Probability and Statistics with Examples using R

Siva Athreya, Deepayan Sarkar, and Steve Tanner

April 25, 2016

Version: – April 25, 2016

CONTENTS

Preface v

1 BASIC CONCEPTS 1
1.1 Definitions and Properties . . . 1
1.1.1 Definitions . . . 1
1.1.2 Basic Properties . . . 2
1.2 Equally Likely Outcomes . . . 8
1.3 Conditional Probability and Bayes' Theorem . . . 12
1.3.1 Bayes' Theorem . . . 17
1.4 Independence . . . 20
1.5 Using R for computation . . . 24

2 SAMPLING AND REPEATED TRIALS 29
2.1 Bernoulli Trials . . . 29
2.1.1 Using R to compute probabilities . . . 34
2.2 Poisson Approximation . . . 37
2.3 Sampling With and Without Replacement . . . 42
2.3.1 The Hypergeometric Distribution . . . 43
2.3.2 Hypergeometric Distributions as a Series of Dependent Trials . . . 43
2.3.3 Binomial Approximation to the Hypergeometric Distribution . . . 45

3 DISCRETE RANDOM VARIABLES 49
3.1 Random Variables as Functions . . . 49
3.1.1 Common Distributions . . . 52
3.2 Independent and Dependent Variables . . . 54
3.2.1 Independent Variables . . . 55
3.2.2 Conditional, Joint, and Marginal Distributions . . . 56
3.2.3 Memoryless Property of the Geometric Random Variable . . . 59
3.2.4 Multinomial Distributions . . . 60
3.3 Functions of Random Variables . . . 63
3.3.1 Distribution of $f(X)$ and $f(X_1, X_2, \ldots, X_n)$ . . . 63
3.3.2 Functions and Independence . . . 67

4 SUMMARIZING DISCRETE RANDOM VARIABLES 71
4.1 Expected Value . . . 71
4.1.1 Properties of the Expected Value . . . 73
4.1.2 Expected Value of a Product . . . 75
4.1.3 Expected Values of Common Distributions . . . 76
4.1.4 Expected Value of $f(X_1, X_2, \ldots, X_n)$ . . . 80
4.2 Variance and Standard Deviation . . . 84
4.2.1 Properties of Variance and Standard Deviation . . . 85
4.2.2 Variances of Common Distributions . . . 87
4.2.3 Standardized Variables . . . 89
4.3 Standard Units . . . 93

4.3.1 Markov and Chebyshev Inequalities . . . 94
4.4 Conditional Expectation and Conditional Variance . . . 97
4.5 Covariance and Correlation . . . 102
4.5.1 Covariance . . . 103
4.5.2 Correlation . . . 105
4.6 Exchangeable Random Variables . . . 107

5 CONTINUOUS PROBABILITIES AND RANDOM VARIABLES 111
5.1 Uncountable Sample Spaces and Densities . . . 111
5.1.1 Probability Densities on $\mathbb{R}$ . . . 113
5.2 Continuous Random Variables . . . 117
5.2.1 Common Distributions . . . 119
5.2.2 A word about individual outcomes . . . 125
5.3 Transformation of Continuous Random Variables . . . 130
5.4 Multiple Continuous Random Variables . . . 137
5.4.1 Marginal Distributions . . . 140
5.4.2 Independence . . . 141
5.4.3 Conditional Density . . . 143
5.5 Functions of Independent Random variables . . . 148
5.5.1 Distributions of Sums of Independent Random variables . . . 149
5.5.2 Distributions of Quotients of Independent Random varibles. . . . 153

6 SUMMARISING CONTINUOUS RANDOM VARIABLES 161
6.1 Expectation, and Variance . . . 161
6.2 Covariance, Correlation, Conditional Expectation and Conditional Variance . . . 168
6.3 Moment Generating Functions . . . 176
6.4 Bivariate Normals . . . 181

7 SAMPLING AND DESCRIPTIVE STATISTICS 187
7.1 The empirical distribution . . . 187
7.2 Descriptive Statistics . . . 188
7.2.1 Sample Mean . . . 188
7.2.2 Sample Variance . . . 189
7.2.3 Sample proportion . . . 189
7.3 Simulation . . . 191
7.4 Plots . . . 194
7.4.1 Empirical Distribution Plot for Discrete Distributions . . . 195
7.4.2 Histograms for Continuous Distributions . . . 197
7.4.3 Hanging Rootograms for Comparing with Theoretical Distributions . . . 197
7.4.4 Q-Q Plots for Continuous Distributions . . . 200

8 SAMPLING DISTRIBUTIONS AND LIMIT THEOREMS 203
8.1 Multi-dimensional continuous random variables . . . 203
8.1.1 Order Statistics and their Distributions . . . 205
8.1.2 χ^2, F and t . . . 208
8.1.3 Distribution of Sampling Statistics from a Normal population . . . 210
8.2 Weak Law of Large Numbers . . . 214
8.3 Convergence in Distribution . . . 216
8.4 Central Limit Theorem . . . 218

8.4.1 Normal Approximation and Continuity Correction 220

9 ESTIMATION AND HYPOTHESIS TESTING 225
9.1 Notations and Terminology for Estimators 225
9.2 Method of Moments 226
9.3 Maximum Likelihood Estimate 227
9.4 Confidence Intervals 229
9.4.1 Confidence Intervals when the standard deviation σ is known 229
9.4.2 Confidence Intervals when the standard deviation σ is unknown 230
9.5 Hypothesis Testing 231
9.5.1 The z-test: Test for sample mean when σ is known 231
9.5.2 The t-test: Test for sample mean when σ is unknown 233
9.5.3 A critical value approach 234
9.5.4 The χ^2-test : Test for sample variance 234
9.5.5 The two-sample z-test: Test to compare sample means 236
9.5.6 The F-test: Test to compare sample variances. 237
9.5.7 A χ^2-test for "goodness of fit" 237

10 LINEAR REGRESSION 241
10.1 Sample Covariance and Correlation 241
10.2 Simple Linear Model 241
10.3 The Least Squares Line 242
10.4 a and b as Random Variables 244
10.5 Predicting New Data When σ^2 is Known 247
10.6 Hypothesis Testing and Regression 249
10.7 Estimaing an Unknown σ^2 249

A WORKING WITH DATA IN R 253
A.1 Datasets in R 253
A.2 Plotting data 254

B SOME MATHEMATICAL DETAILS 255
B.1 Linear Algebra 255
B.2 Jacobian Method 255
B.3 χ^2-goodness of fit test 255

C STRONG LAW OF LARGE NUMBERS 257

D TABLES 261