## STATISTICS

## A CONCISE MATHEMATICAL INTRODUCTION FOR STUDENTS AND SCIENTISTS

DAVID W. SCOTT Rice University Houston, Texas

December 5, 2019

To my parents, John and Nancy Scott.

## Contents

#### Contents

1	Dat	a Ana	lysis and Understanding	11
	1.1	Explo	ring the Distribution of Data	. 11
		1.1.1	Pearson's Father-Son Height Data	. 11
		1.1.2	Lord Rayleigh's Data	. 12
		1.1.3	Discussion	. 13
	1.2	Explo	ring Prediction Using Data	. 13
		1.2.1	Body and Brain Weights of Land Mammals	. 13
		1.2.2	Space Shuttle Flight 25	. 14
		1.2.3	Pearson's Father-Son Height Data Revisited	. 15
		1.2.4	Discussion	. 16
	1.3	Proble	ems	16
	1.0	1 10.01		. 10
<b>2</b>	Cla	ssical ]	Probability	17
	2.1	Exper	iments with Equally Likely Outcomes	. 17
		2.1.1	Simple Outcomes	. 17
		2.1.2	Compound Events and Set Operations	. 18
	2.2	Proba	bility Laws	. 18
		2.2.1	Union and Intersection of Events $A$ and $B$	. 19
			2.2.1.1 Case (i):	. 19
			2.2.1.2 Cases (ii) and (iii):	. 19
			2213 Case (iv):	19
		222	Conditional Probability	. 19
		2.2.2	2.2.2.1 Definition of Conditional Probability	. <u>10</u> 19
			2.2.2.1 Demittion of Conditional Probability With More Than Two	. 15
			Events	10
		<b>り</b>	Independent Events	. 19 
		$_{2.2.0}$		. 20

3

		2.2.4	Bayes Theorem	20
	<b>9</b> 9	2.2.0	rartitions and Iotal Probability	20 22
	2.0	231	With Bonlacomont	$\frac{22}{22}$
		∠.J.1 ევე	With replacement (Dormutations)	22
		2.3.2	Without Replacement (remutations)	$\frac{22}{22}$
		2.3.3	Framples	22 22
		2.3.4	Examples	$\frac{22}{22}$
	24	2.3.0	able Sets: Implications As $n \rightarrow \infty$	$\frac{22}{22}$
	2.4	2 1 1	Solocting Evon or Odd Integers	$\frac{22}{22}$
		2.4.1	Selecting Bational Varsus Irrational Numbers	$\frac{22}{22}$
	25	Z.4.Z Kolmo	selecting Rational versus mational Numbers	22
	2.0 9.6	Doliah	jitur Soniog Vorgus Danallel Networks	22
	2.0	nellau 261	Series Network	$\frac{\Delta \Delta}{22}$
		2.0.1	Parallel Network	20 22
	27	2.0.2 Problo		20 22
	2.1	1 10016	1115	20
3	Ran	dom V	Variables and Models Derived From Classical Prob-	
	abil	ity and	l Postulates	25
	3.1	Rando	m Variables and Probability Distributions: Discrete Uni-	
		form F	Sxample	25
				-0
		3.1.1	Toss of a Single Die	$\frac{1}{25}$
		3.1.1 3.1.2	Toss of a Single Die	$\frac{25}{26}$
	3.2	3.1.1 3.1.2 The U	Toss of a Single Die Toss of a Pair of Dice   Toss of a Pair of Dice Continuous Uni-	25 26
	3.2	3.1.1 3.1.2 The U: form E	Toss of a Single Die Toss of a Pair of Dice   Toss of a Pair of Dice Continuous Uni-   nivariate Probability Density Function: Continuous Uni-   Lxample Continuous Uni-	25 26 26
	3.2	3.1.1 3.1.2 The U: form E 3.2.1	Toss of a Single Die Toss of a Pair of Dice   Toss of a Pair of Dice Continuous Uni-   nivariate Probability Density Function: Continuous Uni-   Example Compute Probabilities   Using the PDF to Compute Probabilities Compute Probabilities	25 26 26 27
	3.2	3.1.1 3.1.2 The U: form E 3.2.1 3.2.2	Toss of a Single Die Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   nivariate Probability Density Function: Continuous Uni-   Example Toss of a PDF to Compute Probabilities   Using the PDF to Compute Relative Odds Toss of a PDF to Compute Relative Odds	25 26 26 27 27
	3.2 3.3	3.1.1 3.1.2 The U: form E 3.2.1 3.2.2 Summ	Toss of a Single Die Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   nivariate Probability Density Function: Continuous Uni-   Example Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Using the PDF to Compute Relative Odds Toss of a Pair of Dice   ary Statistics: Central and Non-Central Moments Toss of a Pair of Dice	25 26 26 27 27 28
	3.2 3.3	3.1.1 3.1.2 The U: form E 3.2.1 3.2.2 Summ 3.3.1	Toss of a Single DieTost of the traction of traction	25 26 26 27 27 28 28
	3.2 3.3	3.1.1 3.1.2 The U: form E 3.2.1 3.2.2 Summ 3.3.1 3.3.2	Toss of a Single Die Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   nivariate Probability Density Function: Continuous Uni-   Example Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Using the PDF to Compute Relative Odds Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Using the PDF to Compute Relative Odds Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Using the PDF to Compute Relative Odds Tos of a Pair of Dice   Toss of a Pair of Dice Tos of a Pair of Dice   Toss of a Pair of Dice Tos of a Pair of Dice   Tos of a Pair of Dice Tos of a Pair of Dice   Tos of a Pair of Dice Tos of a Pair of Dice   Tos of a Pair of Dice	25 26 27 27 28 28 28 28
	3.2 3.3	3.1.1 3.1.2 The U: form E 3.2.1 3.2.2 Summ 3.3.1 3.3.2 3.3.3	Toss of a Single DieTost of the transmission of transmission of the transmission of transmission of the transmission of trans	25 26 27 27 28 28 28 28 28
	3.2 3.3	3.1.1 3.1.2 The U: form E 3.2.1 3.2.2 Summ 3.3.1 3.3.2 3.3.3 3.3.4	Toss of a Single Die Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   nivariate Probability Density Function: Continuous Uni-   Example Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Using the PDF to Compute Relative Odds Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Using the PDF to Compute Relative Odds Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Expectation as a Linear Operator Tose of a Pair of Dice   The Variance of a Random Variable Tose of a Pair of Dice   Standardized Random Variables Tose of Dice	25 26 27 27 28 28 28 28 28 28 28
	3.2 3.3	3.1.1 3.1.2 The U: form E 3.2.1 3.2.2 Summ 3.3.1 3.3.2 3.3.3 3.3.4 3.3.5	Toss of a Single Die Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   nivariate Probability Density Function: Continuous Uni-   Example Using the PDF to Compute Probabilities   Using the PDF to Compute Relative Odds Toss of a Pair of Dice   ary Statistics: Central and Non-Central Moments Toss of a Random Variable   Expectation as a Linear Operator The Variance of a Random Variable   Standardized Random Variables Toss of a Random Variables	25 26 27 27 27 28 28 28 28 28 28 28 28
	3.2 3.3	3.1.1 3.1.2 The U: form E 3.2.1 3.2.2 Summ 3.3.1 3.3.2 3.3.3 3.3.4 3.3.5 3.3.6	Toss of a Single Die Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   nivariate Probability Density Function: Continuous Uni-   Example Using the PDF to Compute Probabilities   Using the PDF to Compute Relative Odds Toss of a Pair of Dice   ary Statistics: Central and Non-Central Moments Toss of a Random Variable   Expectation as a Linear Operator The Variance of a Random Variable   Standardized Random Variables Toss of a Random Variables   Moment Generating Function Toss of a Random Variable	25 26 27 27 28 28 28 28 28 28 28 28 28 28
	3.2 3.3	3.1.1 3.1.2 The U: form E 3.2.1 3.2.2 Summ 3.3.1 3.3.2 3.3.3 3.3.4 3.3.5 3.3.6 3.3.7	Toss of a Single Die Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   nivariate Probability Density Function: Continuous Uni-   Example Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Toss of a Pair of Dice Toss of a Pair of Dice   Using the PDF to Compute Probabilities Toss of a Pair of Dice   Using the PDF to Compute Relative Odds Toss of a Pair of Dice   ary Statistics: Central and Non-Central Moments Toss of a Expectation, Average, and Mean   Expectation as a Linear Operator Toss of a Random Variable   The Variance of a Random Variables Toss of a Random Variables   Higher Order Moments Toss of Measurement   Moment Generating Function Toss of Measurement	25 26 27 27 28 28 28 28 28 28 28 28 28 28 28 28
	3.2 3.3	3.1.1 3.1.2 The U: form E 3.2.1 3.2.2 Summ 3.3.1 3.3.2 3.3.3 3.3.4 3.3.5 3.3.6 3.3.7	Toss of a Single DieToss of a Pair of Dicemivariate Probability Density Function: Continuous Uni-ExampleUsing the PDF to Compute ProbabilitiesUsing the PDF to Compute Relative Oddsary Statistics: Central and Non-Central MomentsExpectation, Average, and MeanExpectation as a Linear OperatorThe Variance of a Random VariableHigher Order MomentsMoment Generating FunctionMoment Scales and Units of Measurement3.3.7.1	25 26 27 27 28 28 28 28 28 28 28 28 28 28 28 28 28
	3.2	3.1.1 3.1.2 The U: form H 3.2.1 3.2.2 Summ 3.3.1 3.3.2 3.3.3 3.3.4 3.3.5 3.3.6 3.3.7	Toss of a Single DieToss of a Pair of DiceToss of a Pair of Dicenivariate Probability Density Function: Continuous Uni-ExampleUsing the PDF to Compute ProbabilitiesUsing the PDF to Compute Relative Oddsary Statistics: Central and Non-Central MomentsExpectation, Average, and MeanExpectation as a Linear OperatorThe Variance of a Random VariableStandardized Random VariablesHigher Order MomentsMoment Generating Function3.3.7.1The Four Measurement Scales3.3.7.2Units of Measurement	25 26 27 27 28 28 28 28 28 28 28 28 28 28 28 28 28

 $\bigodot$  2019 D. W. Scott

	3.5	Waitin	ng Time for a Success: Geometric PMF	29
	3.6	Waitin	ng Time for $r$ Successes: Negative Binomial $\ldots \ldots \ldots$	29
	3.7	Poisso	n Process and Distribution	29
		3.7.1	Moments of the Poisson PMF	30
		3.7.2	Examples	30
	3.8	Waitin	ng Time for Poisson Events: Negative Exponential PDF	30
	3.9	The N	formal Gaussian (Also Known as the Gaussian Distribu-	
		tion)		30
		3.9.1	Standard Normal Distribution	31
		3.9.2	Sums of Independent Normal Random Variables	31
		3.9.3	Normal Approximation to the Poisson Distribution	31
	3.10	Proble	2ms	31
4	Biva	ariate	Random Variables, Transformations, and Simula-	
	tion	.s		33
	4.1	Bivaria	ate Continuous Random Variables	33
		4.1.1	Joint CDF and PDF Functions	33
		4.1.2	Marginal PDF	33
		4.1.3	Conditional Probability Density Function	34
		4.1.4	Independence of Two Random Variables	35
		4.1.5	Expectation, Correlation, and Regression	35
			4.1.5.1 Covariance and Correlation	35
			$4.1.5.2  \text{Regression Function}  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  $	35
		4.1.6	Independence of $n$ Random Variables $\ldots \ldots \ldots \ldots$	35
		4.1.7	Bivariate Normal PDF	35
		4.1.8	Correlation, Independence, and Confounding Variables	35
	4.2	Chang	ge of Variables	35
		4.2.1	Examples: Two Uniform Transformations	35
		4.2.2	One-Dimensional Transformations	36
			4.2.2.1 Example 1: Negative Exponential PDF	36
			4.2.2.2 Example 2: Cauchy PDF	36
			4.2.2.3 Example 3: Chi-Squared PDF With 1 Degree	07
		4.0.0	of Freedom	37
	4.0	4.2.3	Two-Dimensional Transformations	37
	4.3	Simula	ations	37
		4.3.1	Generating Uniform Pseudo-Random Numbers	37
			4.3.1.1 Reproducibility	37
			4.3.1.2 RANDU	37

5

	4.4	4.3.2Probability Integral Transformation374.3.3Event-Driven Simulation37Problems37					
5	Anı	Approximations and Asymptotics					
Ŭ	5.1	Why Do We Like Bandom Samples? 40					
	0.1	5.1.1 When $u(\mathbf{X})$ Takes a Product Form					
		5.1.2 When $u(\mathbf{X})$ Takes a Summation Form					
	5.2	Useful Inequalities					
	0	$5.2.1$ Markov's Inequality $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots 40$					
		$5.2.2$ Chebyshev's Inequality $\ldots \ldots \ldots \ldots \ldots \ldots \ldots 40$					
		5.2.3 Jensen's Inequality					
		5.2.4 Cauchy-Schwarz Inequality					
	5.3	Sequences of Random Variables					
		5.3.1 Weak Law of Large Numbers					
		5.3.2 Consistency of the Sample Variance					
		5.3.3 Relationships Among the Modes of Convergence 41					
		5.3.3.1 Proof of Result $(??)$					
		5.3.3.2 Proof of Result $(??)$					
	5.4	Central Limit Theorem					
		5.4.1 Moment Generating Function for Sums					
		5.4.2 Standardizing the Sum $S_n$					
		5.4.3 Proof of Central Limit Theorem					
	5.5	Delta Method and Variance-Stabilizing Transformations 41					
	5.6	Problems					
6	Par	eameter Estimation 43					
	6.1	Desirable Properties of an Estimator					
	6.2	Moments of the Sample Mean and Variance					
		6.2.1 Theoretical Mean and Variance of the Sample Mean 44					
		6.2.2 Theoretical Mean of the Sample Variance					
		6.2.3 Theoretical Variance of the Sample Variance 44					
	6.3	Method of Moments (MoM)					
	6.4	Sufficient Statistics and Data Compression					
	6.5	Bayesian Parameter Estimation					
	6.6	Maximum Likelihood Parameter Estimation 45					
		6.6.1 Relationship to Bayesian Parameter Estimation 45					
		6.6.2 Poisson MLE Example					

		6.6.3	Normal MLE Example	45
		6.6.4	Uniform MLE Example	45
	6.7	Inform	nation Inequalities and the Cramèr-Rao Lower Bound	45
		6.7.1	Score Function	45
		6.7.2	Asymptotics of the MLE	45
		6.7.3	Minimum Variance of Unbiased Estimators	45
		6.7.4	Examples	45
	6.8	Proble	ems	45
7	Hvr	othesi	is Testing	47
	7.1	Settin	g up a Hypothesis Test	47
		7.1.1	Example of a Critical Region	48
		7.1.2	Accuracy and Errors in Hypothesis Testing	48
	7.2	Best (	Critical Region for Simple Hypotheses	48
		7.2.1	Simple Example Continued	48
		7.2.2	Normal Shift Model with Common Variance	48
	7.3	Best (	Critical Region for a Composite Alternative Hypothesis .	49
		7.3.1	Negative Exponential Composite Hypothesis Test	49
			7.3.1.1 Example	49
			7.3.1.2 Alternative Critical Regions	50
			7.3.1.3 Mount St. Helens Example	51
		7.3.2	Normal Shift Model with Common But Unknown Vari-	
			ance: The $T$ -test $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	52
		7.3.3	The Random Variable $T_{n-1}$	52
			7.3.3.1 Where We Show X and $S^2$ Are Independent .	52
			7.3.3.2 Where We Show That $S^2$ Scaled Is $\chi^2(n-1)$	52
			7.3.3.3 Where We Finally Derive the $T$ PDF	52
		7.3.4	The One-Sample $T$ -test	52
		7.3.5	Example	52
		7.3.6	Other $T$ -tests	52
			$7.3.6.1  \text{Paired } T \text{-test}  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  $	52
			7.3.6.2 Two-Sample $T$ -test	52
			7.3.6.3 Example Two-Sample T-test: Lord Rayleigh's	50
	74	Danan	Data	52 52
	1.4	Repor	Frample When the Null Humothesis Is Dejected	03 52
		$\begin{array}{c} 1.4.1 \\ 7 1 9 \end{array}$	When the Null Hypothesis Is Not Dejected	00 52
		1.4.2 7 1 3	The Power Function	00 52
		1.4.0		55

	7.5	Multiple Testing and the Bonferroni Correction	54
	7.6	Problems	54
8	Cor	fidence Intervals and Other Hypothesis Tests 5	5
	8.1	Confidence Intervals	6
		8.1.1 Confidence Interval for $\mu$ : Normal Data, $\sigma^2$ Known 5	56
		8.1.2 Confidence Interval for $\mu$ : $\sigma^2$ Unknown	6
		8.1.3 Confidence Intervals and $p$ -Values	57
	8.2	Hypotheses About the Variance and the $F$ -Distribution 5	57
		8.2.1 The $F$ -Distribution	57
		8.2.2 Hypotheses About the Value of the Variance 5	68
		8.2.3 Confidence Interval for the Variance	68
		8.2.4 Two-Sided Alternative for Testing $\sigma^2 = \sigma_0^2$ 5	68
	8.3	Pearson's Chi-Squared Tests	68
		8.3.1 The Multinomial PMF	68
		8.3.2 Goodness-of-Fit (GoF) Tests	68
		8.3.3 Two-Category Binomial Case	68
		8.3.4 <i>m</i> -Category Multinomial Case $\ldots \ldots \ldots \ldots \ldots \ldots \ldots $	68
		8.3.5 Goodness-of-Fit Test for a Parametric Model 5	<i>5</i> 8
		8.3.6 Tests for Independence in Contingency Tables 5	68
	8.4	Correlation Coefficient Tests and C.I.'s	<i>6</i> 8
		8.4.1 How to Test if the Correlation $\rho = 0$	58
		8.4.2 Confidence Intervals and Tests for a General Correla-	
		tion Coefficient	59
	8.5	Linear Regression	59
		8.5.1 Least Squares Regression	59
		8.5.2 Distribution of the Least-Squares Parameters 5	59
		8.5.3 A Confidence Interval for the Slope	59
		8.5.4 A Two-Side Hypothesis Test for the Slope	59
		8.5.5 Predictions at a New Value	59
		8.5.6 Population Interval at a New Value	59
	8.6	Analysis of Variance	59
	8.7	Problems	59
0	Ton	as in Statistics	1
I	тор 0 1	MSE and Histogram Bin Width Selection	י <b>ב</b> 1
	9.1	0.1.1 MSE Criterion for Bigged Estimators	ידי 1
		0.1.2 Case Study: Optimal Histogram Bin Widths	ル 31
		5.1.2 Case Study. Optimial instogram Din Widths	1
M	arch 2	2, 2020 8 © 2019 D. W. Sco	tt

9.1.3	Examples with Normal Data	2
9.1.4	Normal Reference Rules for the Histogram Bin Width . 65	3
	9.1.4.1 Scott's Rule	3
	9.1.4.2 Freedman-Diaconis Rule	3
	9.1.4.3 Sturges' Rule	3
	9.1.4.4 Comparison of the Three Rules	3
An Op	imal Stopping Time Problem 64	4
Compo	und Random Variables	õ
9.3.1	Computing Expectations with Conditioning 65	õ
9.3.2	Sum of a Random Number of Random Variables 65	5
Simula	ion and the Bootstrap 65	õ
Multip	e Linear Regression $\ldots \ldots 66$	3
Experi	nental Design $\ldots \ldots 66$	6
Logisti	Regression, Poisson Regression, and the Generalized	
Linear	$Model \dots \dots$	7
Robust	ness	7
Conclu	sions	3
	9.1.3 9.1.4 10 10 10 10 10 10 10 10 10 10 10 10 10	9.1.3 Examples with Normal Data 65   9.1.4 Normal Reference Rules for the Histogram Bin Width 65   9.1.4.1 Scott's Rule 65   9.1.4.2 Freedman-Diaconis Rule 65   9.1.4.3 Sturges' Rule 65   9.1.4.4 Comparison of the Three Rules 65   9.1.4.4 Comparison of the Three Rules 65   9.1.4.4 Comparison of the Three Rules 65   An Optimal Stopping Time Problem 64   Compound Random Variables 65   9.3.1 Computing Expectations with Conditioning 64   9.3.2 Sum of a Random Number of Random Variables 65   Simulation and the Bootstrap 66   Multiple Linear Regression 66   Logistic Regression, Poisson Regression, and the Generalized 67   Robustness 67   Conclusions 67

© 2019 D. W. Scott

10

## Chapter 1

## Data Analysis and Understanding

### 1.1 Exploring the Distribution of Data

#### 1.1.1 Pearson's Father-Son Height Data



Figure 1.1: Displays of the father-son height data collected by Karl Pearson: (L) Box-and-Whiskers plot; (M) Stem-and Leaf plot; (R) Histogram.



Figure 1.2: Histograms of the sons' heights (top row) and fathers' heights (bottom row) using three bin widths: h/2, h, 2h from left to right; see text.

#### 1.1.2 Lord Rayleigh's Data



Figure 1.3: Displays of Lord Rayleigh's 24 measurements of the atomic weight of nitrogen gas. (L) Histogram with 4 bins; (M) A second histogram; (R) Stem-and-Leaf display using the **R** command stem(rayleigh,scale=4).

March 22, 2020

### 1.2 Exploring Prediction Using Data

### 1.2.1 Body and Brain Weights of Land Mammals



Figure 1.4: Scatter diagrams of the raw and log-transformed body and brain weights of 62 land mammals.

13

March 22, 2020





Figure 1.5: Analysis of the number of O-ring failures for the first 24 Space Shuttle launches; see text.

14

### 1.2.3 Pearson's Father-Son Height Data Revisited



Figure 1.6: Father-son height data collected by Karl Pearson

March 22, 2020

© 2019 D. W. Scott

15

### 1.2.4 Discussion

## 1.3 Problems



Figure 1.7: Box-Cox Transformation on natural and log scales

## Chapter 2

## **Classical Probability**

### 2.1 Experiments with Equally Likely Outcomes

### 2.1.1 Simple Outcomes



Figure 2.1: (L) Venn diagram of the classical probability experiment rolling a single die where n = 6. (R) Events A, B, and C are shown as ellipses enclosing the appropriate simple outcomes.

### 2.1.2 Compound Events and Set Operations

## 2.2 Probability Laws



Figure 2.2: Four possible relationships between events A and B.

March 22, 2020

18

#### **2.2.1** Union and Intersection of Events A and B

- 2.2.1.1 Case (i):
- 2.2.1.2 Cases (ii) and (iii):
- 2.2.1.3 Case (iv):
- 2.2.2 Conditional Probability
- 2.2.2.1 Definition of Conditional Probability
- 2.2.2.2 Conditional Probability With More Than Two Events



Figure 2.3: Probability that n students all have different birthdays, plotted using three different scales.

19

March 22, 2020

<sup>© 2019</sup> D. W. Scott

#### 2.2.3 Independent Events

2.2.4 Bayes Theorem

#### 2.2.5 Partitions and Total Probability



Figure 2.4: (L) Venn diagram of partition of  $\Omega$  into m = 6 sets  $A_1, A_2, \ldots, A_6$ ; (M) Set B superimposed upon partition; (R) Set B decomposed into m disjoint events using the partition. In this Figure we use the shorthand notation for intersection, namely,  $BA_i = B \cap A_i$ .

March 22, 2020

Figure 2.5: Illustration of the FPC. Select one of  $\{A, B, C, D\}$ , then one of  $\{a, b\}$ , and finally one of  $\{\alpha, \beta, \gamma\}$ . For each selection at the first step, there are 2 choices at the second step. Finally, for each of the selections after the first two steps, there are 3 choices at the third step. Therefore,  $n(S) = 4 \times 2 \times 3 = 24$  possibilities. It is common to use a tree diagram to visualize the selections on the leaves.



- 2.3 Counting Methods
- 2.3.1 With Replacement
- 2.3.2 Without Replacement (Permutations)
- 2.3.3 Without Replacement Nor Order (Combinations)
- 2.3.4 Examples
- 2.3.5 Extended Combinations (Multinomial)
- 2.4 Countable Sets: Implications As  $n \to \infty$
- 2.4.1 Selecting Even or Odd Integers
- 2.4.2 Selecting Rational Versus Irrational Numbers
- 2.5 Kolmogorov's Axioms
- 2.6 Reliability: Series Versus Parallel Networks



Figure 2.6: A series network (L) and parallel network (R) of n components.

- 2.6.1 Series Network
- 2.6.2 Parallel Network
- 2.7 Problems

© 2019 D. W. Scott

24

## Chapter 3

## Random Variables and Models Derived From Classical Probability and Postulates

- 3.1 Random Variables and Probability Distributions: Discrete Uniform Example
- 3.1.1 Toss of a Single Die



Figure 3.1: (L) The cumulative distribution function for the roll of a single die; and (R) its probability mass function.

#### 3.1.2 Toss of a Pair of Dice



Figure 3.2: (L) The cumulative distribution function for the sum of pips on two dice; and (R) its probability mass function. From the shape of the PMF, this is a **discrete isosceles triangular distribution**.

## 3.2 The Univariate Probability Density Function: Continuous Uniform Example



Figure 3.3: (L) The CDF for a Unif(0, 1) density; and (R) its PDF.

March 22, 2020

26

#### 3.2.1 Using the PDF to Compute Probabilities



Figure 3.4: The shaded areas give the probabilities of the events 1 < X < 2, |X| > 1.5, and 1 < |X| < 2, respectively.

#### 3.2.2 Using the PDF to Compute Relative Odds



Figure 3.5: The CDF and PDF of an isosceles triangular distribution.

March 22, 2020

27

- 3.3 Summary Statistics: Central and Non-Central Moments
- 3.3.1 Expectation, Average, and Mean
- 3.3.2 Expectation as a Linear Operator
- 3.3.3 The Variance of a Random Variable
- 3.3.4 Standardized Random Variables
- 3.3.5 Higher Order Moments
- 3.3.6 Moment Generating Function
- 3.3.7 Measurement Scales and Units of Measurement
- 3.3.7.1 The Four Measurement Scales
- 3.3.7.2 Units of Measurement

### 3.4 Binomial Experiments



Figure 3.6: Binomial PMF for various combinations of n and p.

```
March 22, 2020
```



Figure 3.7: Binomial CDF for n and p as in Figure 3.6.

- 3.5 Waiting Time for a Success: Geometric PMF
- 3.6 Waiting Time for *r* Successes: Negative Binomial
- 3.7 Poisson Process and Distribution



Figure 3.8: The 2 disjoint events that result in x calls in  $[0, t + \delta]$ , ignoring the very small possibility of more than 1 call in  $(t, t + \delta)$ .

29

March 22, 2020

#### 3.7.1 Moments of the Poisson PMF

### 3.7.2 Examples



Figure 3.9: Examples of the Poisson PMF, where  $X \sim Pois(m)$ .

### 3.8 Waiting Time for Poisson Events: Negative Exponential PDF

3.9 The Normal Gaussian (Also Known as the Gaussian Distribution)



Figure 3.10: Examples of the discrete Poisson PMF, Pois(m), and the continuous Normal PDF with the same moments,  $N(\mu = m, \sigma^2 = m)$ .

March 22, 2020

30



Figure 3.11: Gauss on the German Mark bill. Note the Gaussian curve.

#### 3.9.1 Standard Normal Distribution



Figure 3.12: Standard Normal CDF,  $\Phi(x)$ , and PDF,  $\phi(x)$ , for x = 1.

#### 3.9.2 Sums of Independent Normal Random Variables

- 3.9.3 Normal Approximation to the Poisson Distribution
- 3.10 Problems

March 22, 2020

31

## Chapter 4

## Bivariate Random Variables, Transformations, and Simulations

- 4.1 Bivariate Continuous Random Variables
- 4.1.1 Joint CDF and PDF Functions
- 4.1.2 Marginal PDF



Figure 4.1: Joint bivariate PMF. Each arrow displays a probability of  $\frac{1}{10}$ .

### 4.1.3 Conditional Probability Density Function



Figure 4.2: Conditional PDF,  $f_{Y|X=1}(y|1)$ , before normalization.

March 22, 2020

34

- 4.1.4 Independence of Two Random Variables
- 4.1.5 Expectation, Correlation, and Regression
- 4.1.5.1 Covariance and Correlation
- 4.1.5.2 Regression Function
- 4.1.6 Independence of *n* Random Variables
- 4.1.7 Bivariate Normal PDF
- 4.1.8 Correlation, Independence, and Confounding Variables
- 4.2 Change of Variables
- 4.2.1 Examples: Two Uniform Transformations



Figure 4.3: Transformations of a Unif(0, 1) r.v.; see text.

35

### 4.2.2 One-Dimensional Transformations



Figure 4.4: Sample transformations:  $y = x^3$  and  $x = sgn(y) \cdot |y|^{1/3}$ . The range and domain of this transformation A = B = (-1, 1).

#### 4.2.2.1 Example 1: Negative Exponential PDF

#### 4.2.2.2 Example 2: Cauchy PDF



Figure 4.5: Standard Cauchy and Normal PDF's.

March 22, 2020

- 4.2.2.3 Example 3: Chi-Squared PDF With 1 Degree of Freedom
- 4.2.3 Two-Dimensional Transformations
- 4.3 Simulations
- 4.3.1 Generating Uniform Pseudo-Random Numbers
- 4.3.1.1 Reproducibility
- 4.3.1.2 RANDU
- 4.3.2 Probability Integral Transformation



Figure 4.6: Generic PIT diagram. The red strip represents the event  $Y \leq y$  while the blue strip represents the equivalent event  $X \leq F_X^{-1}(y)$ .

4.3.3 Event-Driven Simulation

### 4.4 Problems

## Chapter 5

## Approximations and Asymptotics

- 5.1 Why Do We Like Random Samples?
- 5.1.1 When  $u(\mathbf{X})$  Takes a Product Form
- 5.1.2 When u(X) Takes a Summation Form
- 5.2 Useful Inequalities
- 5.2.1 Markov's Inequality
- 5.2.2 Chebyshev's Inequality
- 5.2.3 Jensen's Inequality<sup>1</sup>



Figure 5.1: Example of a convex function g(x) with two red tangent line segments touching the curve at the black points. A line segment connecting the curve at  $x = x_1$  and  $x = x_2$  is drawn in red; see text.

Marchis22cct2012 may be omitted at a first feading.

- 5.2.4 Cauchy-Schwarz Inequality
- 5.3 Sequences of Random Variables
- 5.3.1 Weak Law of Large Numbers
- 5.3.2 Consistency of the Sample Variance
- 5.3.3 Relationships Among the Modes of Convergence
- **5.3.3.1 Proof of Result** (??)
- **5.3.3.2** Proof of Result  $(??)^2$
- 5.4 Central Limit Theorem
- 5.4.1 Moment Generating Function for Sums
- **5.4.2** Standardizing the Sum  $S_n$
- 5.4.3 Proof of Central Limit Theorem
- 5.5 Delta Method and Variance-Stabilizing Transformations
- 5.6 Problems

<sup>&</sup>lt;sup>2</sup>This section may be omitted at a first reading.

Chapter 6

## **Parameter Estimation**



Figure 6.1: Nine examples of possible Normal fits to a random sample of 50 points. In the first row, the data are displayed using the **R** function rug(x). In the second row, probability histograms hist(x,prob=T) are displayed.

- 6.1 Desirable Properties of an Estimator
- 6.2 Moments of the Sample Mean and Variance
- 6.2.1 Theoretical Mean and Variance of the Sample Mean
- 6.2.2 Theoretical Mean of the Sample Variance
- 6.2.3 Theoretical Variance of the Sample Variance
- 6.3 Method of Moments (MoM)
- 6.4 Sufficient Statistics and Data Compression
- 6.5 Bayesian Parameter Estimation



Figure 6.2: From left to right: a histogram of the 41 data points; the Beta(5,5) prior PDF; and the posterior Beta(593,442) PDF.

44

- 6.6 Maximum Likelihood Parameter Estimation
- 6.6.1 Relationship to Bayesian Parameter Estimation
- 6.6.2 Poisson MLE Example
- 6.6.3 Normal MLE Example
- 6.6.4 Uniform MLE Example
- 6.7 Information Inequalities and the Cramèr-Rao Lower Bound
- 6.7.1 Score Function
- 6.7.2 Asymptotics of the MLE
- 6.7.3 Minimum Variance of Unbiased Estimators
- 6.7.4 Examples
- 6.8 Problems

## Chapter 7

## Hypothesis Testing

### 7.1 Setting up a Hypothesis Test



Figure 7.1: (L) For two roof construction techniques, hypothetical PDF's  $\phi(x|70, 10^2)$  and  $\phi(x|120, 15^2)$  of the minimum wind speed incurred that resulted in roof damage during a hurricane. (R) Illustration of a possible hypothesis-testing decision region for a small sample of n = 2 roofs. Contours of the two bivariate sampling PDF's are shown in green.

- 7.1.1 Example of a Critical Region
- 7.1.2 Accuracy and Errors in Hypothesis Testing
- 7.2 Best Critical Region for Simple Hypotheses
- 7.2.1 Simple Example Continued

#### 7.2.2 Normal Shift Model with Common Variance



Figure 7.2: Critical regions based upon  $\bar{X} > k'$  for testing two shifted Normal PDF's with  $\mu_0 = 0$ ,  $\mu_1 = 10$ , and common  $\sigma = 24$ . (L) n = 9; (R) n = 64. The type I and II errors are shown in red and blue, respectively. The underlying sampling densities and means are shown in green; the densities for  $\bar{X}$  are shown in black.

March 22, 2020

- 7.3 Best Critical Region for a Composite Alternative Hypothesis
- 7.3.1 Negative Exponential Composite Hypothesis Test

#### 7.3.1.1 Example



Figure 7.3: (L) The log-likelihood ratio for a sample of n = 8 negative exponential r.v.'s with  $\beta_0 = 1$ . The levels corresponding to 5% and 1% type I errors are shown. (R) The  $Gamma(8, \beta = 1)$  PDF of  $S_8$ , together with the 95% probability interval (3.62, 14.98). The shaded tail areas have mass 3.176% and 1.824%, respectively, totaling 5%.

March 22, 2020

49





Figure 7.4: Alternative 95% confidence level tests for our example. (L) (3.45, 14.42) has equal tail probabilities of 2.5%; (R) (2.97, 13.63) is the narrowest interval. The tail areas are 1.824% and 3.176%, respectively.

#### 7.3.1.3 Mount St. Helens Example



Figure 7.5: (L) Topo map; (R) Earthquake epicenters.



Figure 7.6: Histograms of times between eruptions (in days) for all 247 eruptions (left frame), the first 147 eruptions (middle frame), and last 100 eruptions (right frame). The blue line depicts a negative exponential fit.

51

March 22, 2020

- 7.3.2 Normal Shift Model with Common But Unknown Variance: The *T*-test
- 7.3.3 The Random Variable  $T_{n-1}$
- 7.3.3.1 Where We Show  $\bar{X}$  and  $S^2$  Are Independent
- **7.3.3.2** Where We Show That  $S^2$  Scaled Is  $\chi^2(n-1)$
- 7.3.3.3 Where We Finally Derive the T PDF
- 7.3.4 The One-Sample *T*-test
- 7.3.5 Example
- 7.3.6 Other *T*-tests
- 7.3.6.1 Paired T-test
- 7.3.6.2 Two-Sample *T*-test
- 7.3.6.3 Example Two-Sample T-test: Lord Rayleigh's Data

![](_page_51_Figure_11.jpeg)

Figure 7.7: Fits to Lord Rayleigh's Data Under the Null and Alternative Hypotheses. A Nobel Prize was awarded for understanding this diagram.

52

- 7.4 Reporting Results: *p*-values and Power
- 7.4.1 Example When the Null Hypothesis Is Rejected
- 7.4.2 When the Null Hypothesis Is Not Rejected
- 7.4.3 The Power Function

![](_page_52_Figure_4.jpeg)

Figure 7.8: (L) Power function with n = 16. The critical region is shown in red, along with two values of the power function at  $\mu = 7.25$  and 14.0. Their complements are examples of type II errors. (R) Effect of sample size on the power function.

53

### 7.5 Multiple Testing and the Bonferroni Correction

### 7.6 Problems

![](_page_53_Figure_2.jpeg)

Figure 7.9: Draft lottery numbers 1–366 by month. The monthly average is the blue line. The overall average of 183.5 is the red dotted line.

## Chapter 8

## Confidence Intervals and Other Hypothesis Tests

### 8.1 Confidence Intervals

8.1.1 Confidence Interval for  $\mu$ : Normal Data,  $\sigma^2$  Known

![](_page_55_Figure_4.jpeg)

![](_page_55_Figure_5.jpeg)

Figure 8.1: (Left frames) A simulation study showing 100 95% confidence intervals for samples of size n = 16 from the N(0,1) PDF, where the intervals that fail to include the true value of  $\mu_0 = 0$  are shown in red. The bottom left frame shows the same 100 confidence intervals sorted for clarity. (Right frames) A simulation study from the alternative hypothesis PDF M(007493, 2) Chosen so that the powfor is 80%. In fact 22049 the W00SCdfs incorrectly include the null hypothesis mean 0.

8.1.3 Confidence Intervals and *p*-Values

# 8.2 Hypotheses About the Variance and the *F*-Distribution

8.2.1 The *F*-Distribution

![](_page_56_Figure_3.jpeg)

Figure 8.2: (L) Examples of the  $F_{10,s}$  PDF for  $1 \le s \le 500$ . (R) Examples of the  $F_{r,100}$  PDF for  $1 \le r \le 500$ .

March 22, 2020

- 8.2.2 Hypotheses About the Value of the Variance
- 8.2.3 Confidence Interval for the Variance
- 8.2.4 Two-Sided Alternative for Testing  $\sigma^2 = \sigma_0^2$
- 8.3 Pearson's Chi-Squared Tests
- 8.3.1 The Multinomial PMF
- 8.3.2 Goodness-of-Fit (GoF) Tests
- 8.3.3 Two-Category Binomial Case
- 8.3.4 *m*-Category Multinomial Case
- 8.3.5 Goodness-of-Fit Test for a Parametric Model
- 8.3.6 Tests for Independence in Contingency Tables
- 8.4 Correlation Coefficient Tests and C.I.'s
- 8.4.1 How to Test if the Correlation  $\rho = 0$

![](_page_57_Figure_12.jpeg)

Figure 8.3: Critical values for R as the sample size increases.

58

- 8.4.2 Confidence Intervals and Tests for a General Correlation Coefficient
- 8.5 Linear Regression
- 8.5.1 Least Squares Regression
- 8.5.2 Distribution of the Least-Squares Parameters
- 8.5.3 A Confidence Interval for the Slope
- 8.5.4 A Two-Side Hypothesis Test for the Slope
- 8.5.5 Predictions at a New Value
- 8.5.6 Population Interval at a New Value
- 8.6 Analysis of Variance

![](_page_58_Figure_9.jpeg)

Figure 8.4: Ten thousand simulations of the *F*-statistic (??) with K = 4,  $n_k = 25$ , and n = 100 under the null hypothesis.  $H_0: \mu_k = \mu_0, k = 1, 2, 3, 4$ .

#### 8.7 Problems

March 22, 2020

 $\bigodot$  2019 D. W. Scott

60

## Chapter 9

## **Topics in Statistics**

- 9.1 MSE and Histogram Bin Width Selection
- 9.1.1 MSE Criterion for Biased Estimators
- 9.1.2 Case Study: Optimal Histogram Bin Widths

![](_page_60_Figure_5.jpeg)

Figure 9.1: For a frequency histogram, the notation used to denote the locations of the bins  $\{B_k\}$ , the bin counts  $\{\nu_k\}$ , and the bin edges  $\{t_k\}$ .

### 9.1.3 Examples with Normal Data

![](_page_61_Figure_1.jpeg)

Figure 9.2: For several sample sizes and the N(0, 1) density, the histogram IMSE curves as the bin width varies on a log-log scale. The red dots locate the best  $h = h_n^*$ .

62

March 22, 2020

- 9.1.4 Normal Reference Rules for the Histogram Bin Width
- 9.1.4.1 Scott's Rule
- 9.1.4.2 Freedman-Diaconis Rule
- 9.1.4.3 Sturges' Rule
- 9.1.4.4 Comparison of the Three Rules

![](_page_62_Figure_5.jpeg)

Figure 9.3: Comparison of the number of bins recommended by Sturges'  $(\bullet)$ , Scott's  $(\bullet)$ , and FD's  $(\bullet)$  Rules for a Beta(5,5) PDF plotted against the exact IMSE values.

![](_page_62_Figure_7.jpeg)

Figure 9.4: Three Histograms of a Beta(5,5) sample with  $n = 2^{21}$ .

63

### 9.2 An Optimal Stopping Time Problem

![](_page_63_Figure_1.jpeg)

Figure 9.5: (L) Optimal stopping point m as a function of the population size n compared to a straight line with slope 1/e; (R) Probability of selecting the best candidate using  $m^*$  compared to p = 1/e.

March 22, 2020

64

### 9.3 Compound Random Variables

- 9.3.1 Computing Expectations with Conditioning
- 9.3.2 Sum of a Random Number of Random Variables
- 9.4 Simulation and the Bootstrap

![](_page_64_Figure_4.jpeg)

Figure 9.6: Simulation and bootstrap analysis of the sample mean and sample median for a N(5, 1) PDF with n = 101 points; see text. Histograms and smoothed histograms are displayed.

March 22, 2020

© 2019 D. W. Scott

65

### 9.5 Multiple Linear Regression

### 9.6 Experimental Design

![](_page_65_Figure_2.jpeg)

Figure 9.7: Surface of predicted average mold strength as a function of the two predictor variables. The contours and curves are shown as dotted lines where the standard deviation of the prediction exceeds 2.5.

March 22, 2020

9.7 Logistic Regression, Poisson Regression, and the Generalized Linear Model

![](_page_66_Figure_1.jpeg)

Figure 9.8: (L) Poisson and (R) Logistic regression models fitted by the **R** function glm to the space shuttle data (red line). Fifty bootstrap fits are superimposed (green lines).

### 9.8 Robustness

![](_page_66_Figure_4.jpeg)

Figure 9.9: (a) n = 9 points satisfying y = 1 + 1.5x with one outlier at x = 1; (b) n = 101 with 40 randomly selected outliers (with negative slope).

67

March 22, 2020

![](_page_67_Figure_0.jpeg)

Figure 9.10: (a) MLE  $N(\bar{x}, s^2)$  and (b)  $L_2E$  Normal fits to a random sample of 400 points from the mixture  $0.75 \times N(0, 1) + 0.25 \times N(3, 1/3^2)$ ; see text.

## 9.9 Conclusions