-10: 0470177942



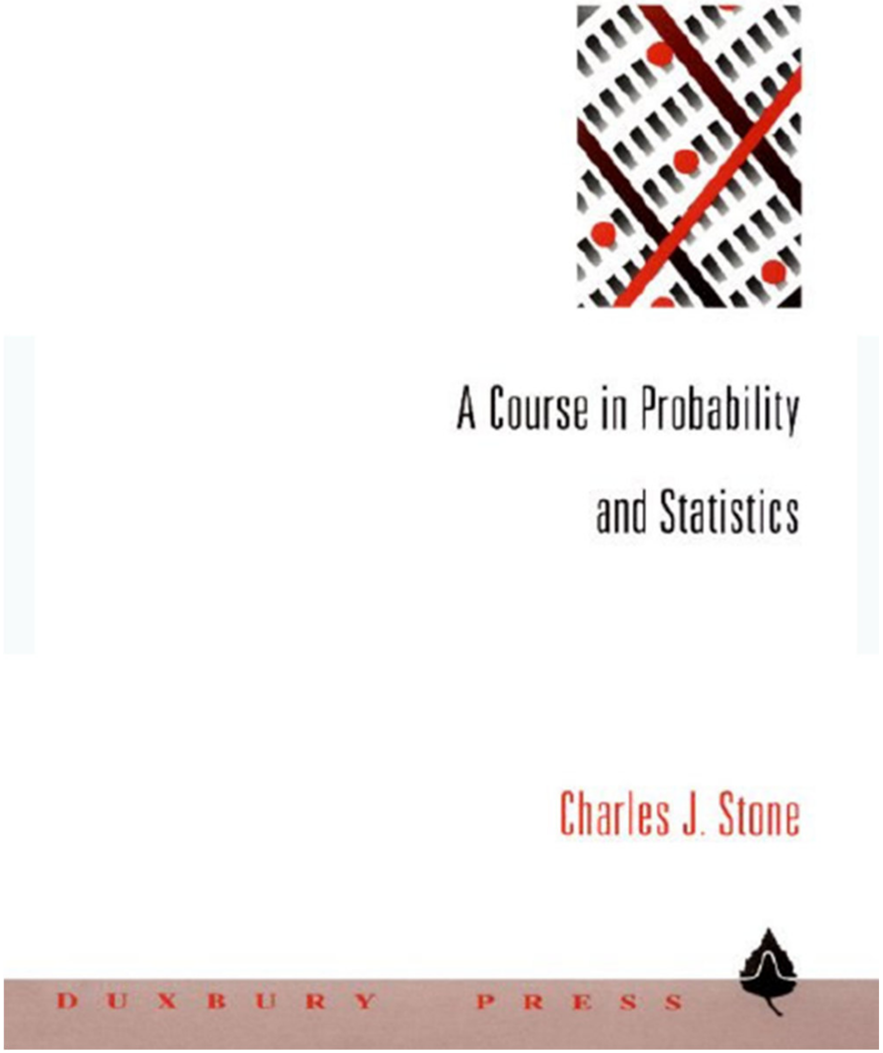
Real Zeros

Linear Equations and Matrix Inversion

Difference Equations

Chebyshev Approximation: Theory, Practice

ISBN-10: 0486652416



Random variables and distributions

PDF

Special discrete & normal models

Examples

ISBN-10: 9780534233280

Online C++ tutorial: http://www.cplusplus.com/doc/tutorial/

C++ Language Tutorial

These tutorials explain the C++ language from its basics up to the newest features of ANSI-C++, including basic concepts such as arrays or classes and advanced concepts such as polymorphism or templates. The tutorial is oriented in a practical way, with working example programs in all sections to start practicing each lesson right away.

[Download the entire tutorial as a PDF file]

Introduction

- Instructions for use

Basics of C++

- Structure of a program
- Variables. Data types.
- Constants
- Operators
- Basic Input/Output

Control Structures

- Control Structures
- Functions (I)
- Functions (II)

Compound Data Types

- Arrays
- Character Sequences
- Pointers
- Dynamic Memory
- Data Structures
- Other Data Types

Object Oriented Programming

- Classes (I)
- Classes (II)
- Friendship and inheritance
- Polymorphism

Advanced Concepts

- Templates
- Namespaces
- Exceptions
- Type Casting
- Preprocessor directives

C++ Standard Library

- Input/Output with files

SYLLABUS and REQUIREMENTS

Applied Computational & Numerical Methods

Course Content and Schedule

This is a short, intensive summer course to be completed in 5 weeks.

The contents include

probability and statistics, error analysis, numerical analysis of data, optimizations, solving systems of equations, algorithms, applications of numerical methods in physical sciences, and a final chapter on the neural network which will be followed by a set of NN examples.

PART I Stochastic Processes and the ROOT Program

Chapter 1 Probability and Statistics: Introduction or Review

- Probability: rules, distributions, error matrix, exercises
- Statistics: mean, variance, correlations, data, problem solving

Chapter 2 Introduction to ROOT and the C Programming Language (optional)

- ROOT: Introduction, installation, getting started
- C/C++: Introduction, examples, debugging, tutorials

Chapter 3 Monte Carlo Techniques

- Random number generations
- Distributions, quality of random variates, Monte Carlo simulations

Chapter 4 Experimental Errors

- Experiments and error estimates
- Statistical, systematic errors, averaging and combining errors, cases

Chapter 5 Data Analysis - Parameter Fitting and Hypothesis Testing

- Interpretation of estimates: meaning, limits and nonphysical estimates
- Maximum likelihood method
- Least Squares, hypothesis testing, minimization, and optimization

PART II Deterministic Processes

Chapter 6 Zeros and Extrema

- Introduction, methods, algorithms, and examples

Chapter 7 Integration of Functions

- Classical formulas and elementary algorithms
- Multidimensional integrals

Chapter 8 Solving Systems of Equations

- Fundamentals and algorithms
- Linear systems of equations, matrix inversion, and **partial differential equations**

PART III Identification, Forecast and Optimization

Chapter 9 The Neural Network Method for Pattern Recognition

- Introduction, methods, algorithms, and examples

Chapter 10 The Genetic Method for System Optimization

- Introduction, methods, algorithms, and examples

PART IV Advanced Topics

Chapter 11 Advanced Topics

- Problem posing simulations
- Global climate changes
- Full body auto collision

PART V Class Presentations of Student Projects

Chapter 12 Presentation of Your Favorite Projects

Applied Computational & Numerical Methods

Weekly Computing Projects

No homework assignments are made.

Students will have opportunities to work out homework style problems in class and after classes (not graded). Together students and the instructor will eventually go over these problems as exercises/examples in class.

4 to 5 projects will be assigned and are due in one week from the date of the assignment. These projects can be run on your own computers. Full instruction on these projects will be detailed in the project assignment.

Project Schedule

Applied Computational & Numerical Methods

The project includes the **basic assignment** for first year graduate students, and **an advanced topic** for more experienced students.

Project Schedule

Applied Computational & Numerical Methods

Week 1 **Getting started with the ROOT program package**
Introduction to the program, installation, setup, running, macros and document Tutorial

Week 2 **Statistical distributions**
Basic assignment: Determination of statistical features of data sets: mean, variance, standard error, error matrix and correlation between two variables
Advanced assignment: Statistical analysis of multivariable data sets and time series

Week 3 **Monte Carlo random variates; Monte Carlo experiments**
Basic assignment: Random number generation with root, statistical features, confidence intervals
Advanced assignment: A Monte Carlo based, statistical experiment to determine the significance of an observation

Project Schedule

Applied Computational & Numerical Methods

Weak 4 **Fits and the regression**

Basic assignment: Fit of functions to data: parameter determination and the goodness of the fit

Advanced assignment: Measurement of the lifetimes of heavy flavored hadrons

Week 5 **Numerical methods**

Partial differential equations

Week 6 **Neural network method**

Basic assignment: Backprop training on data, test of training results, optimization of the forecast capability

Advanced assignment: An optimization for new particle search

Week 7 **Project presentations**

SYLLABUS and REQUIREMENTS
Applied Computational & Numerical Methods

Class Hours *(preliminary--subject to change)*

Tuesday June 13, 17, 20, 27, July 1, 8, 15

2:00 – 3:30 pm   Lecture

3:30 – 4:00 pm   Computational projects: introduction & discussion

The meeting on July 15 is dedicated to student presentations on the projects. I will ask for volunteers soon.

Contact

Prof. Xinchou Lou   xinchoulou@yahoo.com

Why Do I need to learn
Applied Computational & Numerical Methods ?

Background radiation: A physics experiment searches for a new type of cosmic ray. A total of 18 candidates are found in a 6-month period. The background is of purely statistical nature possessing Gaussian distribution and has been evaluated to be 8.5 ± 2.5 by the physicists.

What is the probability that this background fluctuates to 18 or more candidates in the experiment?

Is there a signal?

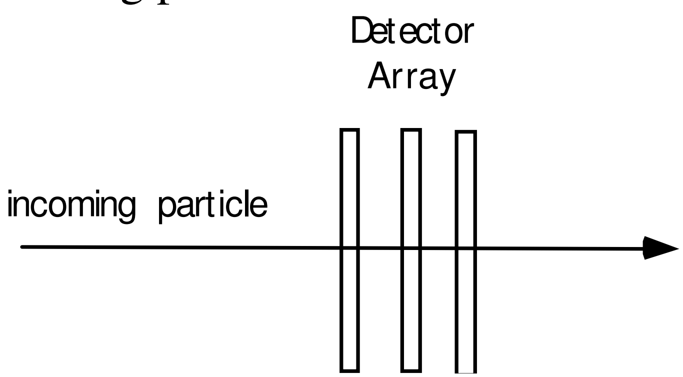
What is the significance of the observation?

Is a new type of cosmic ray discovered by this experiment?

Please fully justify your answer.

Why Do I need to learn
Applied Computational & Numerical Methods ?

Particle Detector: A detector array consists of three detection units (DU), arranged as shown below. Each of the DUs can detect and correctly identify both the proton and the electron, with efficiencies of 90% and 6%, respectively. A 'signal' is defined as correct detection by at least two DUs after a particle has traversed through the detector array. Each of the Dus is sufficiently fast that simultaneously arriving particles can be individually recognized.



- What is the efficiency of detecting a pure proton beam particle?
- What is the efficiency of detecting a pure electron beam particle?
- If the proton and the electron are always arriving in simultaneously, what is the efficiency for the proton-electron pair to be correctly detected and identified?

Why Do I need to learn
Applied Computational & Numerical Methods ?

Most of physics problem does not have analytical solutions, and then numerical solutions are needed:

One-dimensional diffusion equation

$$\frac{\partial}{\partial t}T(x,t)=\kappa\frac{\partial^2}{\partial x^2}T(x,t)$$

T temperature, **κ** thermal diffusion coefficient
If ***k*** is constant which is true in homogeneous media, we have analytical solution, otherwise no analytical solutions when ***k*** = ***k*** (x, t) except in special situations.

Schrödinger equation:

$$ih\frac{\partial}{\partial t}\psi(x,t)=-\frac{h^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x,t)+V(x)\psi(x,t)$$

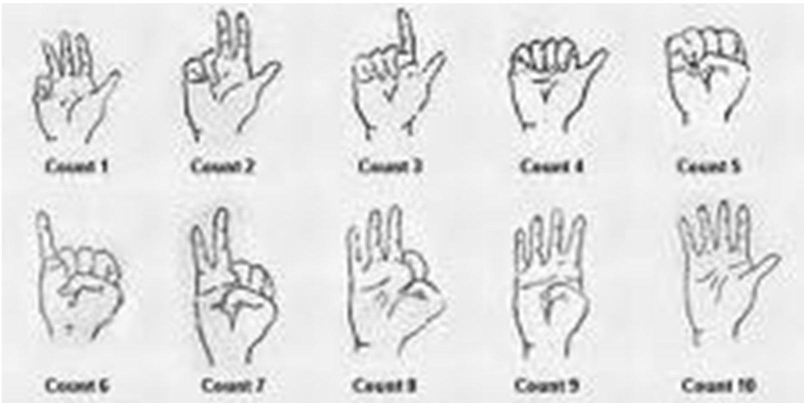
here *V(x)* could be very complicated function.

One-dimensional wave equation:

$$\frac{\partial^2 A}{\partial t^2}=c^2\frac{\partial^2 A}{\partial x^2}$$

If *c* is constant which is true in homogeneous media, we have analytical solution, otherwise no analytical solutions when $c = c(x)$ except in special situations.

Why Do I need to learn
Applied Computational & Numerical Methods ?



Correct data and computing $\Rightarrow$ very serious business

Why Do I need to learn
Applied Computational & Numerical Methods ?

knowledge +
ability to compute

huge advantage

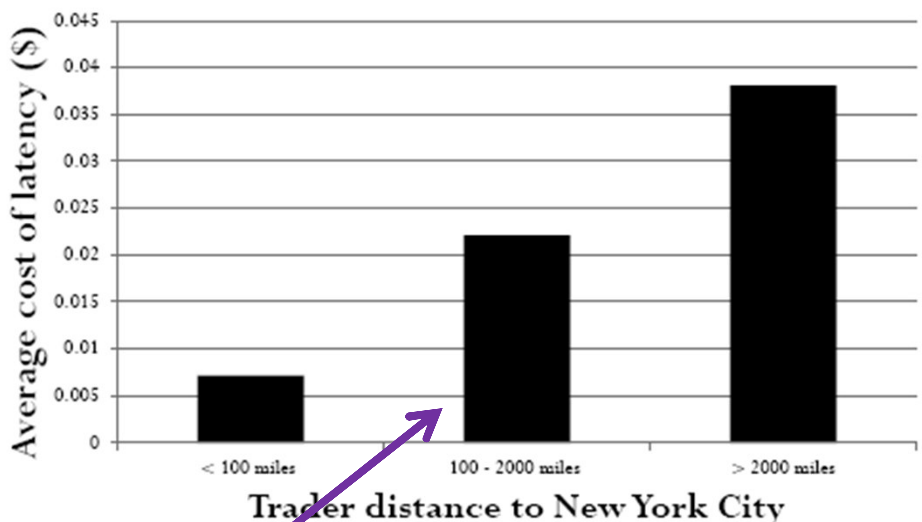
Why Do I need to learn

Applied Computational & Numerical Methods ?

Speed, Distance, and Electronic Trading: New Evidence on Why Location Matters

Latency Cost and Distance to NYC

- Sample firm is headquartered in the New York City area and has 36 branch office locations in 17 states.
- Gainesville is 1,014 miles from NYC!



Hedge-fund manager Adam Senderin his NYC office

Info & speed matter

Definition of Probability

In situations where essential circumstances are kept constant, and repetitions of experiments produce, though different, statistically (following a well-defined distribution) consistent results. The probability of obtaining a certain specified result on performing one of these experiments is then visualized as the ratio:

$$p = \frac{\text{number of occasions on which that result occurs}}{\text{total number of measurements}} \quad (0 \le \mathrm{p} \le 1)$$

In the limit of infinitive number of experiments and measurements the error for the probability is reduced to negligible level, this experimental probability approaches the true underlying probability for the result.

- from theory to data
- use theory to predict/calculate possible outcomes of experiment

Definition of Probability

Quantum states: spin of an electron $\pm\frac{1}{2}$. For an un-polarized electron the probability of finding it in $+\frac{1}{2}$ and $-\frac{1}{2}$ are the same, i.e., 50%. There is no way to predict without absolute certainty which spin state an electron is in without directly measuring it. For a total polarized electron it is then 100% or 0%, depending on the polarization orientation.

Transition among quantum states: heavier cousin of the electron, the muon (m=105 MeV, about 210 time of the electron's mass), is unstable and decays in the following fashion

$$\mu^- \to e^- \overline{\nu_e} \nu_\mu \ (\approx 99\%)$$

$$\mu^- \to e^- \overline{\nu_e} \nu_\mu \gamma \ (\approx 1\%)$$

$$\mu^- \to e^- \overline{\nu_e} \nu_\mu \mathrm{e}^+ \mathrm{e}^- \ (\approx 0.003\%)$$

for a decayed muon we know it most likely to be in the first mode, though we don't know for sure if it is since about 1% of the time it also decays into something else.

Definition of Probability

Random electronics noise: so called white noise, appear random in nature.

Light emission: photon emission, direction and polarization are all random in nature and cannot be described by deterministic algebra.

Radioactive decay: radioactive material has life time

These are all very different from Newtonian/Classical physics where from kinematics and dynamics the exact state of a motion can be determined without uncertainty. Therefore the math language and tools to describe modern physics are very much different as well. Physicists and engineers must use probability to quantify the physical reality

Probability & Statistics are very powerful when combined when computing power and data storage.

Definition of Probability

Situation where a static probability can not be used literally:

http://www.forbes.com/2006/01/09/winners-ride-stovall_in_ss_0106soapbox_inl.html?partner=yahootix

Data from past 36 years for S&P 500, capital appreciation only

	CAGR	Std. Dev.	Risk/Return	F.O.
S&P 500	7.5%	16.3	0.46	N.A.
Worst 10	7.6%	25.7	0.29	64%
Best 10	13.9%	23.6	0.59	72%

CAGR = Capital Appreciation Growth Rate
Std. Dev. = RMS Spread of Prices (a measure of risk)
Risk/Return = CAGR/Std Dev.
F.O. = Frequency of Outperformance, relative to the S&P 500 Index

Performance For 2004's Best And Worst Ten Industries

Best Ten Industries In 2004	% Chg 2004	% Chg 2005	Worst Ten Industries In 2004	% Chg 2004	% Chg 2005
Agricultural Products	46.6	10.5	Aluminum	(17.3)	(5.9)
Hotels, Resorts & Cruise Lines	44.1	0.3	Automobile Manufacturers	(16.9)	(49.2)
Internet Software & Services	66.8	4.0	Broadcasting & Cable TV	(9.1)	(16.9)
Managed Health Care	52.7	42.7	Electronic Manufacturing Services	(17.0)	(11.9)
Oil & Gas Drilling	45.1	53.1	Health Care Facilities	(10.9)	10.4
Oil & Gas Refining & Marketing	61.5	77.3	Insurance Brokers	(24.2)	12.9
Fertilizers & Agricultural Chemicals	93.0	39.6	IT Consulting & Other Services	(31.4)	(42.7)
Internet Retail	80.0	(25.3)	Pharmaceuticals	(9.5)	(5.9)
Steel	58.1	20.7	Semiconductor Equipment	(25.0)	2.0
Wireless Telecommunication Services	57.3	1.6	Semiconductors	(21.3)	11.2
Average	**60.5**	**22.4**	**Average**	**(18.3)**	**(9.6)**

9/10=0.90 **7/10=0.70**

S&P 500 Index returned 3% in 2005

Definition of Probability

So, which S&P 500 industries were the winners and losers in 2005? Take a look at the table below.

S&P 500's Best And Worst Ten Industries

Best Ten Industries In 2005	% Chg	Worst Ten Industries In 2005	% Chg
Oil & Gas Refining & Marketing	77.3	Automobile Manufacturers	(49.2)
Oil & Gas Exploration & Production	65.2	IT Consulting & Other Services	(42.7)
Oil & Gas Drilling	53.1	Photographic Products	(27.4)
Diversified Metals & Mining	48.8	Internet Retail	(25.3)
Oil & Gas Equipment & Services	47.3	Auto Parts & Equipment	(23.8)
Managed Health Care	42.7	Home Furnishings	(19.2)
Construction & Engineering	41.7	Food Distributors	(18.7)
Fertilizers & Agricultural Chemicals	39.6	Computer Storage & Peripherals	(17.1)
Health Care Services	32.2	Broadcasting & Cable TV	(16.9)
Railroads	30.9	Brewers	(15.3)

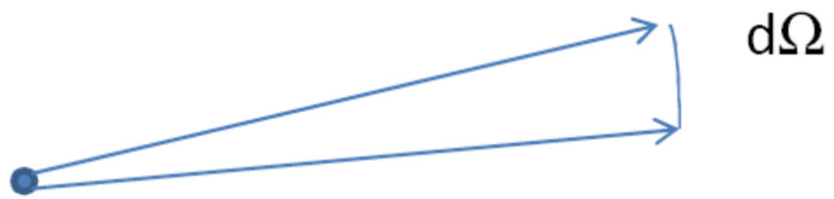
"…. There is no guarantee that what worked in the past will work in the future. Plus, no technique works all the time. We'll just have to wait and see."

Informal assignment to class:

What happened past several years (2005-2010)?

Examples of Probability

A radioactive source radiates gamma rays isotropically into space. For a detector (ε=100%) covering a solid angle of $d\Omega$, the probability of detecting the gamma ray will be $d\Omega/4\pi$ in any given decay, independent of the position of the counter.



Examples of Probability

An excited state has a very long lifetime τ, the probability that it has de-excited at time t is given by

$$p_t(de-excited) = \int_0^t \frac{e^{-t/\tau}}{\tau} dt$$

The probability that it will remain in the excited state is

$$p_t(excited) = 1 - \int_0^t \frac{e^{-t/\tau}}{\tau} dt$$

After one lifetime $t = \tau$,

$$\mathbf{p_t(de\text{-}excited) = 1\text{-}e^{-1} = 0.632}$$

and

$$\mathbf{p_t(de\text{-}excited) = e^{-1} = 0.3679}$$

random process: coin toss in a fair way

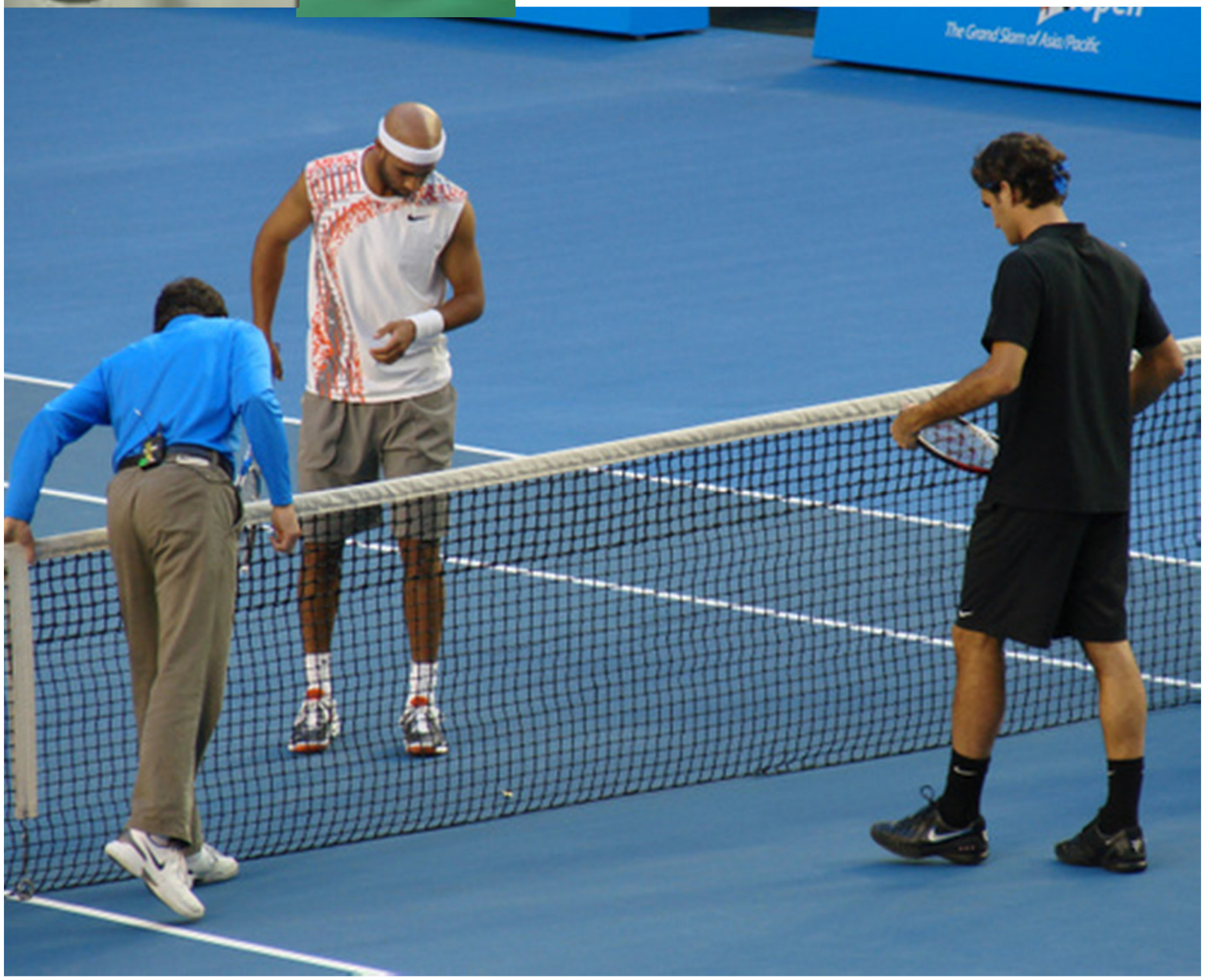
a simple **binomial** problem

"an event with exactly two possible outcomes"

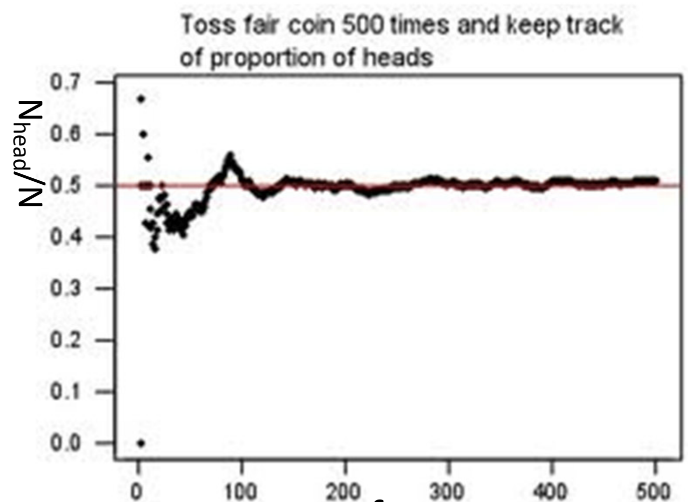
- $p_{head} + p_{tail} = 1.0$
- $p_{head} = p_{tail}$

$$\Downarrow$$

$$p_{head} = 0.50, \ p_{tail} = 0.50$$



James Blake vs. Roger Federer, 2008 Australian Open



larger data sample
$\Rightarrow$ better accuracy

random process: coin toss in a fair way (?)

also a binomial problem

European Roulette

- $\sum_{n=0}^{32} p_n = 1.0$
- $p_i = p_j \quad (0 \le i, j \le 32)$

$$\Downarrow$$

$$p_n = \frac{1}{33}$$

$$p_{n \neq 0} = \frac{32}{33}, \ p_0 = \frac{1}{33}$$

Rules of Probability

Rule 1: $0 \leq p \leq 1$

- p=0 means implies that a particular event ***never*** occurs
- p=1 means implies that a particular event ***always*** occurs

Rule 2: $P(A+B) \leq P(A) + P(B)$

- the probability **P(A+B)** that at least one of the events A or B occurs is equal to or smaller than the sum of individual probabilities P(A) and P(B)

- the equality stands when A and B are exclusive, whereas when A and B have common elements, the inequality applies

Example: throwing a dice-

$$P(3 \text{ or even}) = P(3) + P(2) + P(4) + P(6) = 4/6$$

$$P(\text{smaller than } 3.5 \text{ or even}) = P(1) + P(2) + P(3) + P(4) + P(6) = 5/6$$

[instead of P(1)+P(2)+P(3)+P(2)+P(4)+P(6)]

Rules of Probability

Rule 3: P(AB) = P(A/B)P(B)
= P(B/A)P(A)

the probability P(AB) of obtaining **both** A and B
and the conditional probability P(A/B) of A given B

conversely rule 3 defines P(A/B) = P(AB)/P(B)

If the occurance of B does not affect whether or not A occurs, then P(A/B) = P(A), and A and B are said to be **independent.** In this case P(AB) = P(A)P(B).

Rules of Probability

Example:

A = it is **rainy day**

B = **season** of the year

In a desert environment raining should be the same for any days c the week (without previous knowledge of the forecast, of course), therefore A and B are independent.

However if the place is in Dallas instead, B=April, A and B tends to be correlated here at Dallas, as we get more rain in Spring. i.e., **P(A/Spring) > P(A/Fall)** for Dallas.

Rules of Probability

Example

Detector efficiency for particles of two different species. The thickness and the detecting medium/electronics determine the probability if a particular kind of particle will be detected.

- A detecting particle a, probability **p(A)**
  B detecting particle b, probability **p(B)**

They are **independent** of each other

- Detecting both $p(A \text{ and } B)=p(A)\cdot p(B)$

  Detecting a but not b $p(A \text{ and "not } B\text{"})=p(A)\cdot[1-p(B)]$

  Detecting b but not a $p(\text{"not } A\text{" and } B)=[1-p(A)]\cdot p(B)$

  Detecting neither $p(\text{"not } A\text{" and "not } B\text{"})=[1-p(A)]\cdot[1-p(B)]$

Understanding Combinations

- Ordered combinations:

  a set of N different (numbers or objects) selected from a total sample of M (M=>N), the total number of possible combinations (taken into account their order) is

$$M(M-1)(M-2)....(M-N+1)=M!/(M-N)!$$

- Non-ordered combinations N from M

  factor due to ordering = N!

  total possible combination (with no regard to ordering)

$$\frac{M!}{(M-N)!\cdot N!}$$

Probability – Binomial Distribution

Understanding Binomial Distribution

- Situation: Conduct a fixed number(N) of **independent** trials, each of which can have only two possible outcomes:

yes (probability=p) /no (probability=1-p)

- For **n** yes, there must be (**N-n**) no. The probability for a single combination is then

$p^n \cdot (1-p)^{N-n}$ **per combination**

- Consider there are N!/n!/(N-n)! combinations (non-order) the probability of finding n occurances(succeses) is

$$p(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}$$

if we conduct only a fixed number (N) independent trials.

Understanding Binomial Distribution

To Understand the distribution:

- p^n is the probability of obtaining successes on n specific attempts;
- $(1-p)^{(N-n)}$ failure on remaining N-n attempts;
- the factorial term gives the number of permutations of n successes and N-n failures.

Understanding Binomial Distribution

For example: N=3 (p=1/2)

$$n=0,\quad \frac{3!}{0!3!}=1\ \ \langle FFF\rangle,$$

only 1 possible combination, $\text{p}=\frac{1}{8}$

$$n=1,\quad \frac{3!}{1!2!}=3\ \ \langle SFF, FSF, FFS\rangle,$$

3 possible combinations, $\text{p}=\frac{3}{8}$

$$n=2,\quad \frac{3!}{1!2!}=3\ \ \langle SSF, FSF, FSS\rangle,$$

3 possible combinations, $\text{p}=\frac{3}{8}$

$$n=3,\quad \frac{3!}{3!0!}=1\ \ \langle SSS\rangle,$$

1 possible combination, $\text{p}=\frac{1}{8}$

Probability – Binomial Distribution

mean value (expectation)

$$n_{mean} = \sum_n n \times p(n) = Np$$

variance of the distribution

$$\sigma^2 = Np(1-p) \cong N(\frac{n_{mean}}{N})(1-\frac{n_{mean}}{N})$$, when p is small $\sigma^2 \cong N(\frac{n_{mean}}{N}) \to n_{mean}$

The Standard Error (deviation) $\sigma \cong \sqrt{n_{mean}}$

The relative measurement error $\sigma / n_{mean} \cong \frac{1}{\sqrt{n_{mean}}}$

Which gets smaller relative to n_{mean} when n_{mean} increases.

Probability – Poisson Distribution

when $p = n/N << 1$, Np=constant, and $N >> 1$-- sufficiently large number of trials:

$$\frac{N!}{(N-n)!} = N(N-1)(N-2)....(N-n+1) \approx N^n$$

$$(1-\frac{<n>}{N})^{(N-n)} \approx e^{-(\frac{<n>}{N})(N-n)} \cong e^{-<n>}$$

Then Binomial distribution

$$p(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}$$

would be approximated to

$$p(n) \approx \frac{N^n}{n!} p^n e^{-<n>} = \frac{(Np)^n}{n!} e^{-<n>}$$
$$= \frac{<n>^n}{n!} e^{-<n>}$$

which gives the Poisson distribution

$$p(n) = \frac{<n>^n}{n!} e^{-<n>}$$

Probability – Poisson Distribution

(1) **mean = $<n>$**

Mean of n can be determined by weighting n with p(n)

$$\text{mean } \bar{\text{n}} = \sum_{n=0}^{\infty} n \times p(n) = \sum_{n=0}^{\infty} n \times \frac{<n>^n}{n!} e^{-<n>}$$

$$= 1 \times <n> e^{-<n>} + 2 \times \frac{<n>^2}{2!} e^{-<n>} + 3 \times \frac{<n>^3}{3!} e^{-<n>} + \ldots$$

$$= <n> e^{-<n>} \left[1 + <n> + \frac{<n>^2}{2!} + \ldots \right]$$

$$= <n> e^{-<n>} \times e^{<n>}$$

$$= <n>$$

Probability – Poisson Distribution

(2) variance = <n>, and standard deviation/error $\sigma = \sqrt{<n>}$

$$E(n(n-1)) = \sum_{n=1}^{\infty} n(n-1)\left[\frac{<n>^n}{n!} e^{-<n>}\right]$$

$$= <n>^2 e^{-<n>} \sum_{n=0}^{\infty} \frac{<n>^n}{n!}$$

$$= <n>^2 e^{-<n>} e^{<n>}$$

$$= <n>^2$$

also

$$E(n(n(-1)) = E(n^2) - E(n) = E(n^2) - <n>$$

therefore

$$E(n^2) = <n>^2 + <n>$$

The **variance** is

$$\sigma_n^2 = \sum_{n=1}^{\infty} p(n) \times [n - <n>]^2$$

$$= \sum_{n=1}^{\infty} p(n) \times \left[n^2 - 2<n>n + <n>^2\right]$$

$$= \sum_{n=1}^{\infty} p(n) \times n^2 - 2<n> \sum_{n=1}^{\infty} p(n) \times n + <n>^2$$

$$= <n>^2 + <n> - 2<n>^2 + <n>^2$$

$$= <n>$$

Probability – Gaussian Distribution

- **when <n> is large** (σ^2=<n> therefore n is always very close to <n>)
- **p=constant, N->∞**

Using the (2) variance = <n>, and standard deviation/error $\sigma = \sqrt{<n>}$

$$E(n(n-1)) = \sum_{n=1}^{\infty} n(n-1)\left[\frac{<n>^n}{n!}e^{-<n>}\right]$$

$$= <n>^2 e^{-<n>} \sum_{n=0}^{\infty} \frac{<n>^n}{n!}$$

$$= <n>^2 e^{-<n>} e^{<n>}$$

$$= <n>^2$$

also

$$E(n(n(-1)) = E(n^2) - E(n) = E(n^2) - <n>$$

therefore

$$E(n^2) = <n>^2 + <n>$$

$$N! \cong \sqrt{2\pi N}\, N^N e^{-N}$$

a Poisson distribution can be approximated to

Probability – Gaussian Distribution

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

$$\cong \frac{[n+(<n>-n)]^n}{\sqrt{2\pi n} n^n e^{-n}} e^{-<n>}$$

$$= \frac{[n+\Delta]^n}{\sqrt{2\pi n}} n^{-n} e^{n-<n>}$$

$$= \frac{[1+\frac{\Delta}{n}]^n}{\sqrt{2\pi n}} e^{-\Delta}$$

where **Δ = <n>-n** is the deviation from the mean <n>.

Remember that a Taylor expansion

$$n\ln(1+\frac{\Delta}{n}) = n\{\frac{\Delta}{n} - \frac{1}{2}(\frac{\Delta}{n})^2 + \frac{1}{3}(\frac{\Delta}{n})^3 + ...\}$$

$$\cong \Delta - \frac{\Delta^2}{2n}$$

and thus

$$(1+\frac{\Delta}{n})^n \cong e^{\Delta} e^{-\frac{\Delta^2}{2n}}$$

Probability – Gaussian Distribution

The above approximated Poisson distribution can be expressed as

$$P(n) \cong e\frac{[1+\frac{\Delta}{n}]^{n}}{\sqrt{2\pi n}}e^{-\Delta}$$

$$\cong \frac{e^{\Delta-\frac{\Delta^2}{2n}}}{\sqrt{2\pi n}}\ e^{-\Delta}$$

$$= \frac{1}{\sqrt{2\pi n}}e^{\Delta-\Delta-\frac{\Delta^2}{2n}}$$

$$= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\Delta^2}{2\sigma^2}}$$

where **σ^2 = <n> ≅ n**, so long as n is close to <n>

Probability – Gaussian Distribution

- Gaussian with one variable

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\cdot\sigma^2}}$$

and

$$n(x) = N \cdot f(x)$$

- Means = $\bar{x}$ (=<n>) and variance = σ^2
- Confidence levels

Understanding Gaussian distribution:

(1) function f(x) is a probability density, not a probability,

(2) $\int_{x_l}^{x_h} f(x)dx$ represents the probability for x to be between x_l and x_h,

value of the integral can be looked up in the Gaussian Int. Table

(3) function n(x) is also a probability density (**event/population density**)

(4) the shape is symmetric around $\bar{x}$

(5) ($\bar{x}$-σ, $\bar{x}$+σ) represents 68% of the area, or possibility, is

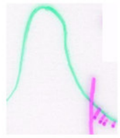
refered to as root of mean squared (rms)

(6) confidence levels: 1.64σ <=> 90%,

1.96σ <=> 95%,

2.58σ<=>99%.

Probability – Gaussian Distribution



Upper Limit

$CL = \int_{-\infty}^{X_{CL}} f_G(x)dx$, $f_G(x)$ is a standard Gaussian

0.0	.5000	.5040	.5080	.5120	5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554,	.6591	.6628	.6664	.6700	.6736	.6772	,6808	.6844	.6879
0.5	.6915"	.6950	.6985	,7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	,9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

Probability Distributions

N=11 — **binomial** ↓ — **Poisson** ↓ — **Gaussian** ↓

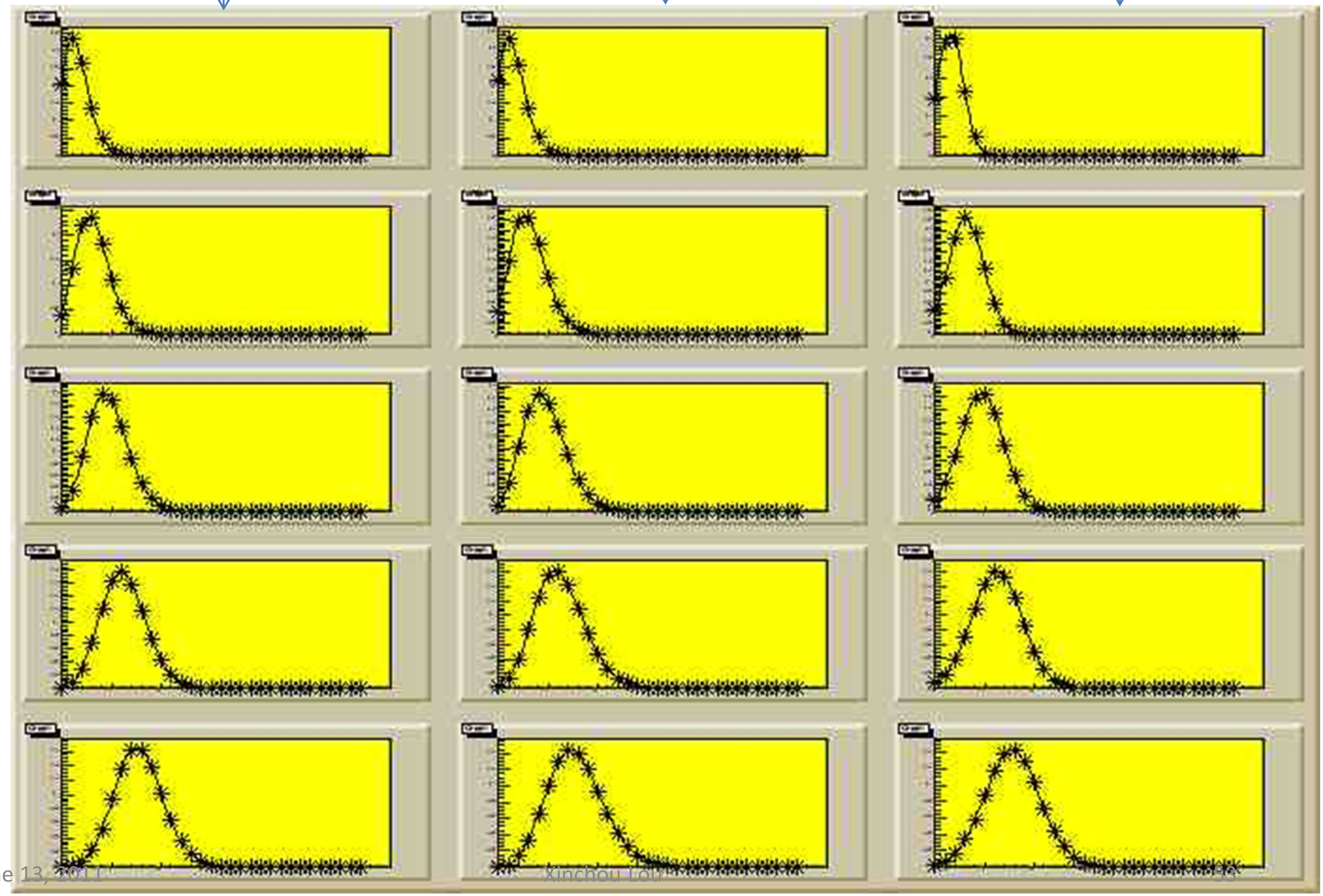
P=0.05

P=0.10

P=0.15

P=0.20

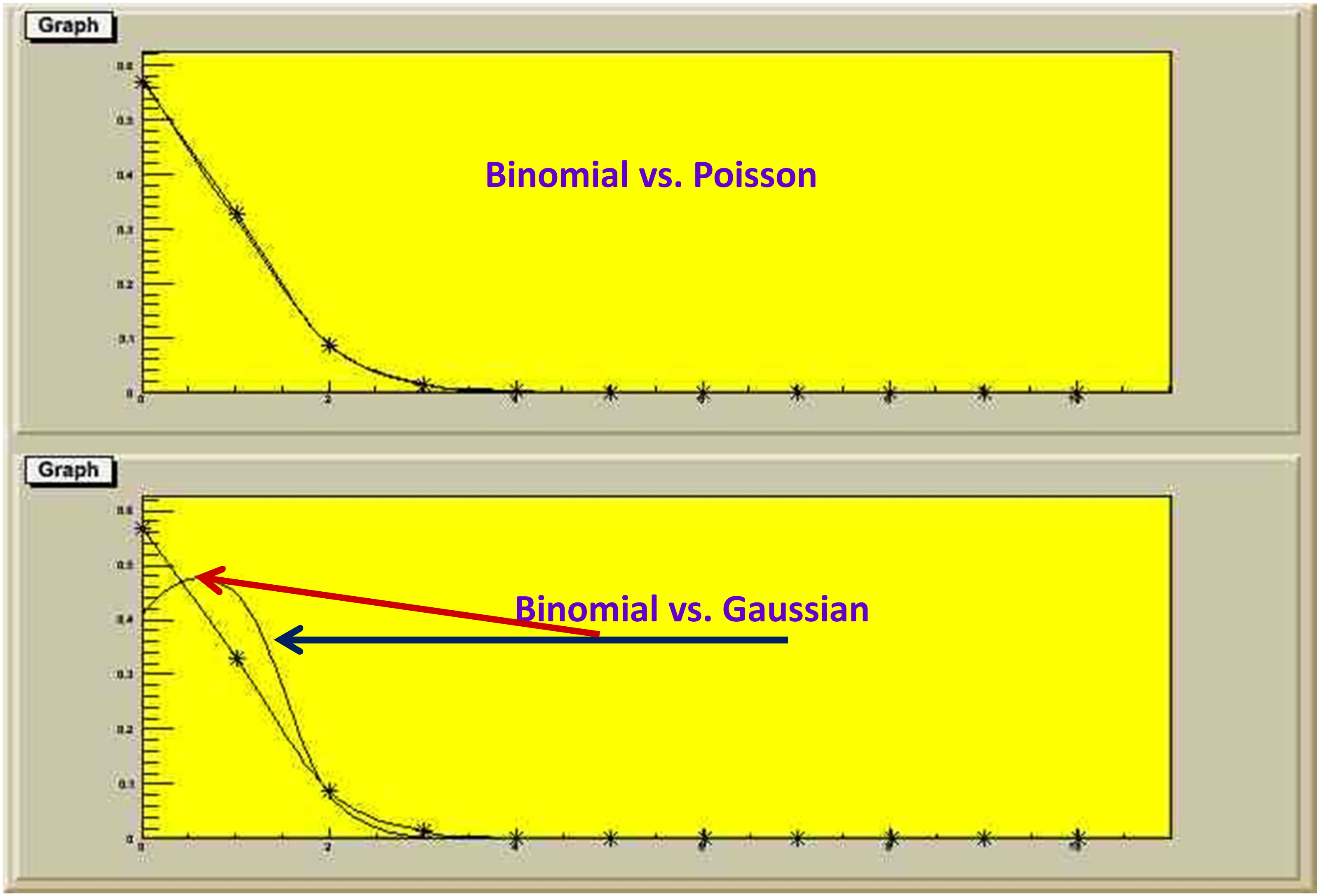
P=0.25



n

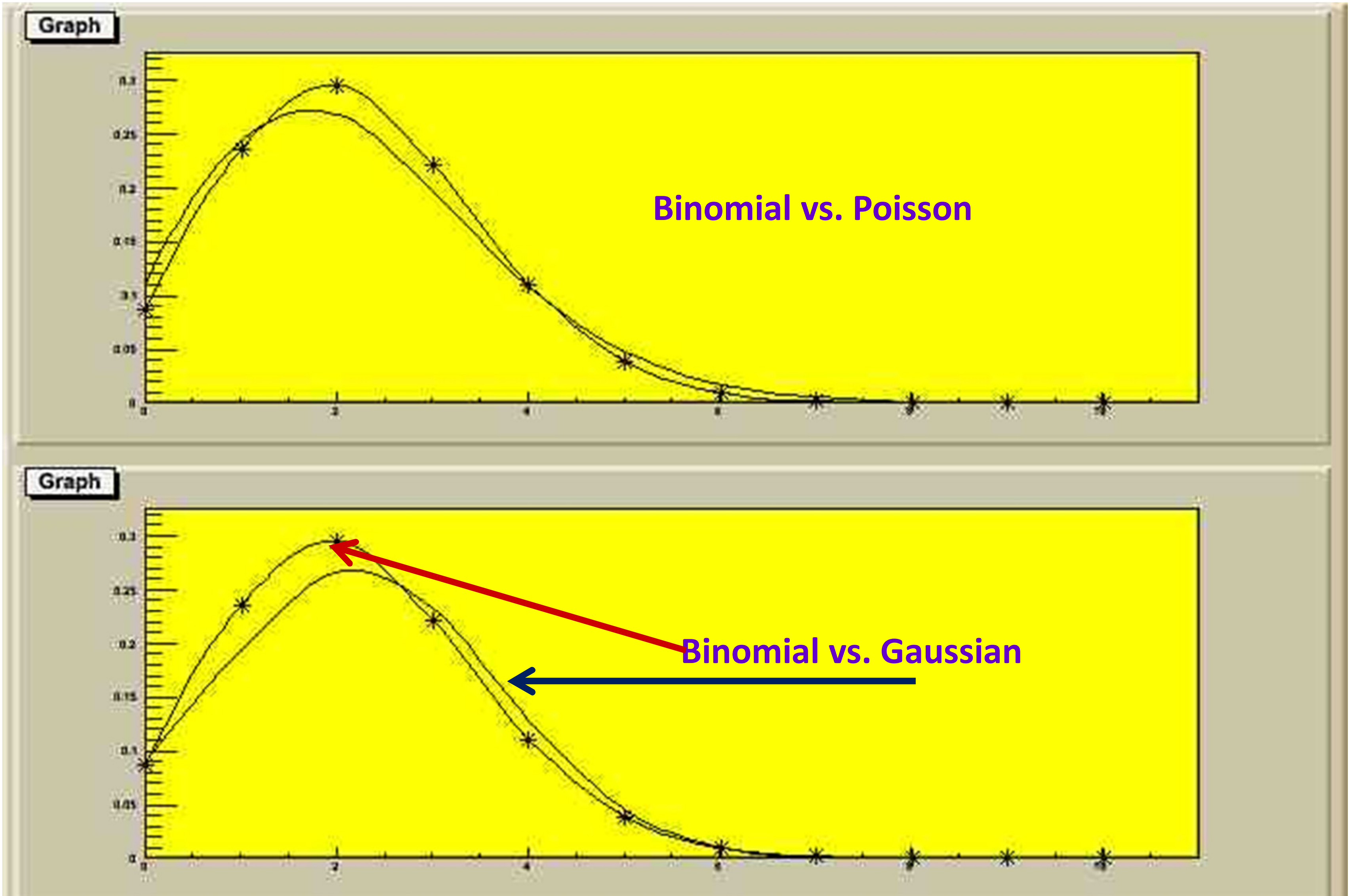
Probability Distributions

N=11, p=0.05

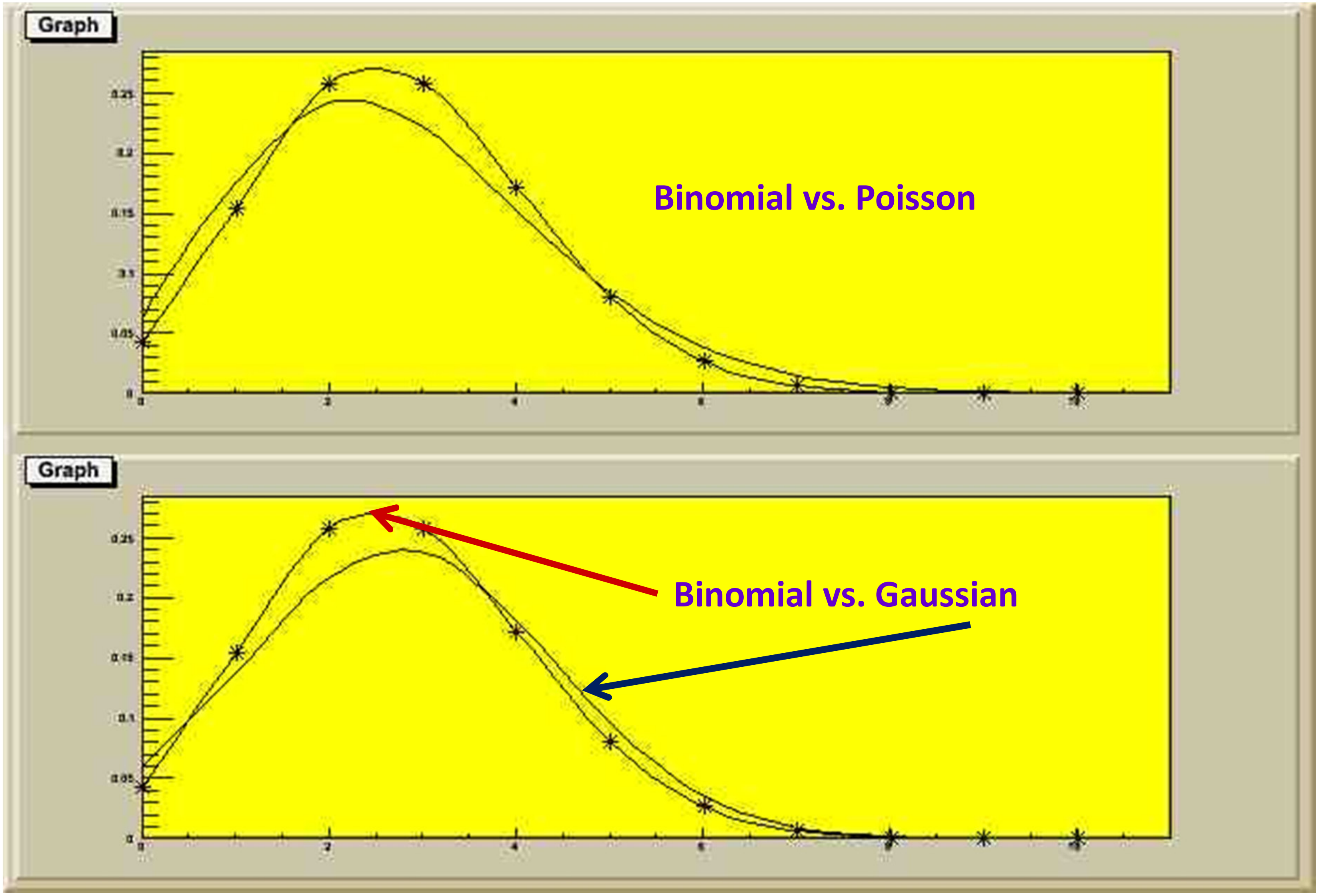


Probability Distributions

N=11, p=0.20, mean=2.2

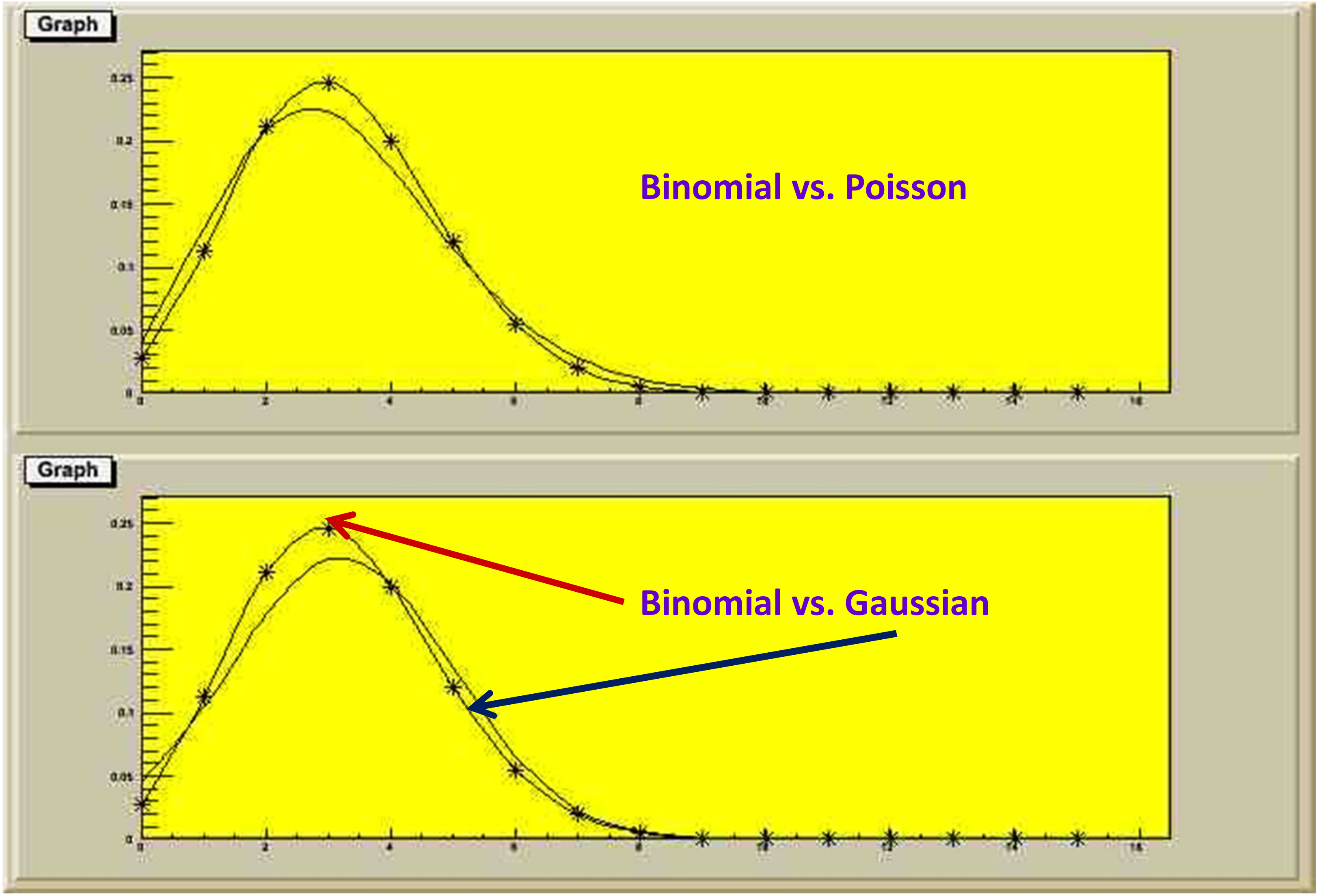


N=11, p=0.25, mean=2.75



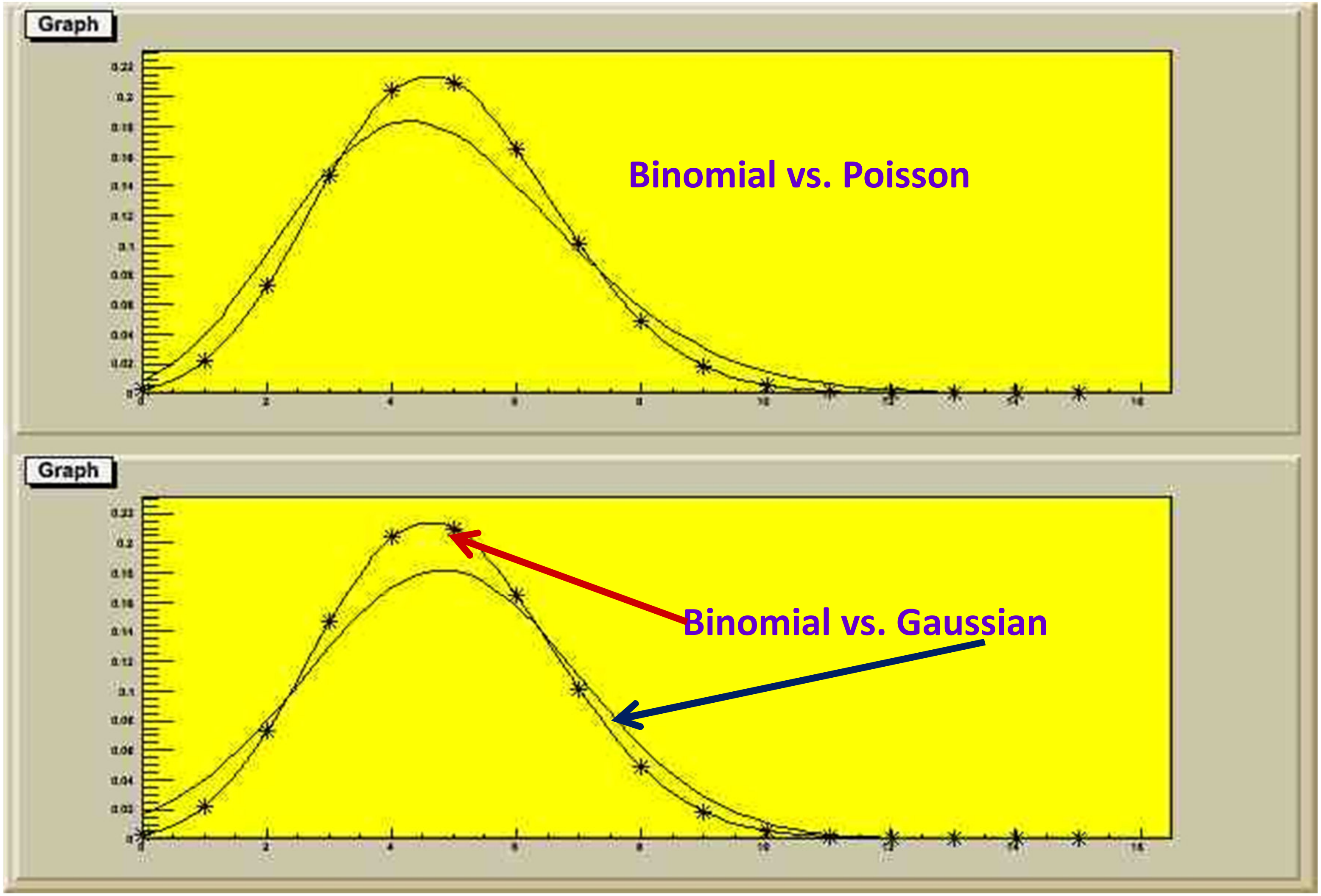
Probability Distributions

N=16, p=0.20, mean=3.2



Probability Distributions

N=16, p=0.30, mean=4.8



Good approximation by Gaussian distribution when **mean=Np≅7**

Summary of Lecture I

Organization

- Syllabus, TOC
- Ref. Books and Sites
- Projects

What have been covered

- Probability
- Binomial Distribution
- Poisson Distribution
- Gaussian Distribution & Gaussian approximation
- Mean, variances
- Examples