ECON 307/STAT 310
Probability and Statistics

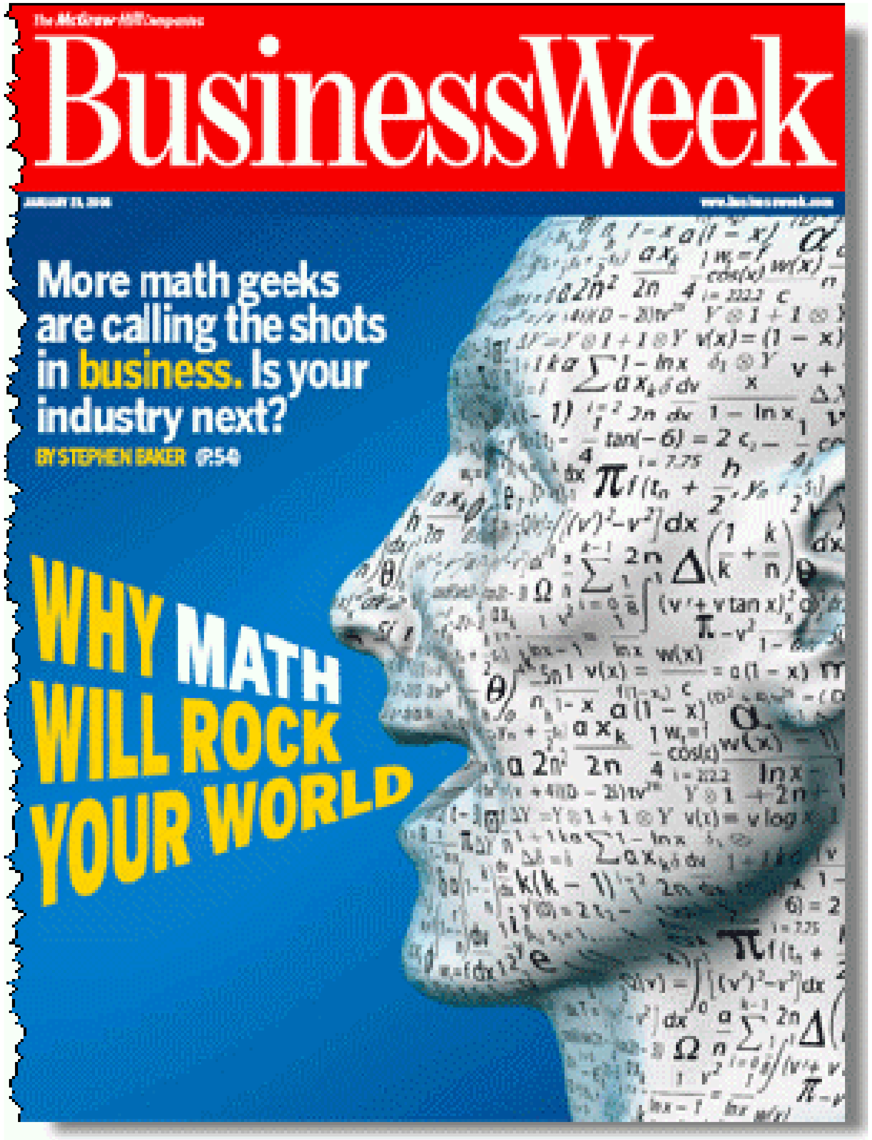
☞ *or* ☜

Introduction to Mathematical Statistics

Fall 2017

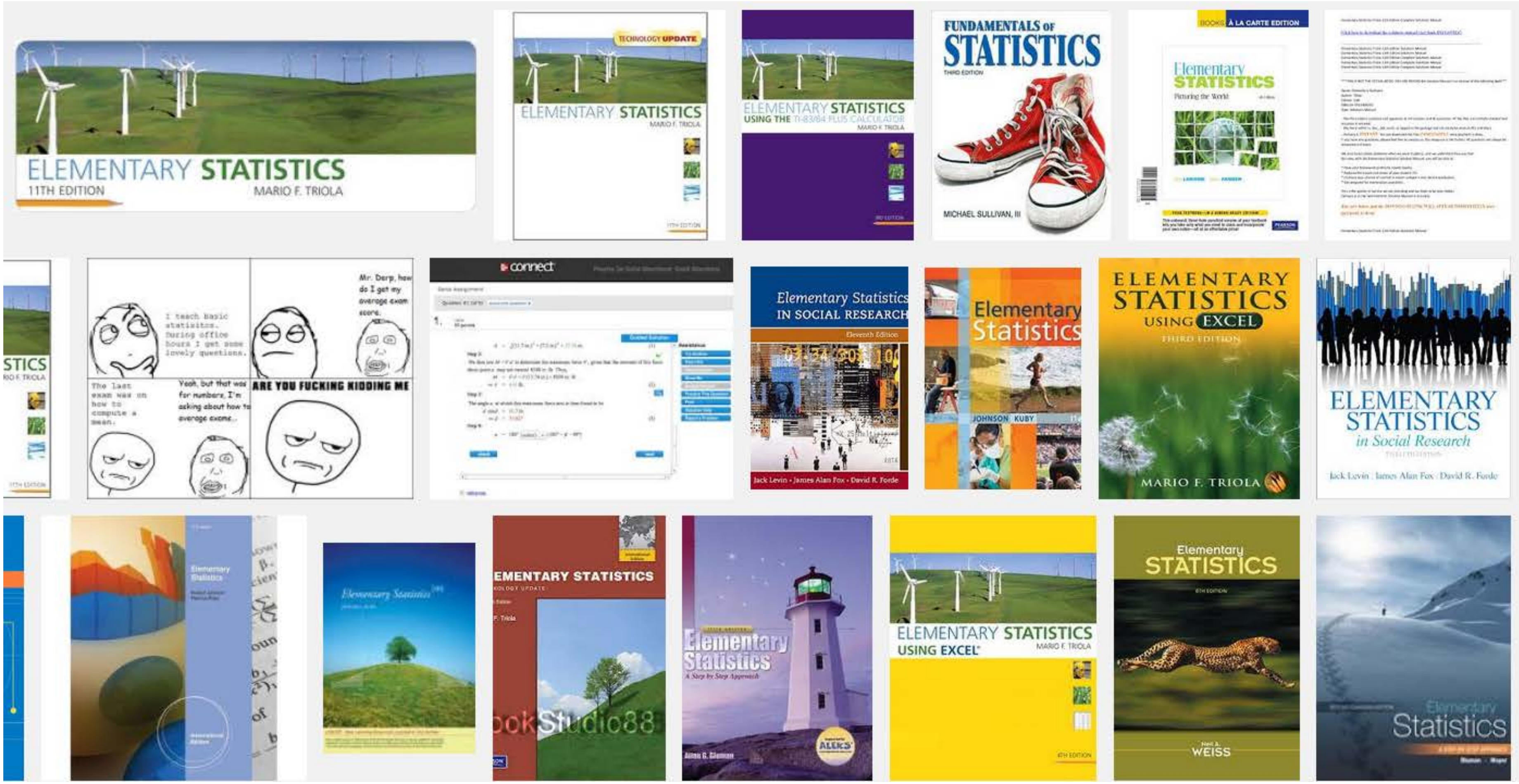
Everyone's Favorite Subject

Big Math



January 23, 2006

Elementary Statistics



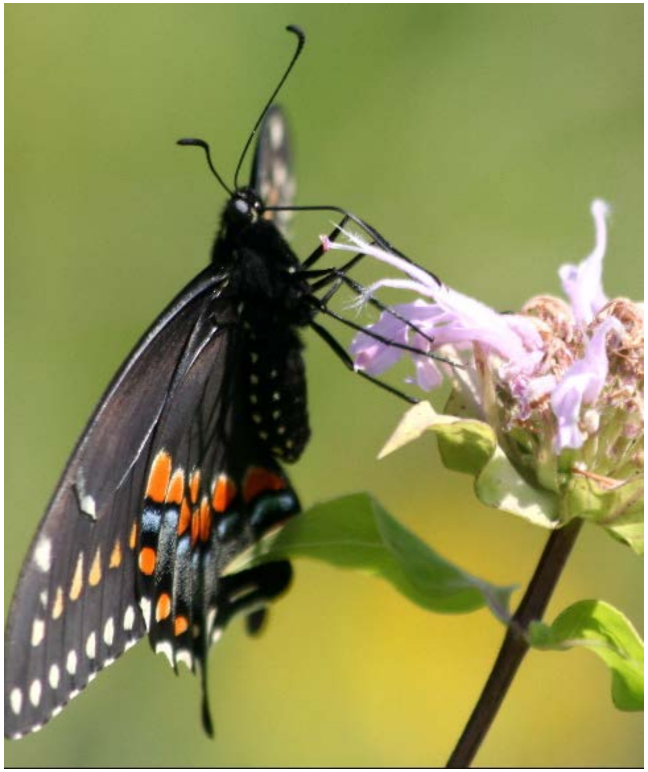
joy OF COOKING

Logistics and Expectations

- Fast Paced
  - Not an abbreviated exercise
  - Concise as possible
  - Required base knowledge for most of stat courses to come
- What to expect
  - Get and read the book(s)!
  - Lectures
  - Class participation
  - Homework
  - Exams

Understanding

- Dr. Dobelman's website:
  - How to find
  - http://dobelman.rice.edu
- Course Syllabus
- Canvas
- Pace yourself



Mathematical Statistics

$$
\mathrm{E}(X)=\int X dF(x)=\begin{cases}\int xf(x)dx & x \text{ is continuous}\\ \Sigma x_i p_i & x \text{ is discrete}\end{cases}
$$

Statistics

- What is “Statistics”
  - vs. what “are” statistics
    - [sample] average
    - [sample] proportion
- What is *N*?
- Why *talk* about probability and statistics?

Statistics

- Branch of mathematics (BOM) which studies methods for the **calculation of probabilities**
- BOM - collection and interpretation of **quantitative data** and the use of probability theory to **estimate population parameters**
- OED
  - The branch of political science concerned with the collection, classification, and discussion of (esp. numerical) facts bearing on **the condition of a state or community.**

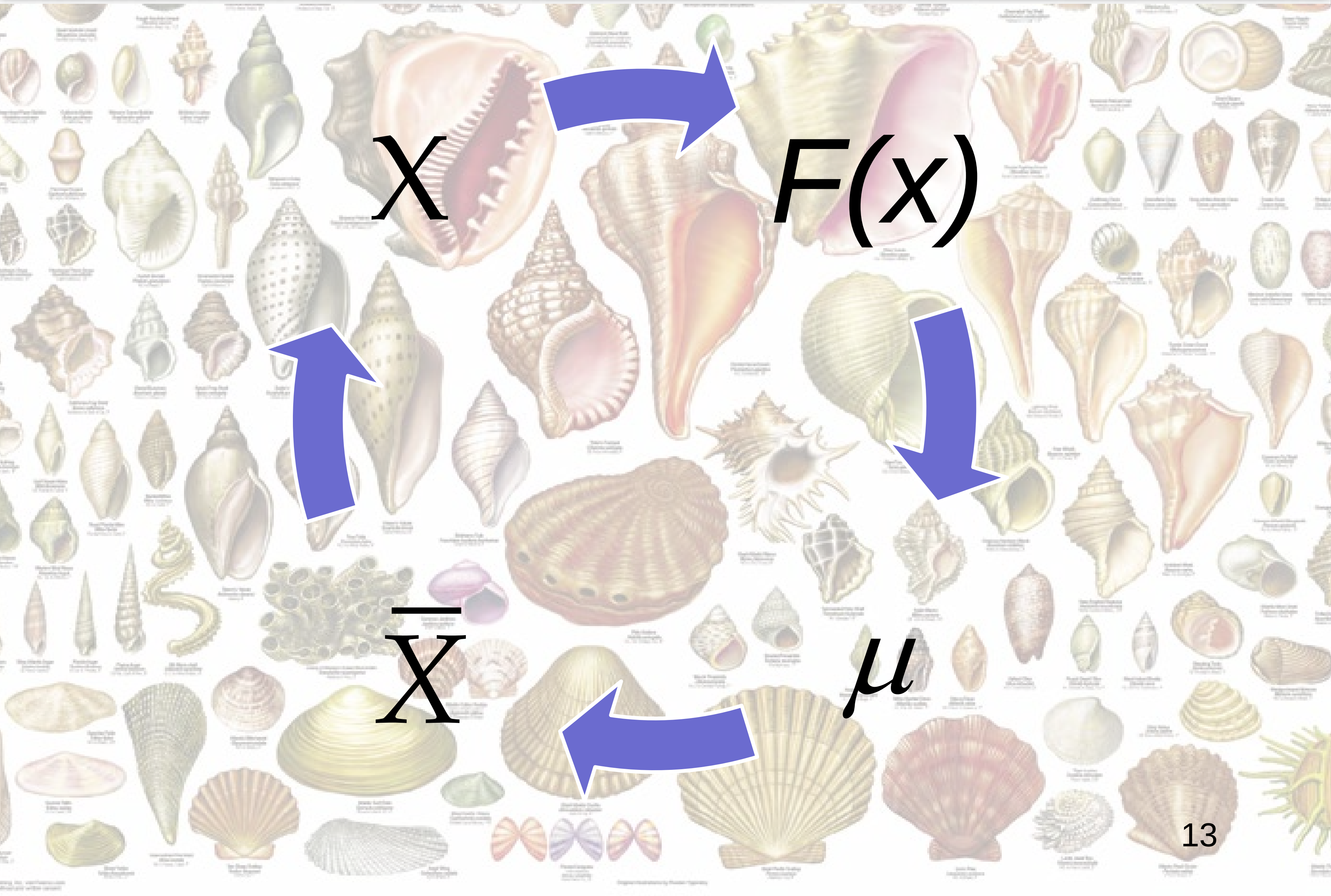
- The branch of science or mathematics concerned with the analysis and interpretation of **numerical data** and appropriate ways of gathering such data.
- Also, the systematic **collection and arrangement** of numerical facts or data of any kind

Don’t Like Statistics

- Why don’t people like statistics?
  - Some people “hate” statistics
  - Too hard to understand or learn
  - Doesn’t make sense/not intuitive
  - Too many formulas
  - Do not understand the applicability
  - Easy to misuse or “lie”

- Population versus samples
- Descriptive Statistics
  - Summarizing, describing, data reduction
- Probability (populations)
  - How the population behaves
  - Probabilities of obtaining outcomes
- Statistics and Inference (samples)
  - Data used to make statements or decisions about the universe from which the data are obtained

Population v. Sample



RICE UNIVERSITY

Descriptive

- Stem and Leaf
- Histogram
- Univariate summary statistics

114,023 EarthquakeMag.	
Mean	1.98
Median	1.90
Mode	1.90
Standard Deviation	0.8803
Variance	0.7749
Kurtosis	0.4995
Skewness	0.4353
Standard Error	0.0026
Range	7.3
Minimum	-0.6
Maximum	6.7
97% Conf. Level	0.0057

2 Days, Japan, 2/10/01

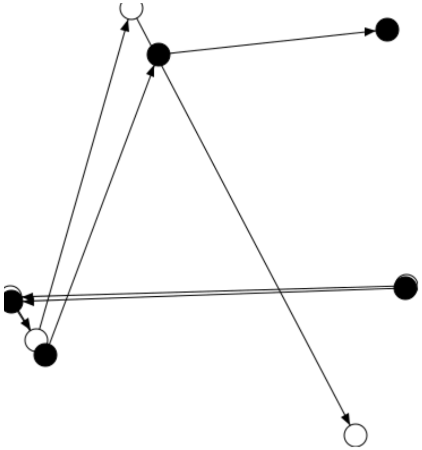
Terms of the binomial distribution for $n=20$ and values of p from 0.1 to 0.5 (each term has been multiplied by 10 000)

Number of successes	$p=0.1$ $q=0.9$	$p=0.2$ $q=0.8$	$p=0.3$ $q=0.7$	$p=0.4$ $q=0.6$	$p=0.5$ $q=0.5$
0	1216	115	8	—	—
1	2702	576	68	5	—
2	2852	1369	278	31	2
3	1901	2054	716	123	11
4	898	2182	1304	350	46
5	319	1746	1789	746	148
6	89	1091	1916	1244	370
7	20	545	1643	1659	739
8	4	222	1144	1797	1201
9	1	74	654	1597	1602
10	—	20	308	1171	1762
11	—	5	120	710	1602
12	—	1	39	355	1201
13	—	—	10	146	739
14	—	—	2	49	370
15	—	—	—	13	148
16	—	—	—	3	46
17	—	—	—	—	11
18	—	—	—	—	2
19	—	—	—	—	—
20	—	—	—	—	—

- Probability $\sup\left|F_n(x)-F(x)\right| \xrightarrow{a.s.} 0$
  - Mathematical foundation/basis for results in statistics
- Statistics $u^{\alpha}P(X>u)=\beta^{\alpha};\ f_{\hat{\alpha}}(a)=?$
  - Branch of scientific method that deals with [numerical] properties of populations that occur in nature (or our imaginations), of natural phenomena
  - Natural phenomena includes all the happenings of the external world, human or not
  - Estimators of parameters

Laplacian Determinism

- Laplacian Dæmon – 19th century ideal
  - Perfect knowledge of the past and the system → perfect prediction



- Poincare complication (per Mirowski, 1990)
  - Imperfect knowledge of past (minor errors) → wildly discrepant future predictions
  - 60 years before Mandelbrot

Keep in mind

- Understand variability
- Violin story
- Fundamental theorem of probability

$$\bar{X}_n \xrightarrow{a.s.} \mu$$

$$\sqrt{n}\,\frac{\bar{X}_n - \mu_X}{\sigma_X} \xrightarrow{d} \Phi(z)$$

- Fundamental theorem of statistics

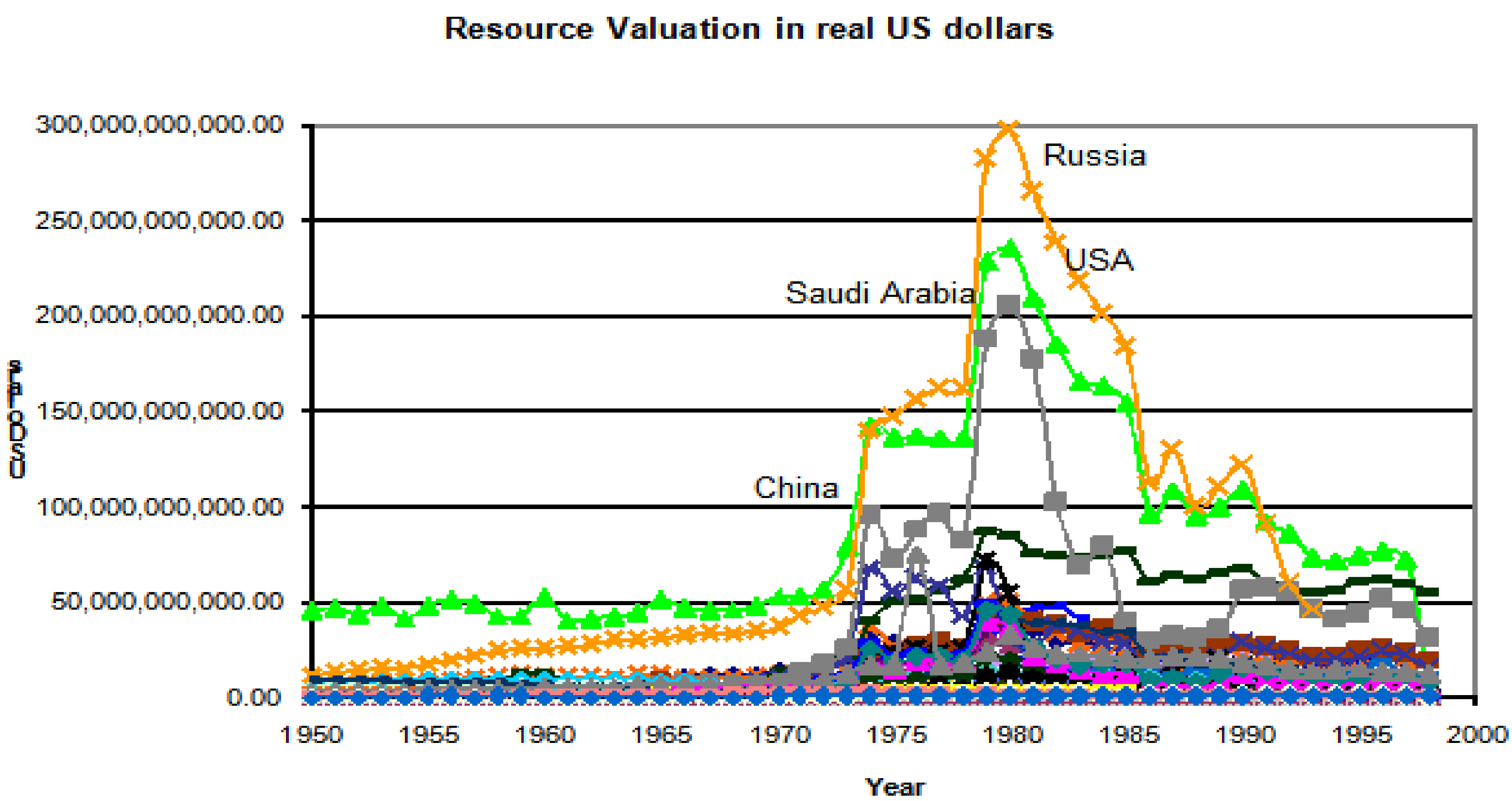
$$\sup\left|F_n(x) - F(x)\right| \xrightarrow{a.s.} 0$$

- Table of Contents

- Supposed to remember - Basic notation and definitions

RICE UNIVERSITY

How to Handle This?



MathStat Contents

1. Probability and Distributions
2. Multivariate Distributions
3. Some Special Distributions
4. Some Elementary Statistical Inferences
5. Consistency and Limiting Distributions
6. Maximum Likelihood Methods
7. Sufficiency
8. Optimal Tests of Hypotheses
9. Inferences about Normal Models
10. Nonparametric and Robust Statistics
11. Bayesian Statistics

$$\left(\Omega,\mathcal{F},P\right)\xrightarrow{\mathrm{X}(\omega)}\left(\mathbb{R},\mathcal{B},F\right)$$

$$\mathrm{P}(X\in A)=?$$

$$\mathrm{E}(g(X))=\int_{\Omega} g\left(X(\omega)\right)d\,\mathrm{P}(\omega)=\int_{\mathbb{R}} g\left(X(\omega)\right)dF(x)$$

- Set theory
- The probability set function
- Conditional probability and Independence
- Random variables
  - Discrete random variables
  - Discrete random variables - Transformations

- Continuous random variables
- Continuous random variables - Transformations

- Expectation of a random variable
- Some special expectations
- Important Inequalities

$$(\Omega, \mathcal{F}, P) \xrightarrow{\begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix}} \left(\mathbb{R}^2, \mathcal{B}_2, F_{X,Y}\right)$$

$$f_{X,Y} = ?$$

$$\Sigma = \mathrm{E}\big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big) = \iint_{\mathbb{R}^2} \big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big)\, dF_{XY}$$

$$(U, V) = g(X, Y)$$

$$u = g_1(x, y)$$
$$v = g_2(x, y)$$

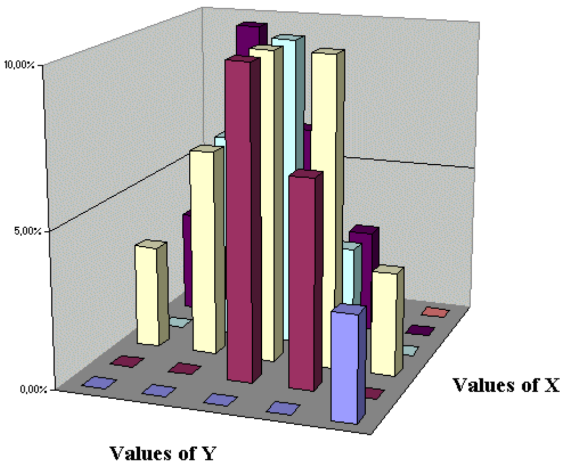
$$f_{U,V} = ?$$

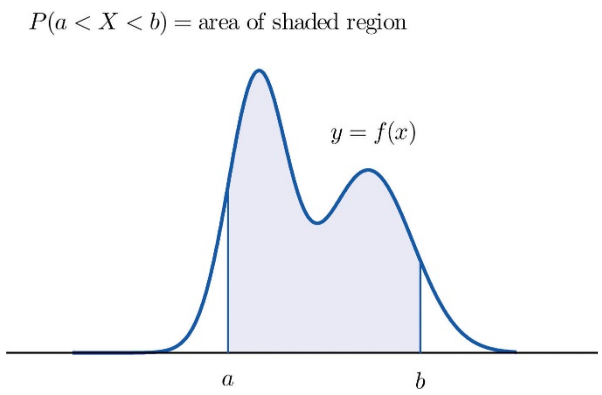
- Distribution of two random variables
- Distribution of two random variables - Expectation
- Transformations - Bivariate random variables
- Conditional distributions and expectations
- The Correlation Coefficient

- Independent random variables
- Extension to several random variables
- Covariance matrices
- Transformations for several random variables
- Linear combinations of random variables

3. Some Special Distributions

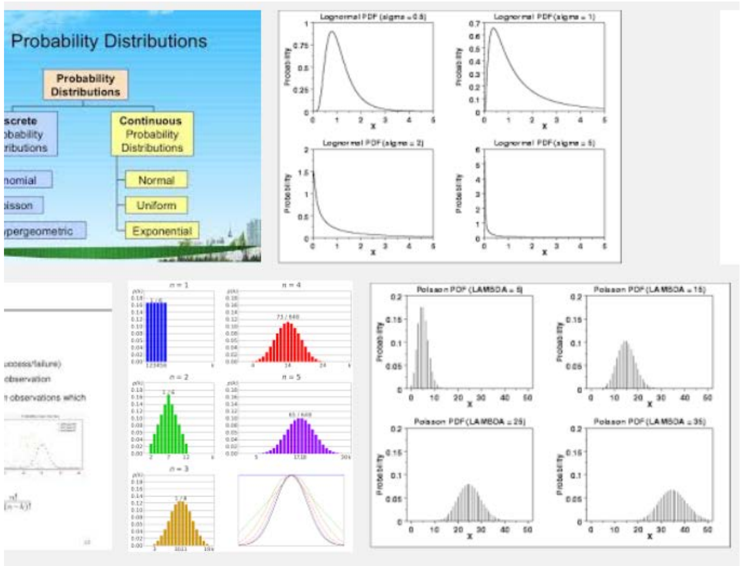
$$\mathrm{P}(X=x)=f_X(x)=p_x$$



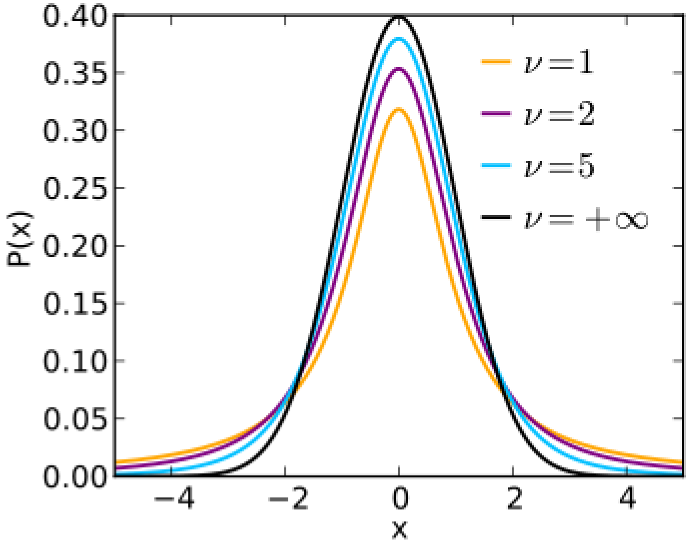


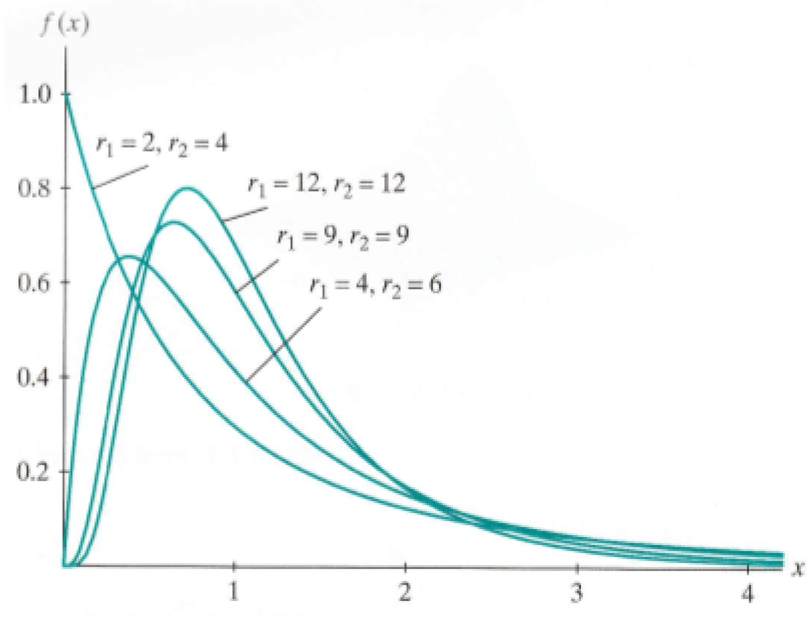
$$\mathrm{P}(X\le x)=F_X(x)=\int_{-\infty}^{x}dF=\int_{-\infty}^{x}f(x)dx$$

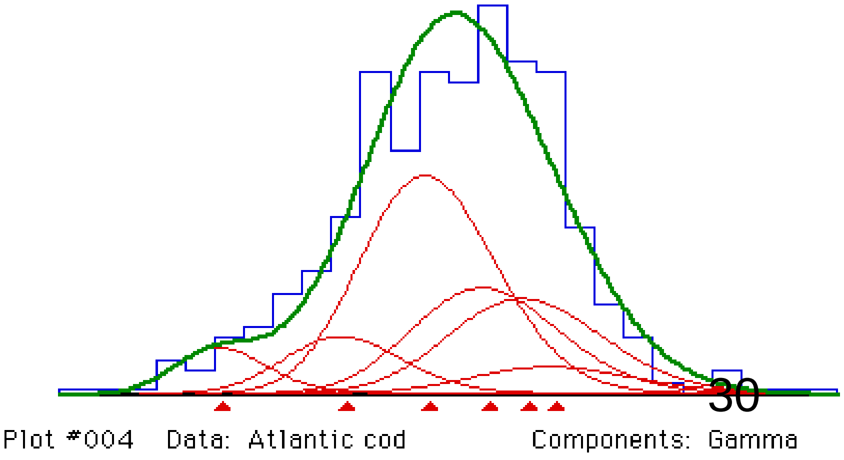
- Binomial and related distributions
- The Poisson distribution
- The gamma, chi squared and beta distributions
- The Normal distribution



- The multivariate normal distribution
- The t-distribution
- The F-distribution
- Student's theorem
- Mixture distributions







$$X_1, \ldots, X_n$$

$$X_i \text{ iid } F_X(x) \text{ with } f_X(x \mid \theta)$$

$$W(X) = W(X_1, \ldots, X_n) \equiv \bar{X} = \hat{\theta}$$

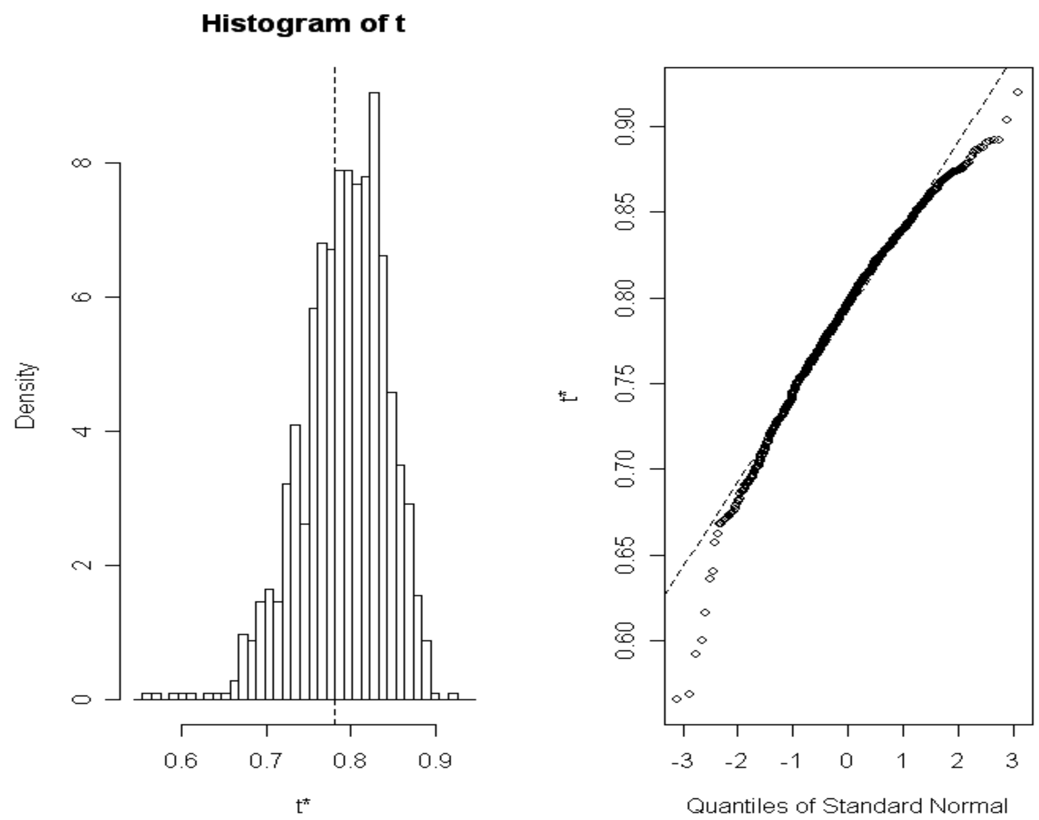
$$\mathrm{P}\big(L(X) \leq \theta \leq U(X)\big) = 1 - \alpha = 95\%$$

$$H_0 : \mu = 100$$
$$H_1 : \mu \neq 100$$

- Sampling and statistics
- Histogram estimates of pdmf's
- Confidence intervals
- Confidence intervals for differences in means
- Confidence intervals for differences in proportions
- Confidence intervals for parameters of discrete distributions

- Order statistics
- Quantiles and confidence intervals for quantiles
- Introduction to hypothesis testing
- Additional comments about statistical tests
- Chi-Squared tests
- Monte Carlo

- Bootstrap procedures
- Percentile bootstrap confidence intervals
- Bootstrap testing procedures



$$\left.\begin{array}{l} X_n \xrightarrow{a.s.} X \\ X_n \xrightarrow{L^p} X \\ X_n \xrightarrow{c} X \end{array}\right\} \Rightarrow X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$$

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma_X} \xrightarrow{d} \phi$$

- Convergence in probability
- Convergence in distribution
- Convergence in distribution - Bounded in probability
- Convergence in distribution - Delta method
- Convergence in distribution - MGF technique
- Central limit theorem
- Multivariate extensions

For a sequence of random variables X_n
with measures α_n,
cumulative distribution functions F_n,
and characterstic functions ϕ_n ,
we have the following notions of convergence :

in rth mean / in L^r , r >1

$$\lim_{n\to\infty} \int (| X_n(\omega) - X(\omega) |)^r \, d\alpha(\omega) = 0$$

in mean / in L^1

$$\lim_{n\to\infty} \int | X_n(\omega) - X(\omega) | \, d\alpha(\omega) = 0$$

almost surely / with probability 1

$$\mathbb{P}\left[\lim_n X_n(\omega) = X(\omega)\right] = 1$$

in probability / in measure

$$\forall \epsilon > 0, \ \lim_{n\to\infty} \mathbb{P}[| X_n(\omega) - X(\omega) | > \epsilon] = 0$$

in distribution / weakly / in law

$$\lim_{n\to\infty} \alpha_n[I] = \alpha[I]$$
$\forall$ measurable I with $\alpha[\partial I] = 0$

$$\lim_{n\to\infty} F_n(t) = F(t)$$
$\forall\, t \in \mathbb{R}$ where $F(t)$ is continuous

$$\lim_{n\to\infty} \phi_n(t) = \phi(t)$$
$\forall\, t \in \mathbb{R}$, ϕ is continuous at 0

$$\lim_{n\to\infty} \int f(\omega) \, d\alpha_n(\omega) = \int f(\omega) \, d\alpha(\omega)$$
$\forall$ f continuous and bounded

If
$\exists$ Y a r.v. such that
$\forall n$, $\mathbb{P}[| X_n | < Y] = 1$
and $\mathbb{E}\,Y < \infty$

If
$X_n = \sum_{i=1}^{n} Z_i$
for Z_i independent r.v.s
OR
$\forall \epsilon > 0$,
$\sum_{i=1}^{\infty} P(| X_n(\omega) - X(\omega) | > \epsilon) < \infty$

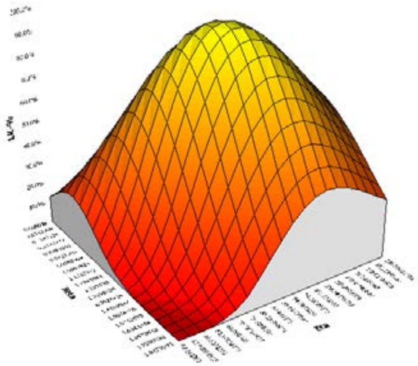
If
$\exists\, c \in \mathbb{R}$ such that
$\forall n$, $\mathbb{P}[| X_n | < c] = 1$

If
converges
to a constant

$$\hat{\theta} = W(X_1, \ldots, X_n)$$

$$L(\theta \mid x) = f_X(x \mid \theta)$$

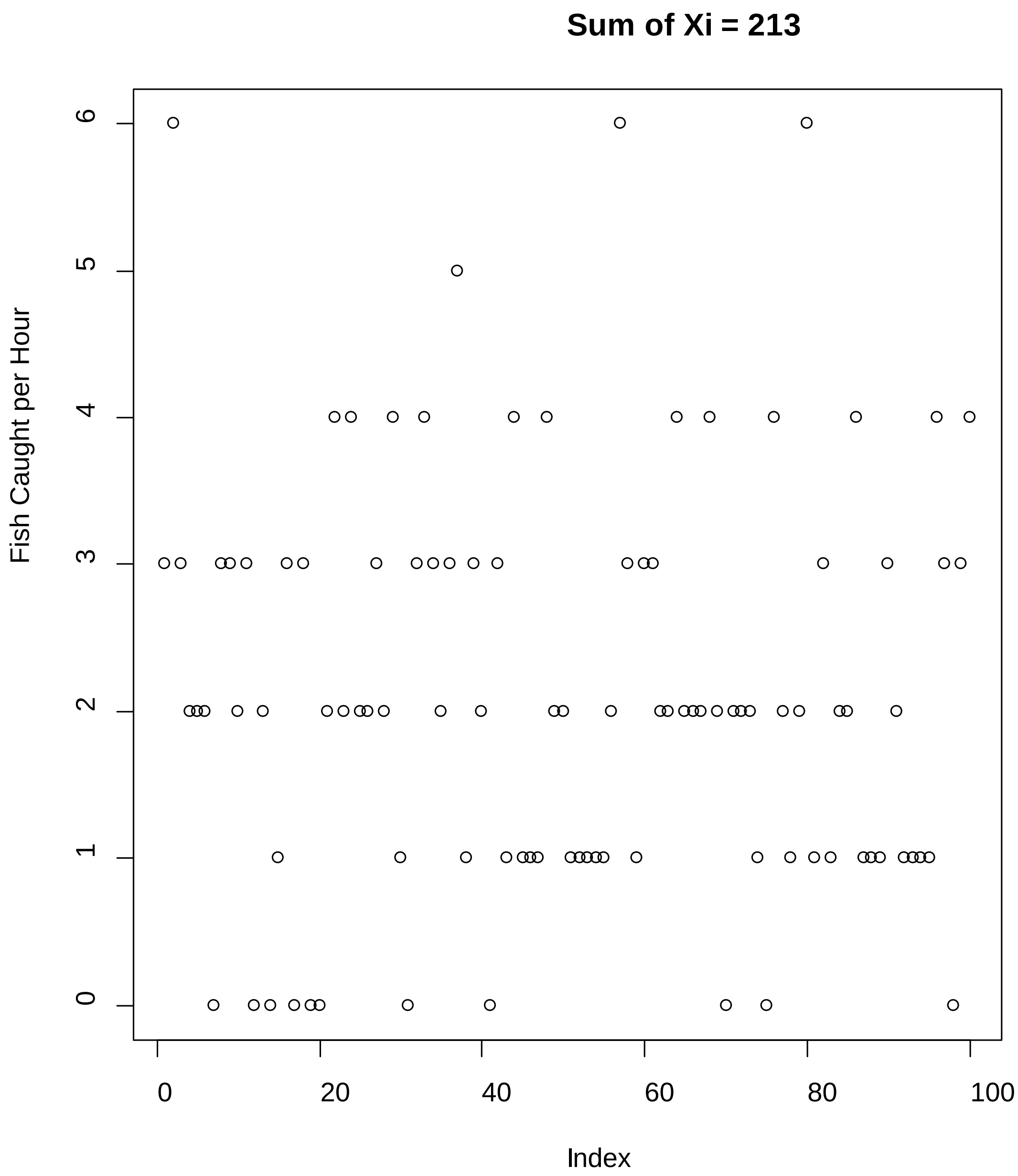
Likelihood Function Surface



$$\mathrm{Var}(W) \geq \frac{\mathrm{E}(\dot{g}(\theta))^2}{I_n(\theta)}$$

- Maximum Likelihood estimation
- Cramer-Rao lower bound and efficiency
- Maximum likelihood tests
- Multi parameter case: Estimates
- Multi parameter case: Testing
- the EM algorithm

7. Sufficiency



$$\lambda = ?$$

7. Sufficiency

- Measures of quality of estimators
- A sufficient statistic for a parameter
- Properties of a sufficient statistic
- Completeness and uniqueness
- The Exponential class of distributions

$$f_X(x \mid \theta) = e^{w'(\theta)T(x)+H(x)+C(\theta)}$$

7. Sufficiency

- Functions of a parameter
- The case of several parameters
- Minimal sufficiency and auxiliary statistics
- Sufficiency, completeness and independence

8. Hypothesis Testing

- Hypothesis

$$H_0 : \theta = \theta_0 \qquad H_0 : \theta = \theta_0$$
$$H_1 : \theta = \theta_1 \qquad H_1 : \theta \neq \theta_0$$

$$H_0 : f = f_0$$
$$H_1 : f \neq f_0$$

- Test statistic $T = T(X_1, X_2, \ldots, X_n)$
- Critical region (to reject H0)

$$R = \{x : \text{Reject}\, H_0\}$$

- Significance level (α)

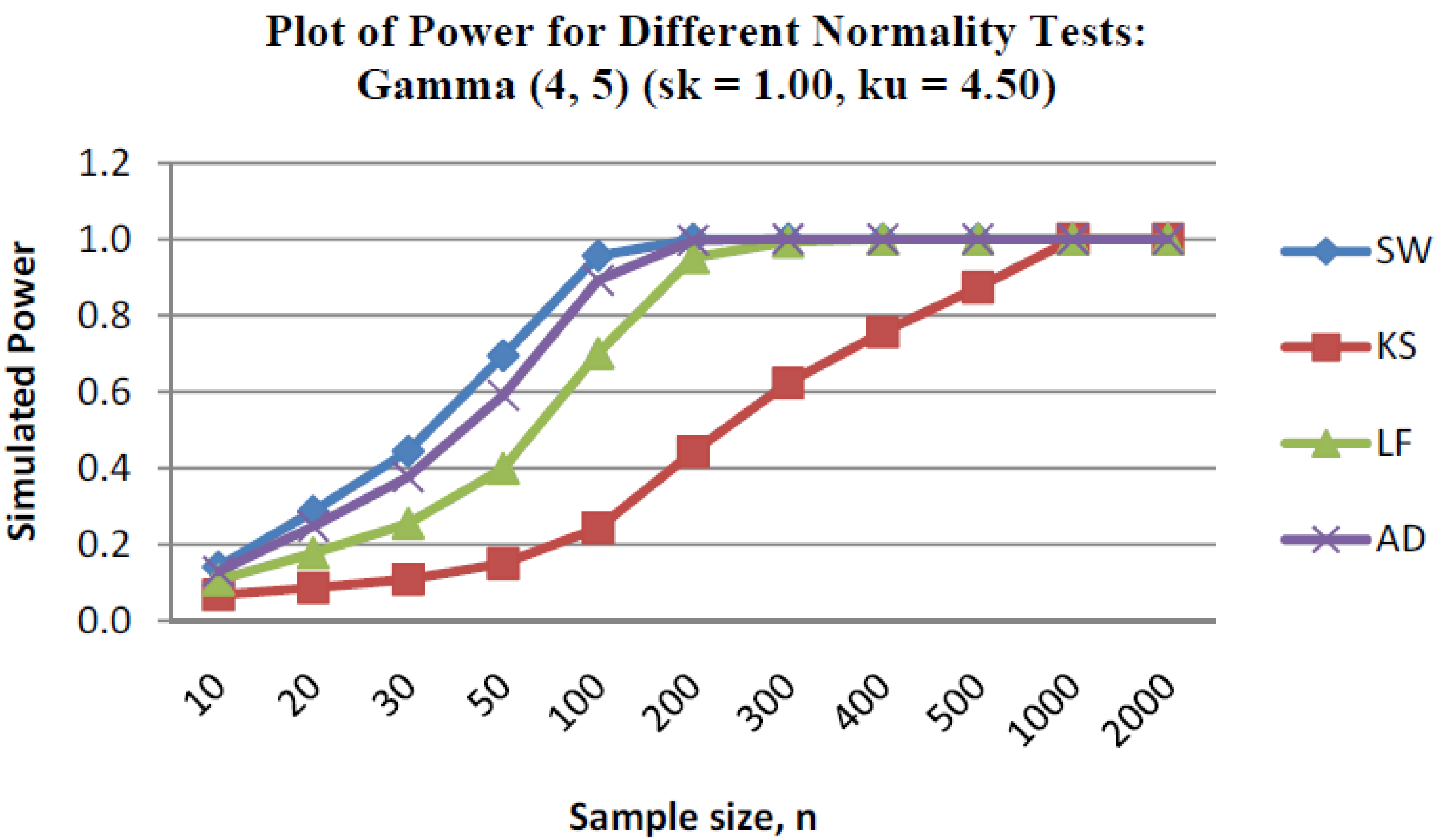
- Power of the test (β)

$$\beta(\theta) = \mathrm{P}(\text{reject } H_0)\,\forall\,\theta$$

$$\beta(\theta') \geq \beta(\theta)\,\forall\,\theta' \in \Theta_1 \text{ vs. } \theta \in \Theta_0$$

$$\Lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(x)}{\sup_{\theta \in \Theta} L(x)}$$

$$R = \{x : \text{Reject } H_0 \text{ if } \lambda \leq c\}$$


Plot of Power for Different Normality Tests:
Gamma (4, 5) (sk = 1.00, ku = 4.50)
Simulated Power
1.2
1.0
0.8
0.6
0.4
0.2
0.0
10
20
30
50
100
200
300
400
500
1000
2000
Sample size, n
SW
KS
LF
AD

- Most powerful tests
- Uniformly most powerful tests
- Likelihood ratio tests
- Sequential probability ratio test
- Minimax procedure
- Classification procedure

$$f_X(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{\frac{-1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)}$$

$$X^T \Sigma^{-1} X$$

$$(X-\mu)^T \Sigma^{-1} (X-\mu)$$

$$Y = X\beta + \varepsilon$$

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$$

- Quadratic forms
- One-way ANOVA
- Noncentral chi square and F-distributions
- Multiple comparisons
- The analysis of variance
- A regression problem
- A test of independence

- The distribution of certain quadratic forms
- The independence of certain quadratic forms

STATISTICAL

INDIAN INSTITUTE

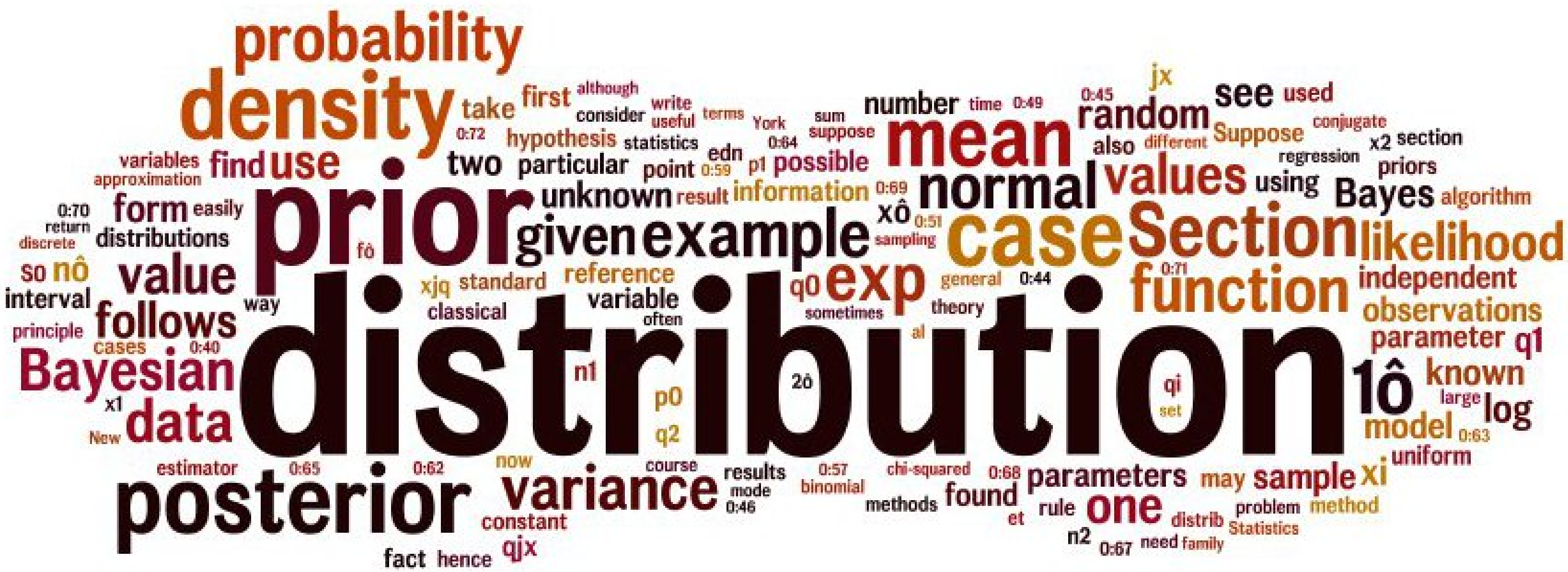
भिन्नेष्वैक्यस्य दर्शनम्

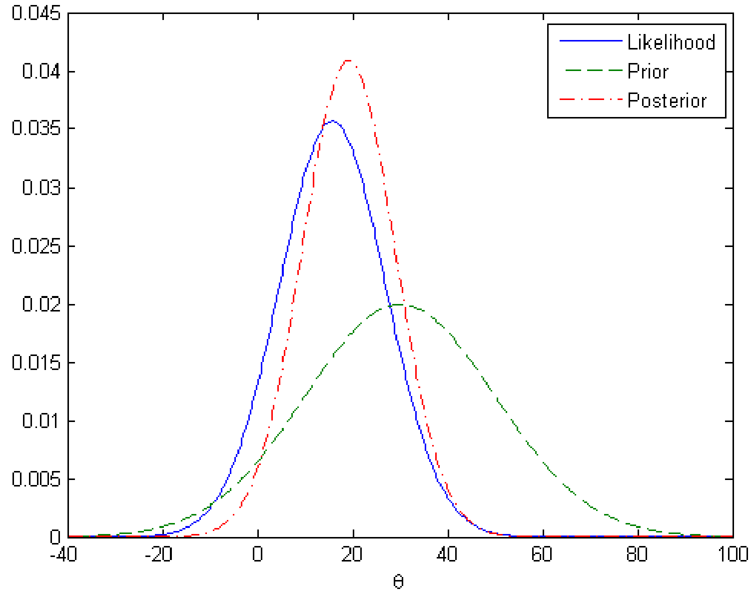
UNITY IN DIVERSITY

- Location models
- Sample median and the sign test
  - ARE
  - estimating equations base on the sign test
  - CI for median
- Signed-rank Wilcoxon
  - ARE
  - estimating equations based on Signed-rank Wilcoxon
  - CI for median

- Mann-Whitney-Wilcoxon procedure
  - ARE
  - estimating equations based on Mann-Whitney-Wilcoxon
  - CI for shift parameter Δ
- General rank scores
  - Efficacy
  - estimating equations based on general scores
  - optimization: best estimates

- Adaptive procedures
- Simple linear model
- Measures of association
  - Kendall's tau τ
  - Spearman's rho ρ
- Robust concepts
  - location model
  - linear model




Likelihood
Prior
Posterior
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
-40
-20
0
20
40
60
80
100
θ

- Subjective probability
- Bayesian procedures:
  - Prior and posterior distributions
  - Bayesian point estimates
  - Bayesian interval estimation
  - Bayesian testing procedures
  - Bayesian sequential procedures
- More Bayesian terminology and ideas
- Gibbs sampler
- Modern empirical Bayes

Appendices

- Mathematical comments
  - Regularity conditions
  - Sequences
- R-Functions
- Tables of Distributions
- List of Common Distributions
- References
- Answers to Selected Exercises