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.

 © 2019 D. W. Scott

Contents

 © 2019 D. W. Scott

 © 2019 D. W. Scott

 © 2019 D. W. Scott

 © 2019 D. W. Scott

 © 2019 D. W. Scott

 © 2019 D. W. Scott

 © 2019 D. W. Scott

Chapter 1

Data Analysis and Understanding

1.1 Exploring the Distribution of Data

1.1.1 Pearson's Father-Son Height Data

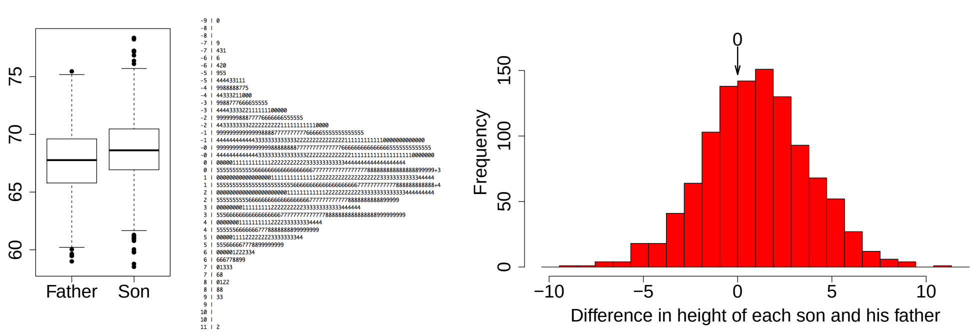


Figure 1.1: Displays of the father-son height data collected by Karl Pearson: (L) Box-and-Whiskers plot; (M) Stem-and Leaf plot; (R) Histogram.

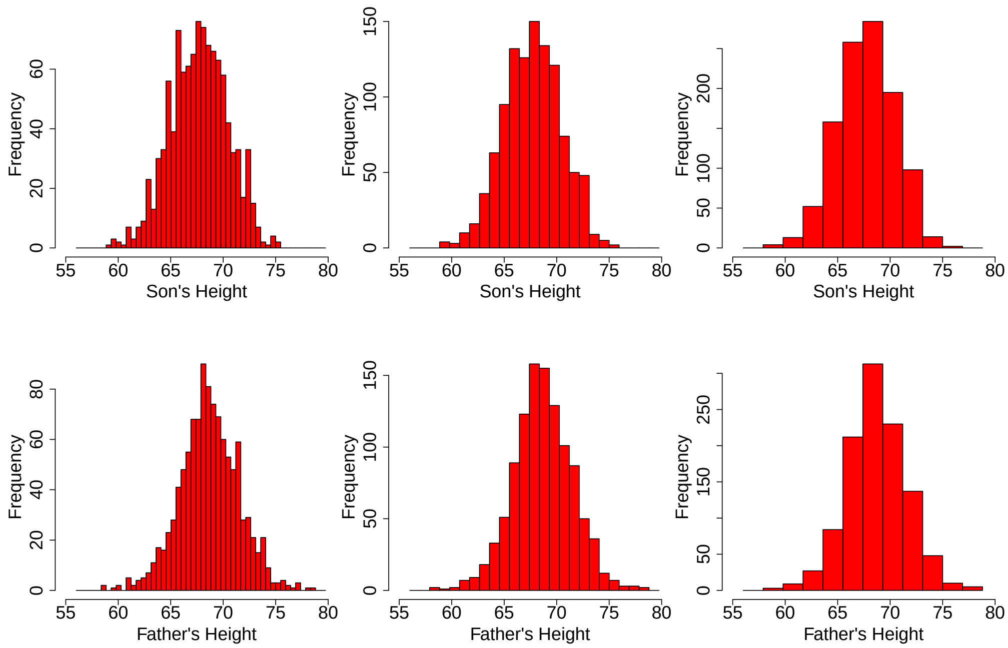
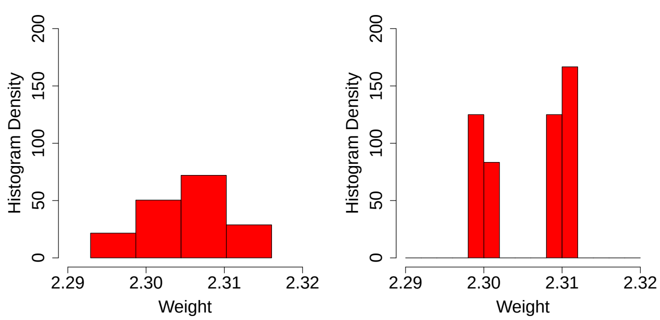


Figure 1.2: Histograms of the sons' heights (top row) and fathers' heights (bottom row) using three bin widths: $h/2$, h, $2h$ from left to right; see text.

1.1.2 Lord Rayleigh's Data




```
2298 | 25799
2299 | 4
2300 | 57
2301 | 48
2302 |
2303 |
2304 |
2305 |
2306 |
2307 |
2308 |
2309 | 69
2310 | 01112223334
2311 | 6
```

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)`.

 © 2019 D. W. Scott

1.1.3 Discussion

1.2 Exploring Prediction Using Data

1.2.1 Body and Brain Weights of Land Mammals

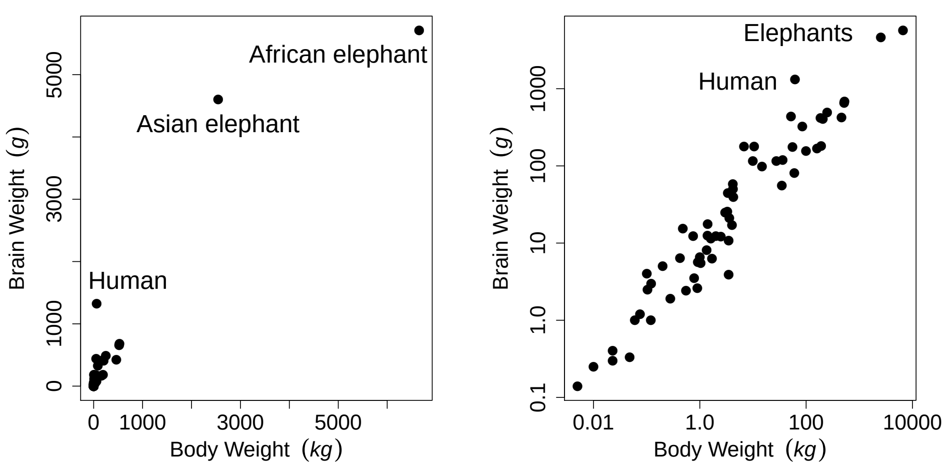


Figure 1.4: Scatter diagrams of the raw and log-transformed body and brain weights of 62 land mammals.

 © 2019 D. W. Scott

1.2.2 Space Shuttle Flight 25

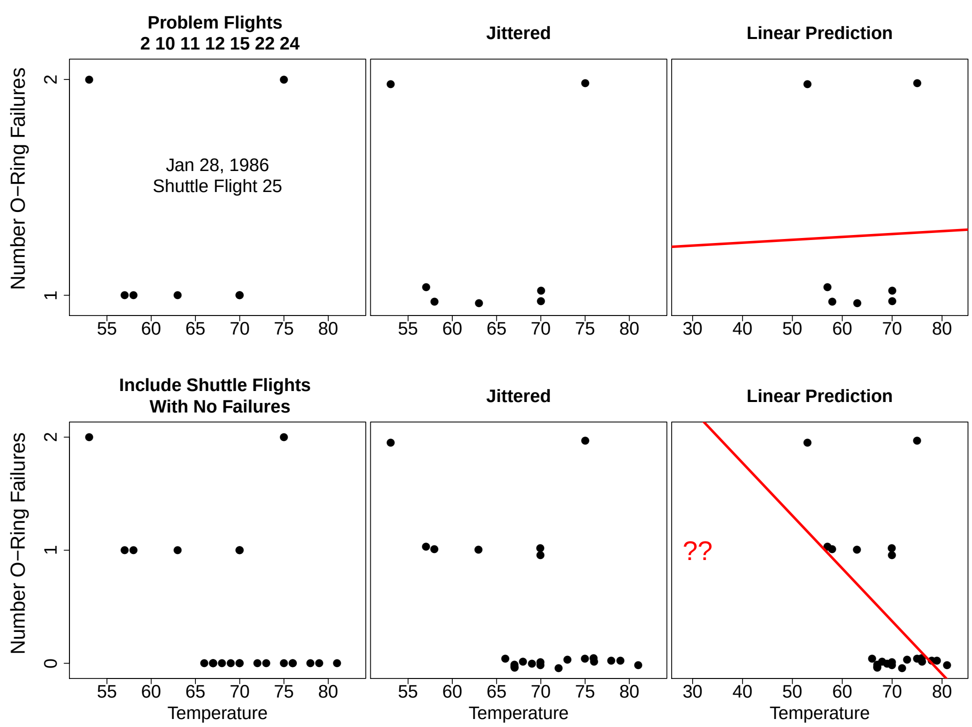


Figure 1.5: Analysis of the number of O-ring failures for the first 24 Space Shuttle launches; see text.

1.2.3 Pearson's Father-Son Height Data Revisited

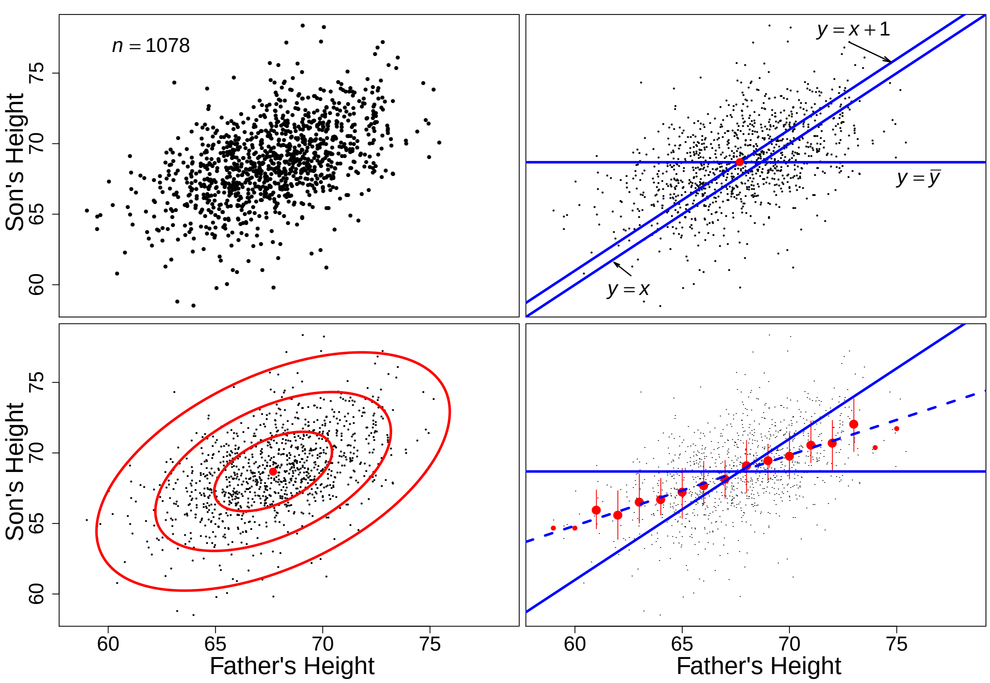


Figure 1.6: Father-son height data collected by Karl Pearson

 © 2019 D. W. Scott

1.2.4 Discussion

1.3 Problems

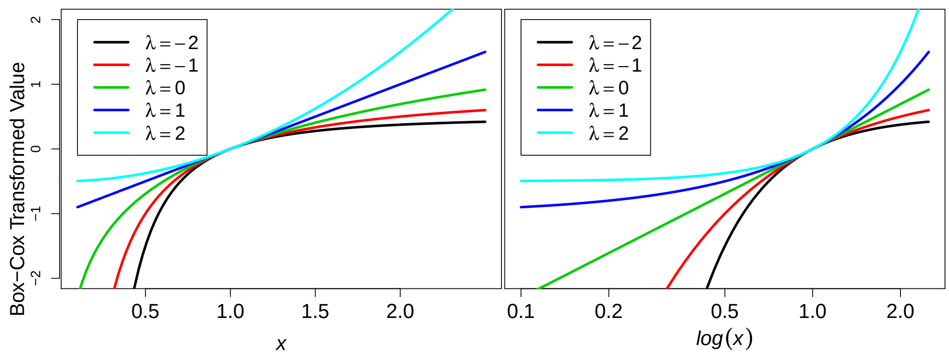


Figure 1.7: Box-Cox Transformation on natural and log scales

Chapter 2

Classical Probability

2.1 Experiments with Equally Likely Outcomes

2.1.1 Simple Outcomes

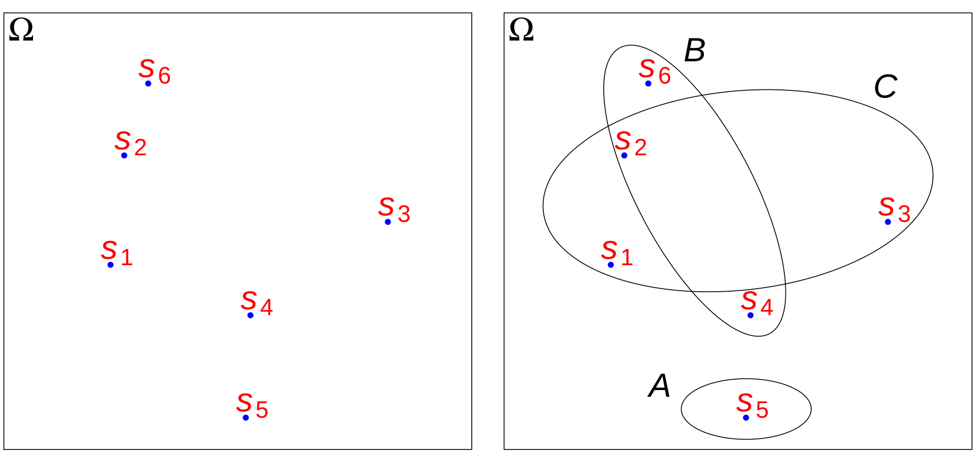


Figure 2.1: (L) Venn diagram of the classical probability experiment rolling a single die where $n = 6$. (R) Events A, B, and C are shown as ellipses enclosing the appropriate simple outcomes.

2.1.2 Compound Events and Set Operations

2.2 Probability Laws

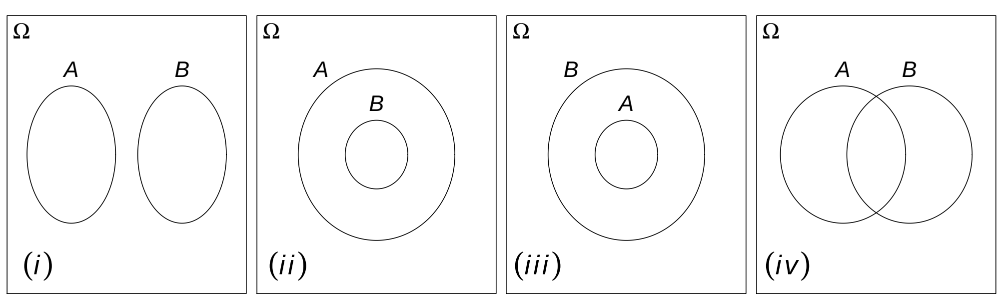


Figure 2.2: Four possible relationships between events A and B.

 © 2019 D. W. Scott

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

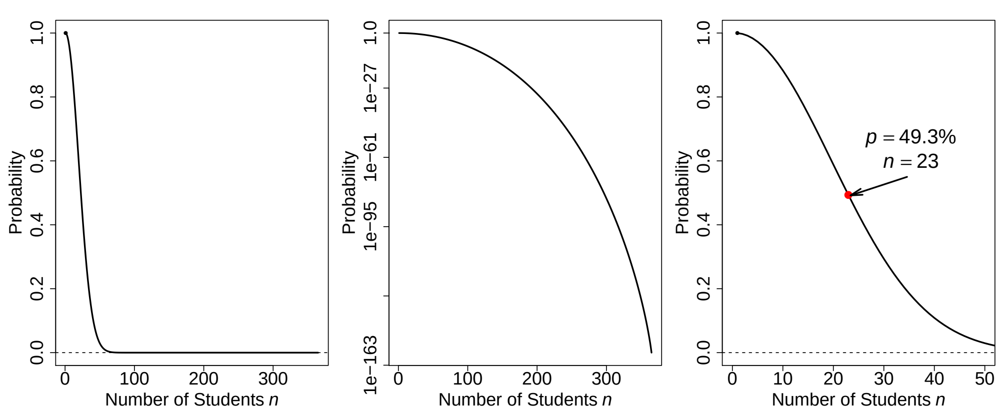


Figure 2.3: Probability that n students all have different birthdays, plotted using three different scales.

 © 2019 D. W. Scott

2.2.3 Independent Events

2.2.4 Bayes Theorem

2.2.5 Partitions and Total Probability

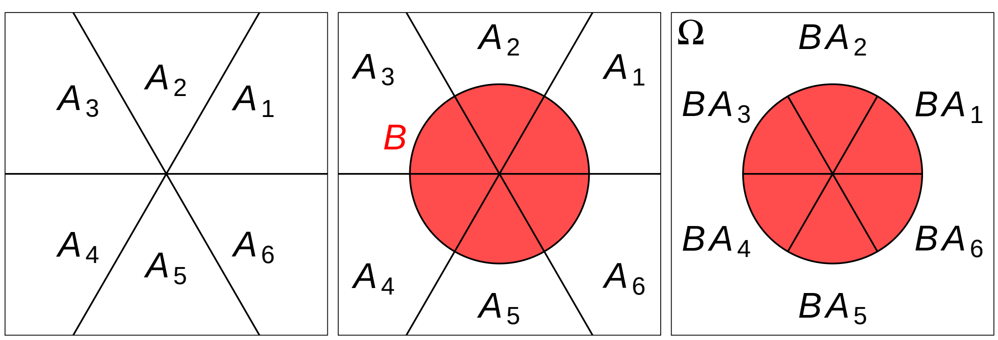
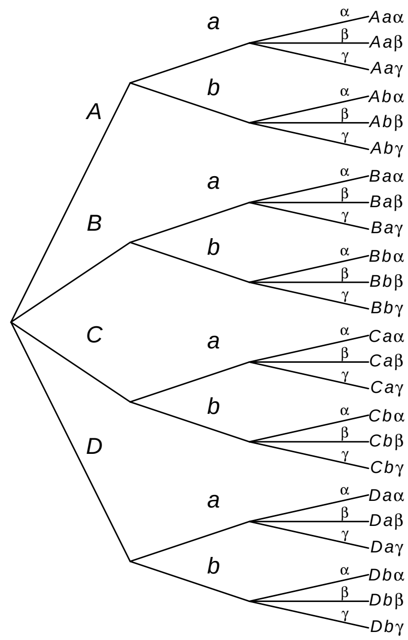


Figure 2.4: (L) Venn diagram of partition of Ω into $m = 6$ sets $A_1, A_2, \ldots, A_6$; (M) Set B superimposed upon partition; (R) Set B decomposed into m disjoint events using the partition. In this Figure we use the shorthand notation for intersection, namely, $BA_i = B \cap A_i$.

 © 2019 D. W. Scott

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 selection process, recording 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)

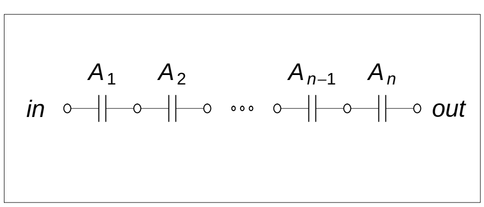
2.4 Countable Sets: Implications As $n \to \infty$

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



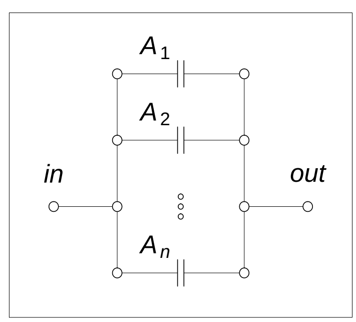


Figure 2.6: A series network (L) and parallel network (R) of n components.

2.6.1 Series Network

2.6.2 Parallel Network

2.7 Problems

 © 2019 D. W. Scott

 © 2019 D. W. Scott

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

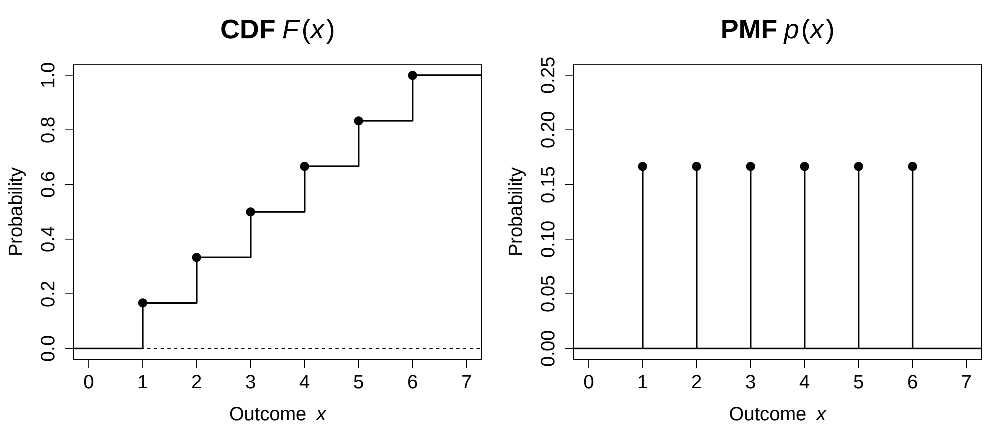


Figure 3.1: (L) The cumulative distribution function for the roll of a single die; and (R) its probability mass function.

3.1.2 Toss of a Pair of Dice

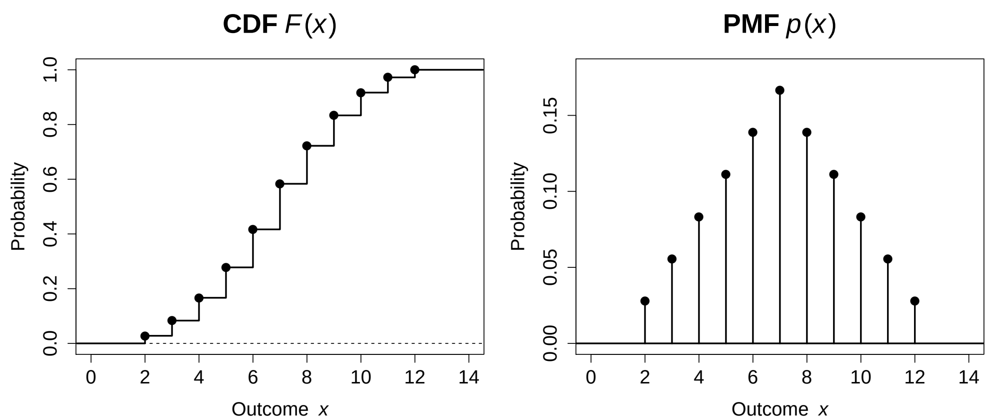


Figure 3.2: (L) The cumulative distribution function for the sum of pips on two dice; and (R) its probability mass function. From the shape of the PMF, this is a **discrete isosceles triangular distribution**.

3.2 The Univariate Probability Density Function: Continuous Uniform Example

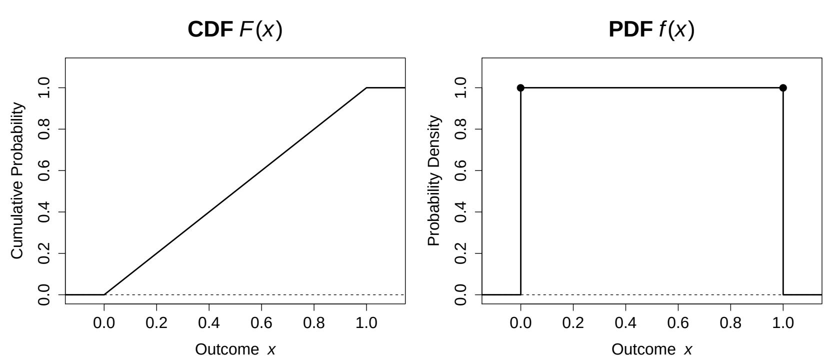


Figure 3.3: (L) The CDF for a $Unif(0,1)$ density; and (R) its PDF.

 © 2019 D. W. Scott

3.2.1 Using the PDF to Compute Probabilities

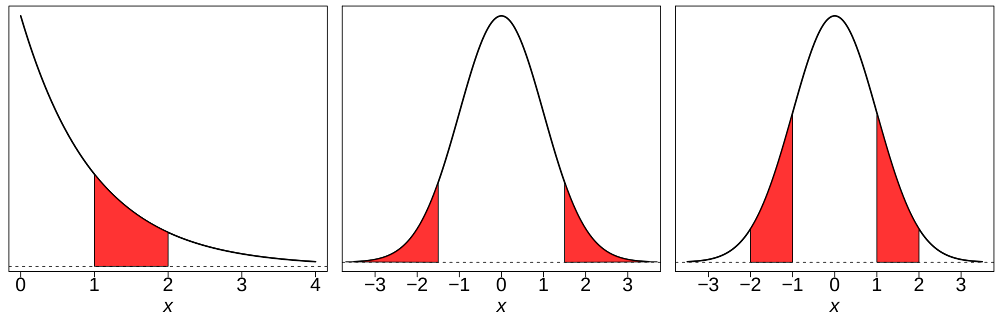


Figure 3.4: The shaded areas give the probabilities of the events $1 < X < 2$, $|X| > 1.5$, and $1 < |X| < 2$, respectively.

3.2.2 Using the PDF to Compute Relative Odds

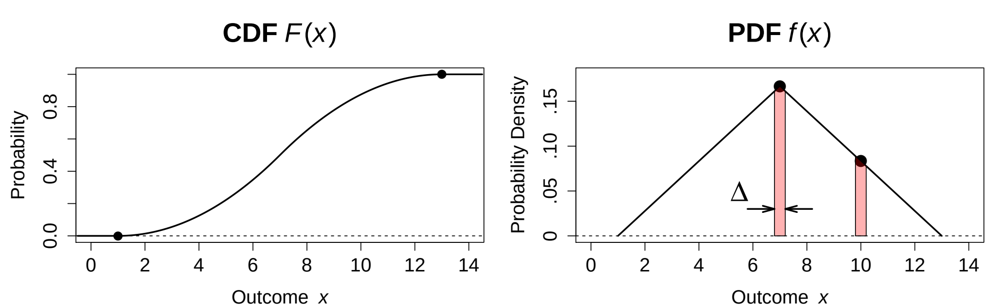


Figure 3.5: The CDF and PDF of an isosceles triangular distribution.

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

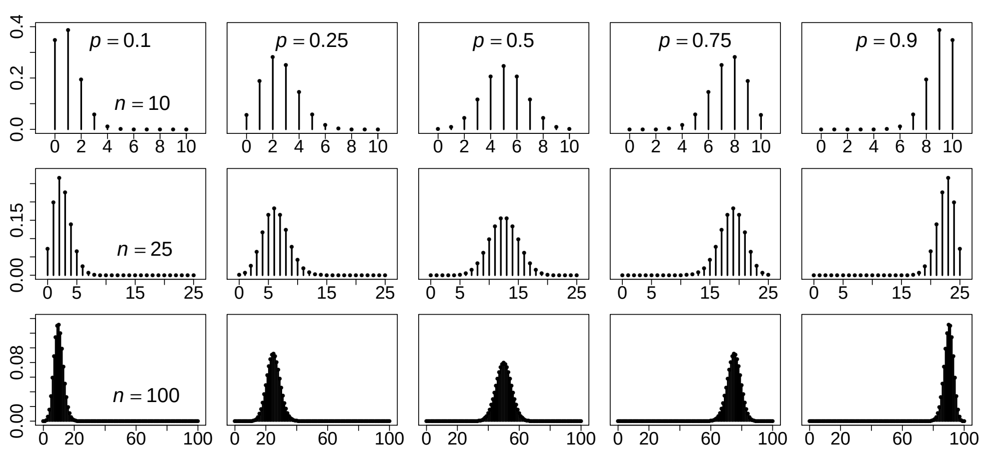


Figure 3.6: Binomial PMF for various combinations of n and p.

 © 2019 D. W. Scott

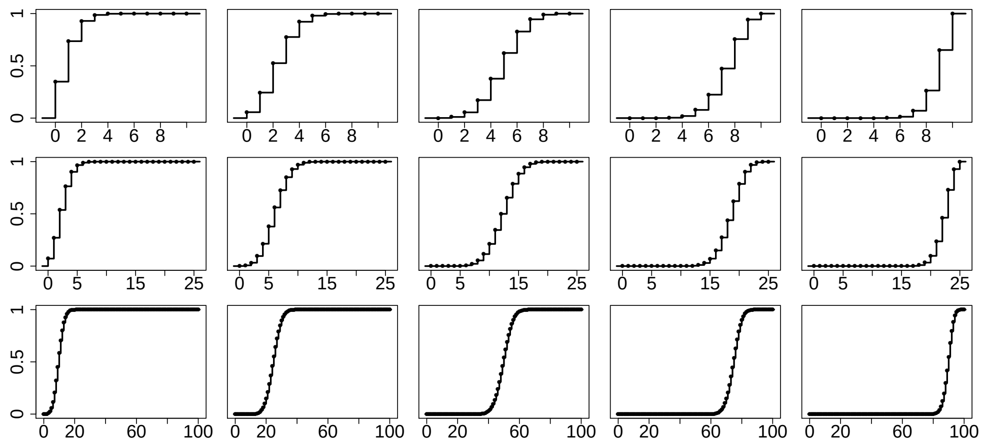


Figure 3.7: Binomial CDF for n and p as in Figure 3.6.

3.5 Waiting Time for a Success: Geometric PMF

3.6 Waiting Time for r Successes: Negative Binomial

3.7 Poisson Process and Distribution

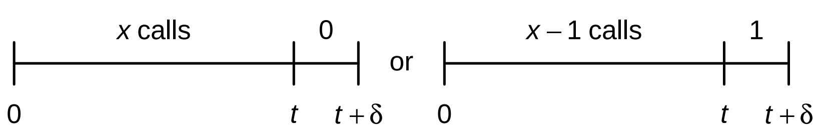


Figure 3.8: The 2 disjoint events that result in x calls in $[0, t + \delta]$, ignoring the very small possibility of more than 1 call in $(t, t + \delta)$.

 © 2019 D. W. Scott

3.7.1 Moments of the Poisson PMF

3.7.2 Examples

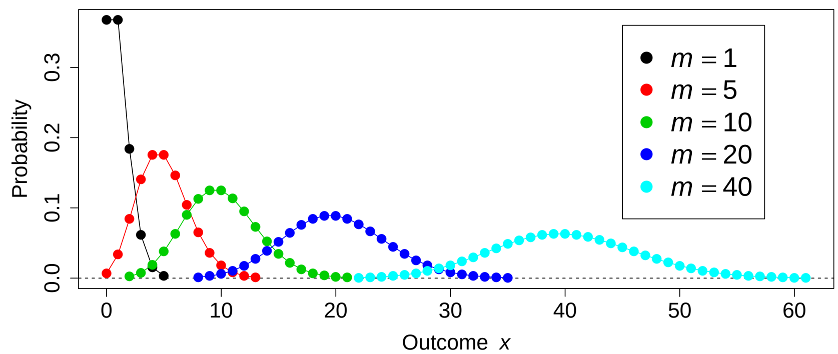


Figure 3.9: Examples of the Poisson PMF, where $X \sim Pois(m)$.

3.8 Waiting Time for Poisson Events: Negative Exponential PDF

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

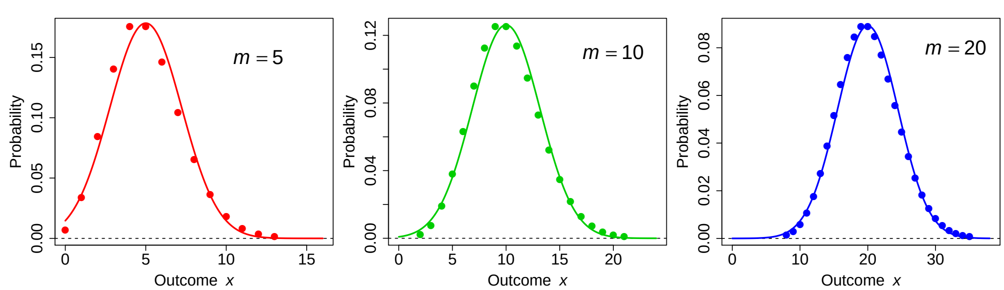


Figure 3.10: Examples of the discrete Poisson PMF, $Pois(m)$, and the continuous Normal PDF with the same moments, $N(\mu = m, \sigma^2 = m)$.

 © 2019 D. W. Scott

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

3.9.1 Standard Normal Distribution

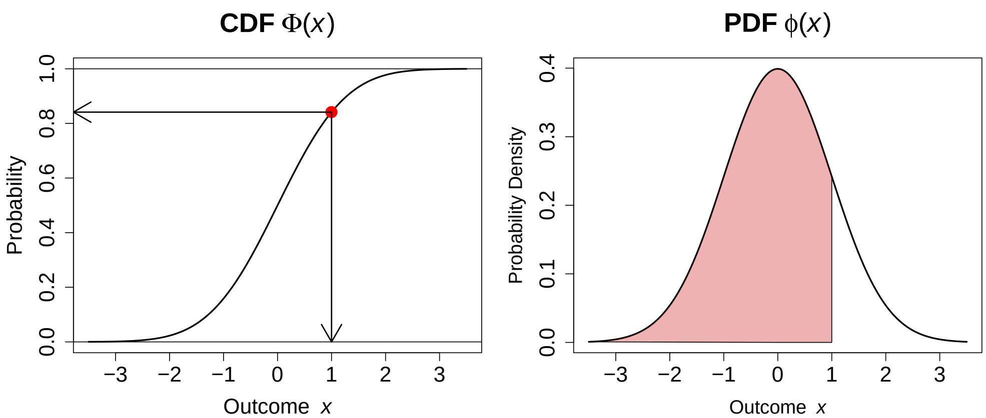


Figure 3.12: Standard Normal CDF, $\Phi(x)$, and PDF, $\phi(x)$, for $x = 1$.

3.9.2 Sums of Independent Normal Random Variables

3.9.3 Normal Approximation to the Poisson Distribution

3.10 Problems

 © 2019 D. W. Scott

 © 2019 D. W. Scott

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

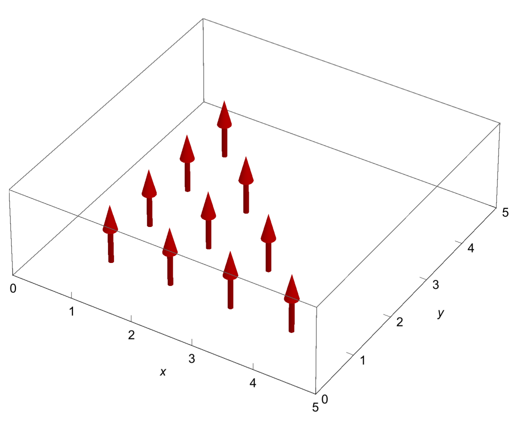


Figure 4.1: Joint bivariate PMF. Each arrow displays a probability of $\frac{1}{10}$.

4.1.3 Conditional Probability Density Function

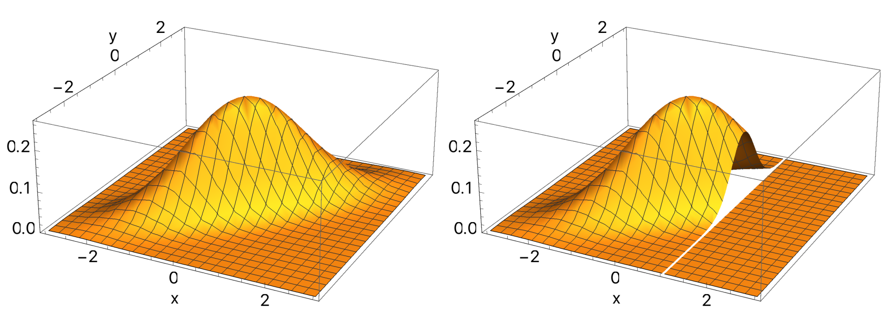


Figure 4.2: Conditional PDF, $f_{Y|X=1}(y|1)$, before normalization.

 © 2019 D. W. Scott

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

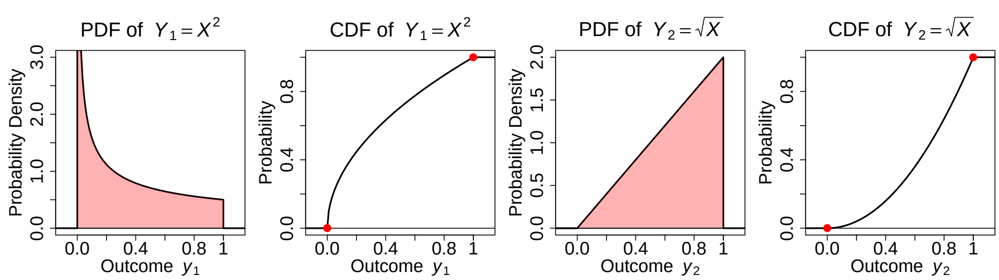


Figure 4.3: Transformations of a $Unif(0, 1)$ r.v.; see text.

© 2019 D. W. Scott

4.2.2 One-Dimensional Transformations

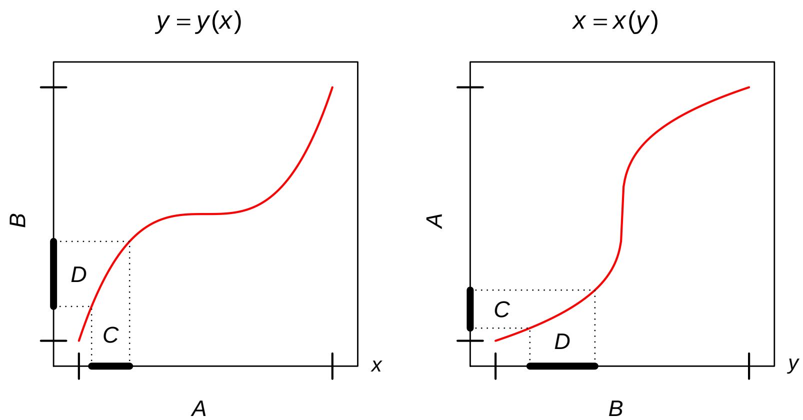


Figure 4.4: Sample transformations: $y = x^3$ and $x = sgn(y) \cdot |y|^{1/3}$. The range and domain of this transformation $A = B = (-1, 1)$.

4.2.2.1 Example 1: Negative Exponential PDF

4.2.2.2 Example 2: Cauchy PDF

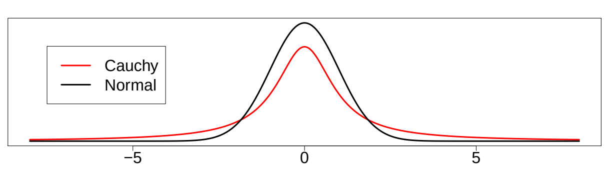


Figure 4.5: Standard Cauchy and Normal PDF's.

 © 2019 D. W. Scott

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

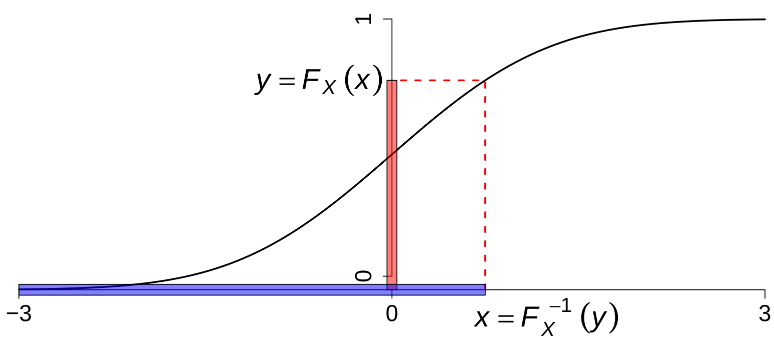


Figure 4.6: Generic PIT diagram. The red strip represents the event $Y \leq y$ while the blue strip represents the equivalent event $X \leq F_X^{-1}(y)$.

4.3.3 Event-Driven Simulation

4.4 Problems

 © 2019 D. W. Scott

 © 2019 D. W. Scott

Chapter 5

Approximations and Asymptotics

5.1 Why Do We Like Random Samples?

5.1.1 When $u(\boldsymbol{X})$ Takes a Product Form

5.1.2 When $u(\boldsymbol{X})$ 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[1]

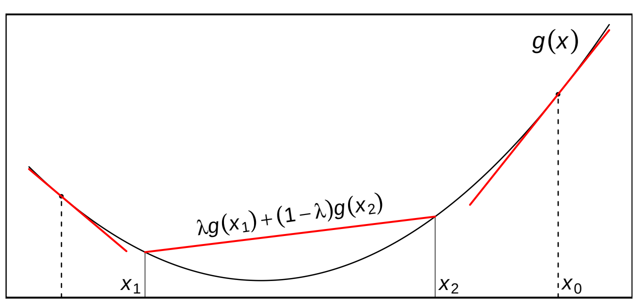


Figure 5.1: Example of a convex function $g(x)$ with two red tangent line segments touching the curve at the black points. A line segment connecting the curve at $x = x_1$ and $x = x_2$ is drawn in red; see text.

[1] This section may be omitted at a first reading.

 © 2019 D. W. Scott

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

[2]This section may be omitted at a first reading.

 © 2019 D. W. Scott

Chapter 6

Parameter Estimation

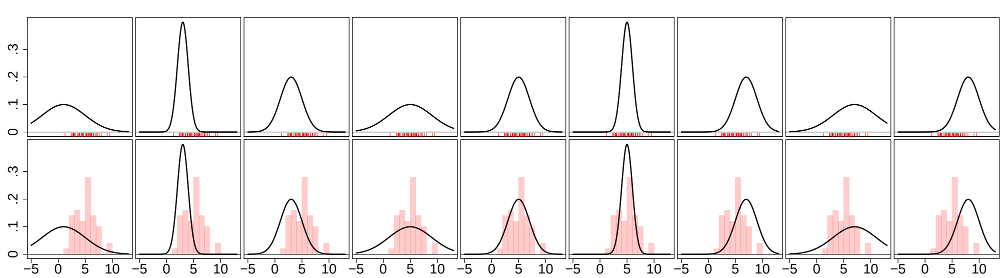


Figure 6.1: Nine examples of possible Normal fits to a random sample of 50 points. In the first row, the data are displayed using the **R** function `rug(x)`. In the second row, probability histograms `hist(x,prob=T)` are displayed.

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

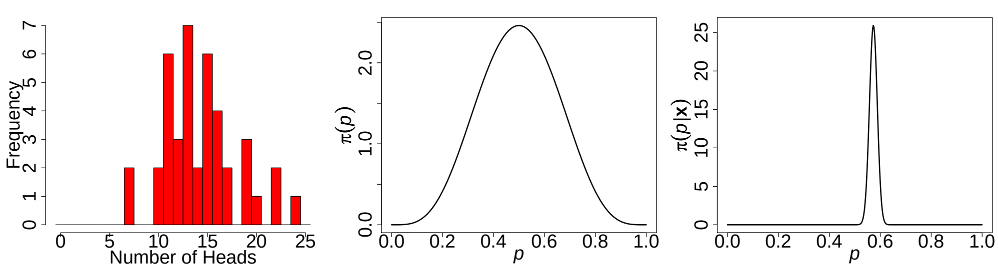


Figure 6.2: From left to right: a histogram of the 41 data points; the $Beta(5, 5)$ prior PDF; and the posterior $Beta(593, 442)$ PDF.

 © 2019 D. W. Scott

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

© 2019 D. W. Scott

 © 2019 D. W. Scott

Chapter 7

Hypothesis Testing

7.1 Setting up a Hypothesis Test

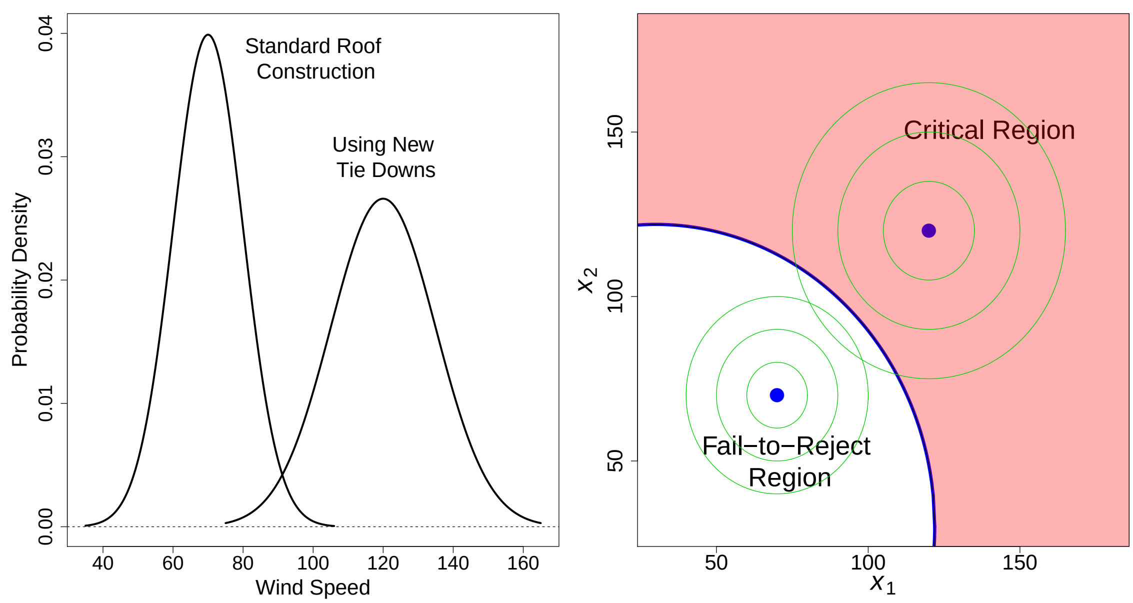


Figure 7.1: (L) For two roof construction techniques, hypothetical PDF's $\phi(x|70, 10^2)$ and $\phi(x|120, 15^2)$ of the minimum wind speed incurred that resulted in roof damage during a hurricane. (R) Illustration of a possible hypothesis-testing decision region for a small sample of $n = 2$ roofs.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

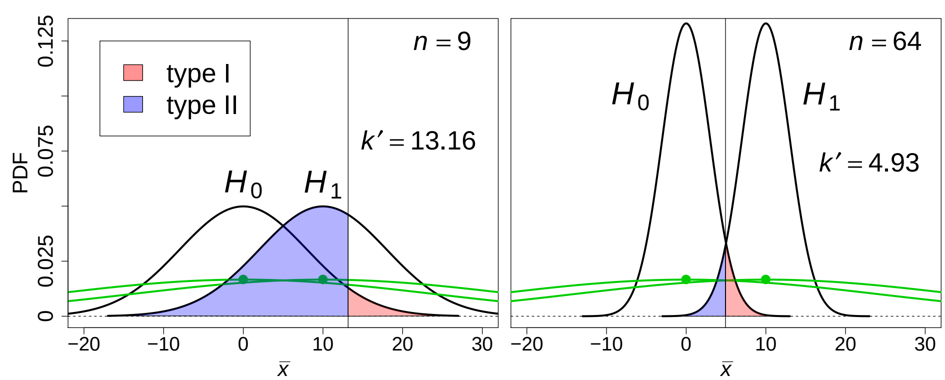


Figure 7.2: Critical regions based upon $\bar{X} > k'$ for testing two shifted Normal PDF's with $\mu_0 = 0$, $\mu_1 = 10$, and common $\sigma = 24$. (L) $n = 9$; (R) $n = 64$. The type I and II errors are shown in red and blue, respectively. The underlying sampling densities and means are shown in green;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

7.3 Best Critical Region for a Composite Alternative Hypothesis

7.3.1 Negative Exponential Composite Hypothesis Test

7.3.1.1 Example

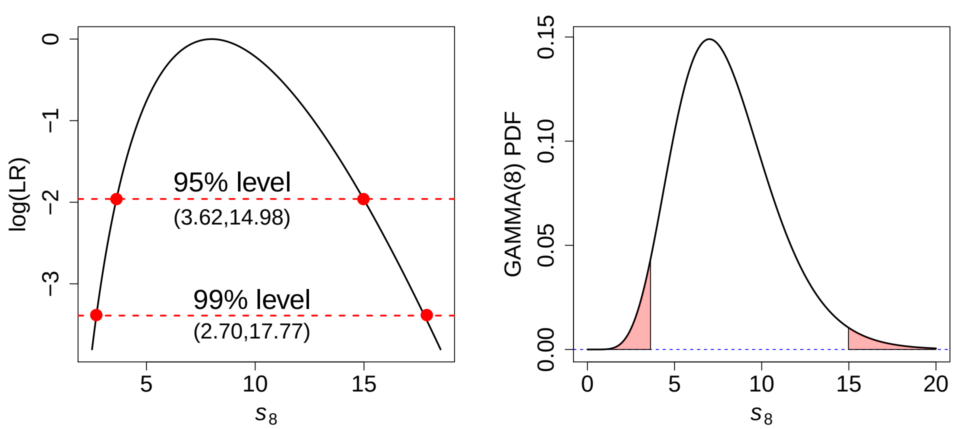


Figure 7.3: (L) The log-likelihood ratio for a sample of $n = 8$ negative exponential r.v.'s with $\beta_0 = 1$. The levels corresponding to 5% and 1% type I errors are shown. (R) The $Gamma(8, \beta = 1)$ PDF of S_8, together with the 95% probability interval $(3.62, 14.98)$. The shaded tail areas have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

 © 2019 D. W. Scott

7.3.1.2 Alternative Critical Regions

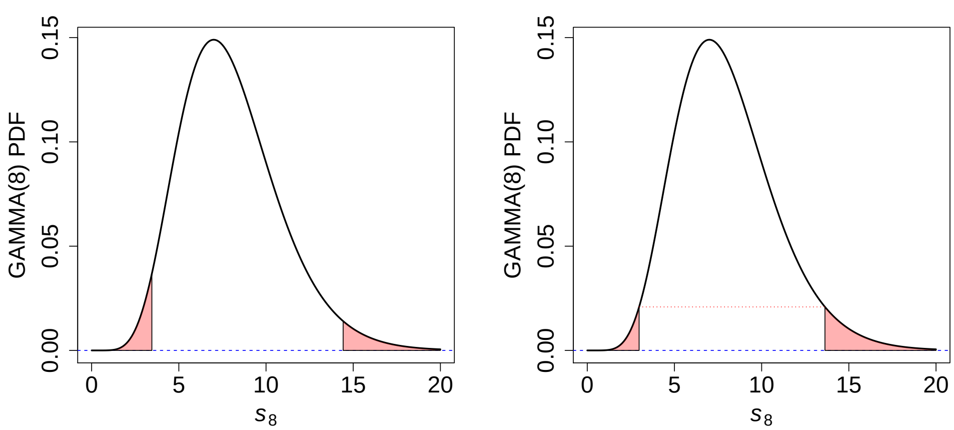


Figure 7.4: Alternative 95% confidence level tests for our example. (L) $(3.45, 14.42)$ has equal tail probabilities of 2.5%; (R) $(2.97, 13.63)$ is the narrowest interval. The tail areas are 1.824% and 3.176%, respectively.

 © 2019 D. W. Scott

7.3.1.3 Mount St. Helens Example

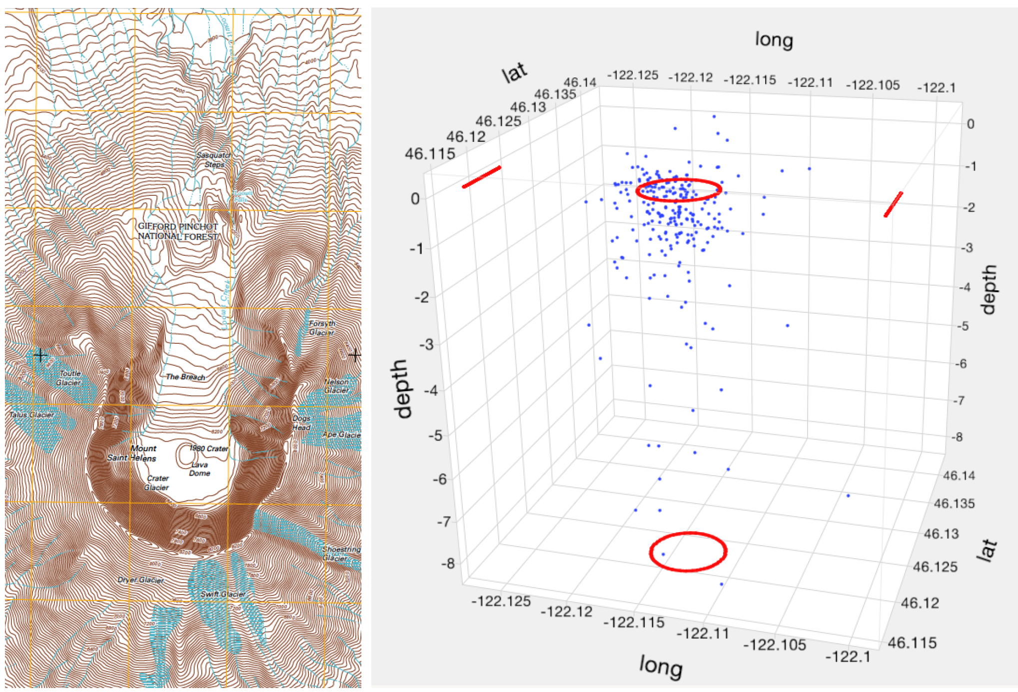


Figure 7.5: (L) Topo map; (R) Earthquake epicenters.

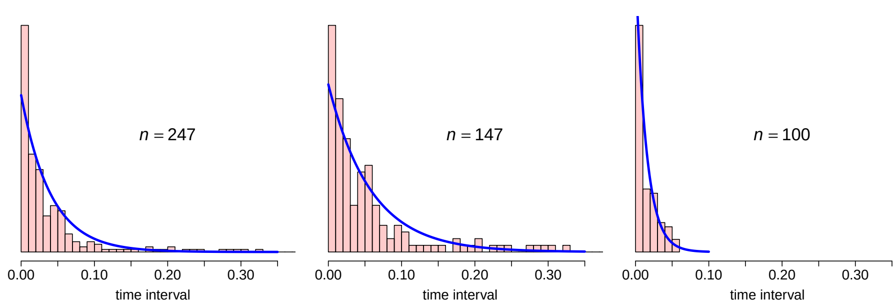


Figure 7.6: Histograms of times between eruptions (in days) for all 247 eruptions (left frame), the first 147 eruptions (middle frame), and last 100 eruptions (right frame). The blue line depicts a negative exponential fit.

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

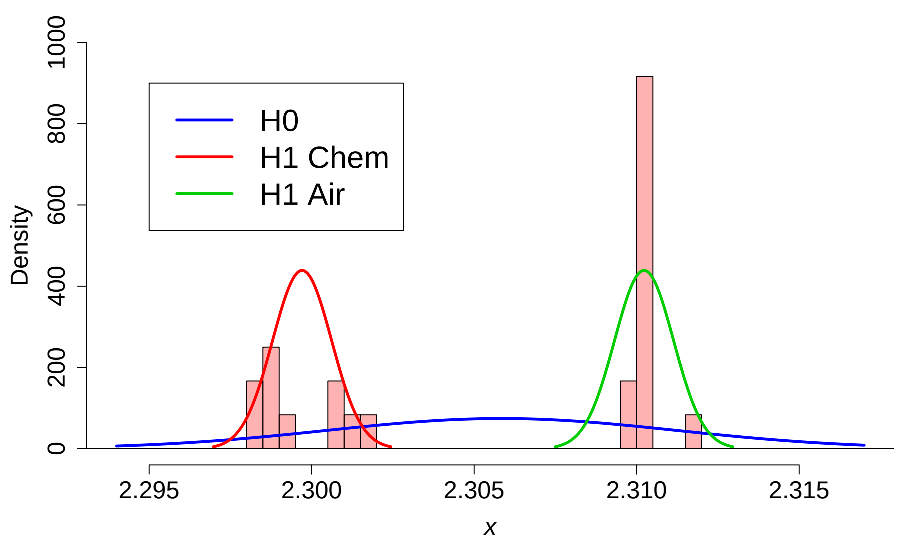


Figure 7.7: Fits to Lord Rayleigh's Data Under the Null and Alternative Hypotheses. A Nobel Prize was awarded for understanding this diagram.

 © 2019 D. W. Scott

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

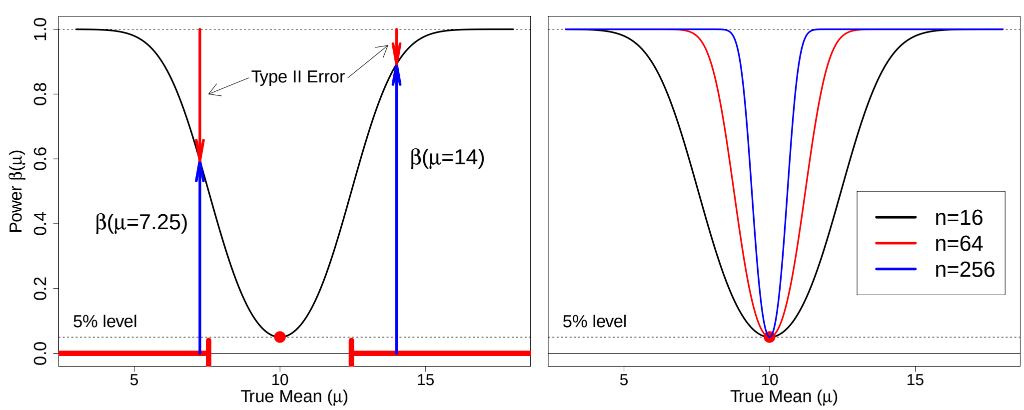


Figure 7.8: (L) Power function with $n = 16$. The critical region is shown in red, along with two values of the power function at $\mu = 7.25$ and 14.0. Their complements are examples of type II errors. (R) Effect of sample size on the power function.

7.5 Multiple Testing and the Bonferroni Correction

7.6 Problems

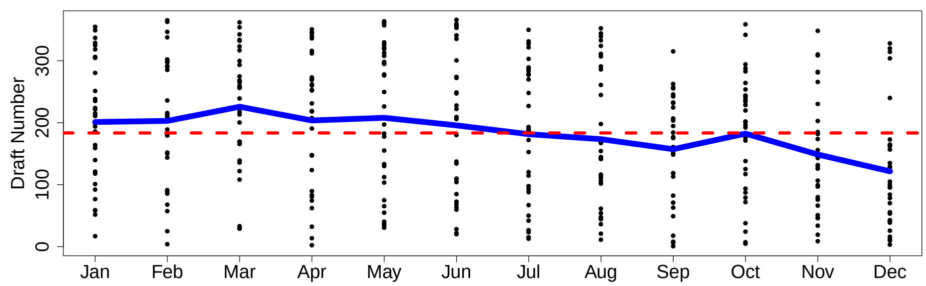


Figure 7.9: Draft lottery numbers 1–366 by month. The monthly average is the blue line. The overall average of 183.5 is the red dotted line.

 © 2019 D. W. Scott

Chapter 8

Confidence Intervals and Other Hypothesis Tests

8.1 Confidence Intervals

8.1.1 Confidence Interval for μ: Normal Data, σ^2 Known

8.1.2 Confidence Interval for μ: σ^2 Unknown

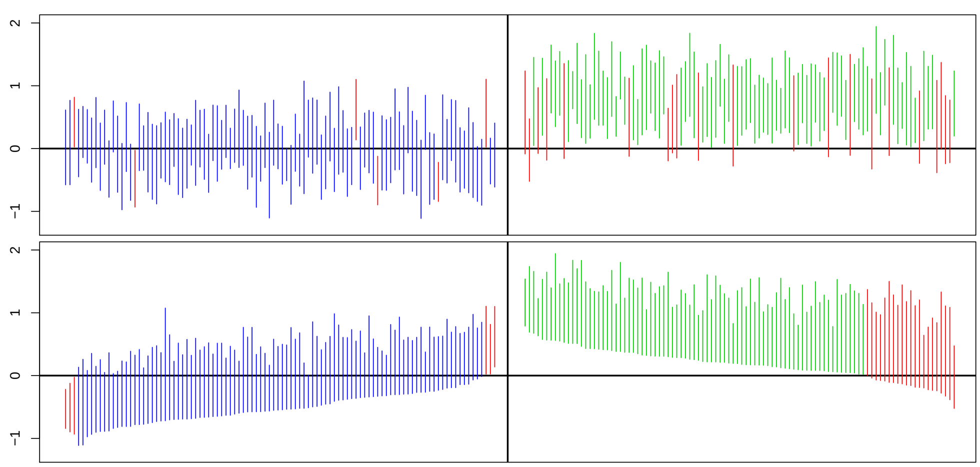


Figure 8.1: *(Left frames)* A simulation study showing 100 95% confidence intervals for samples of size $n = 16$ from the $N(0,1)$ PDF, where the intervals that fail to include the true value of $\mu_0 = 0$ are shown in red. The bottom left frame shows the same 100 confidence intervals sorted for clarity. *(Right frames)* A simulation study from the alternative hypothesis PDF $N(0.7493, 1)$, chosen so that the power is 80%. In fact, 21 of the 100 CIs 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

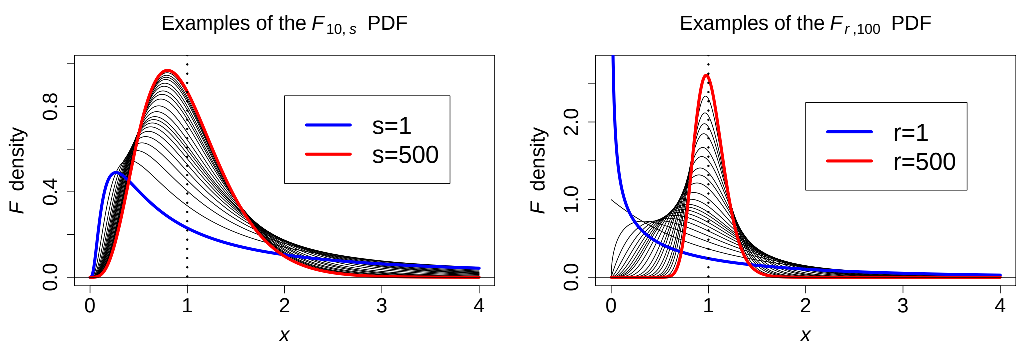


Figure 8.2: (L) Examples of the $F_{10,s}$ PDF for $1 \leq s \leq 500$. (R) Examples of the $F_{r,100}$ PDF for $1 \leq r \leq 500$.

 © 2019 D. W. Scott

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 $\rho = 0$

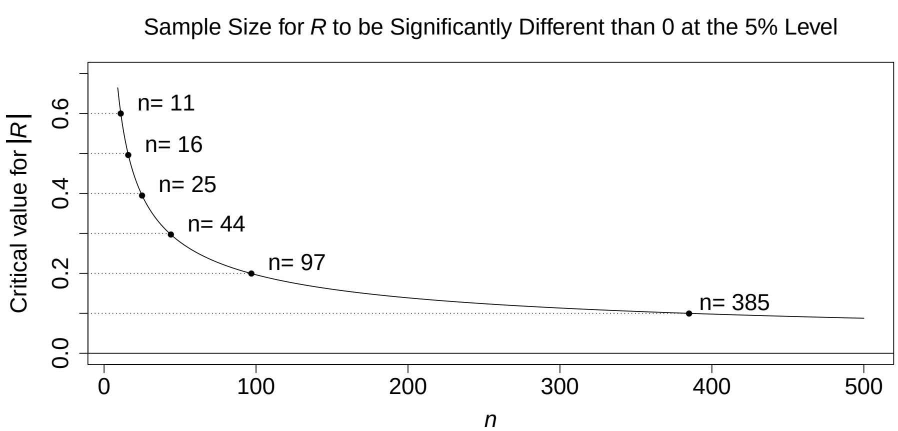


Figure 8.3: Critical values for R as the sample size increases.

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

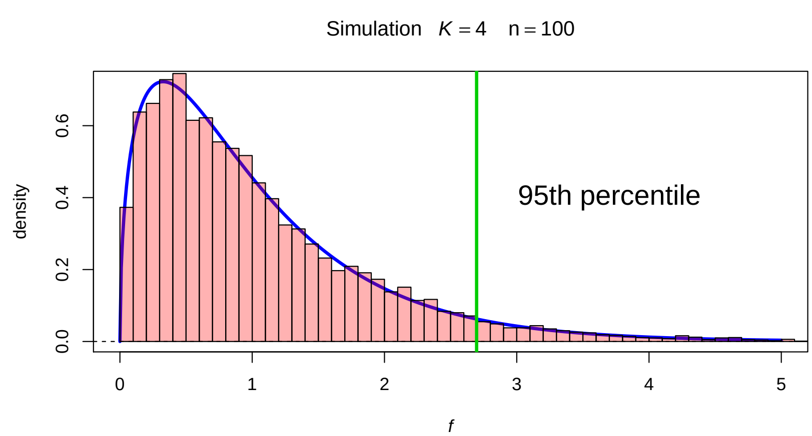


Figure 8.4: Ten thousand simulations of the F-statistic (**??**) with $K = 4$, $n_k = 25$, and $n = 100$ under the null hypothesis. $H_0 : \mu_k = \mu_0$, $k = 1, 2, 3, 4$.

8.7 Problems

 © 2019 D. W. Scott

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

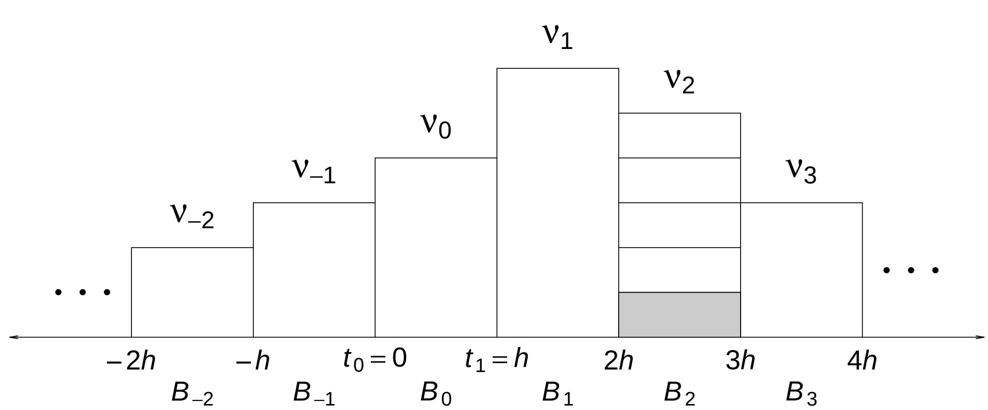


Figure 9.1: For a frequency histogram, the notation used to denote the locations of the bins $\{B_k\}$, the bin counts $\{\nu_k\}$, and the bin edges $\{t_k\}$.

9.1.3 Examples with Normal Data

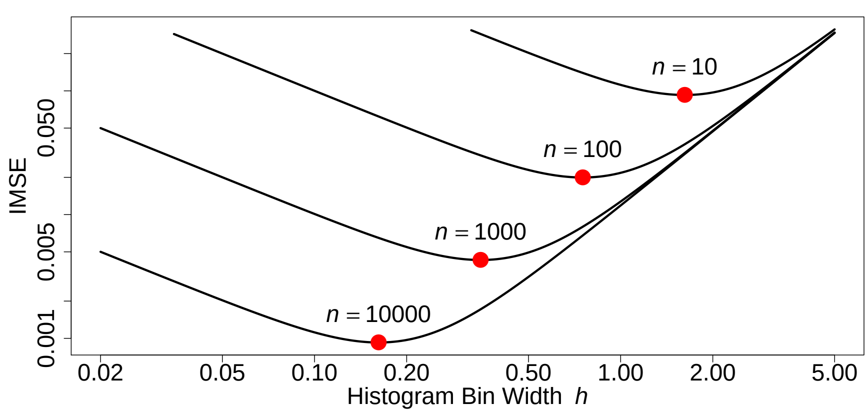


Figure 9.2: For several sample sizes and the $N(0,1)$ density, the histogram IMSE curves as the bin width varies on a log-log scale. The red dots locate the best $h = h_n^*$.

 © 2019 D. W. Scott

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

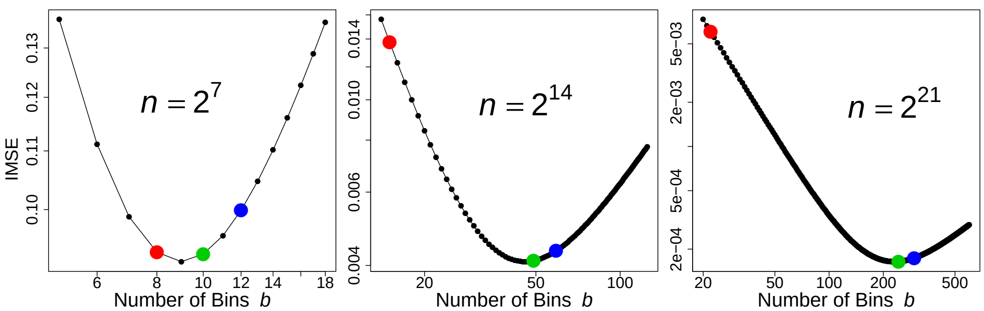


Figure 9.3: Comparison of the number of bins recommended by Sturges' (●), Scott's (●), and FD's (●) Rules for a $Beta(5, 5)$ PDF plotted against the exact IMSE values.

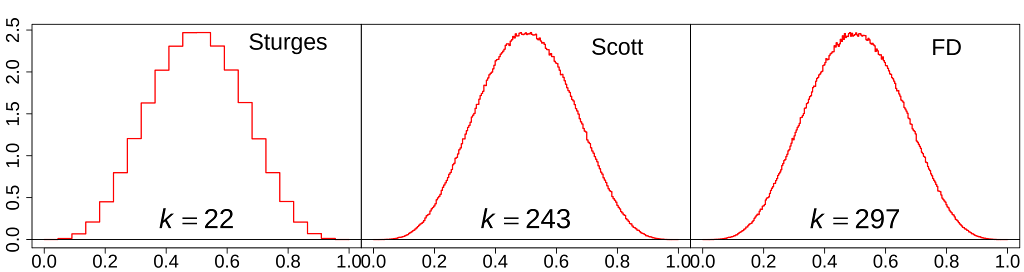


Figure 9.4: Three Histograms of a Beta(5, 5) sample with $n = 2^{21}$.

 © 2019 D. W. Scott

9.2 An Optimal Stopping Time Problem

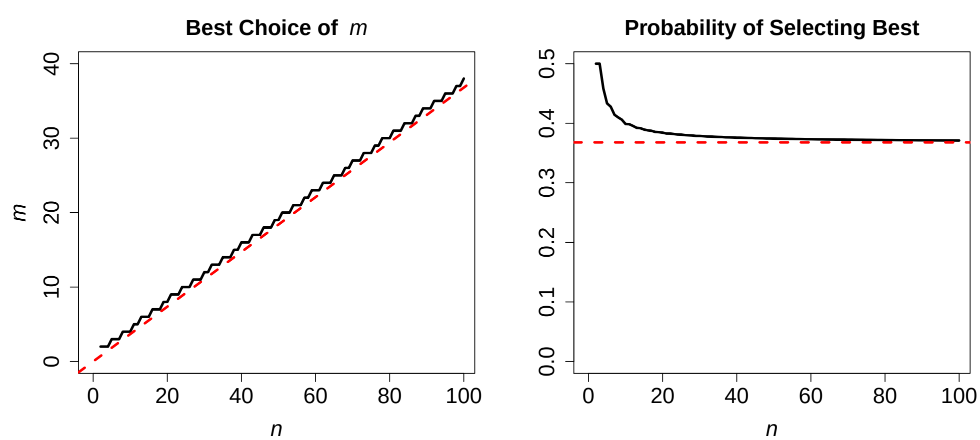


Figure 9.5: (L) Optimal stopping point m as a function of the population size n compared to a straight line with slope $1/e$; (R) Probability of selecting the best candidate using m^* compared to $p = 1/e$.

 © 2019 D. W. Scott

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

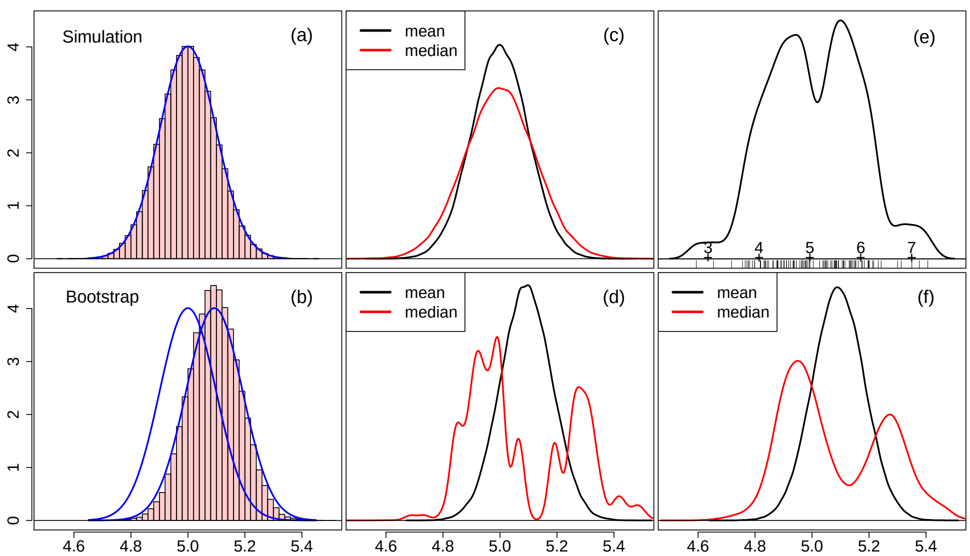


Figure 9.6: Simulation and bootstrap analysis of the sample mean and sample median for a $N(5, 1)$ PDF with $n = 101$ points; see text. Histograms and smoothed histograms are displayed.

 © 2019 D. W. Scott

9.5 Multiple Linear Regression

9.6 Experimental Design

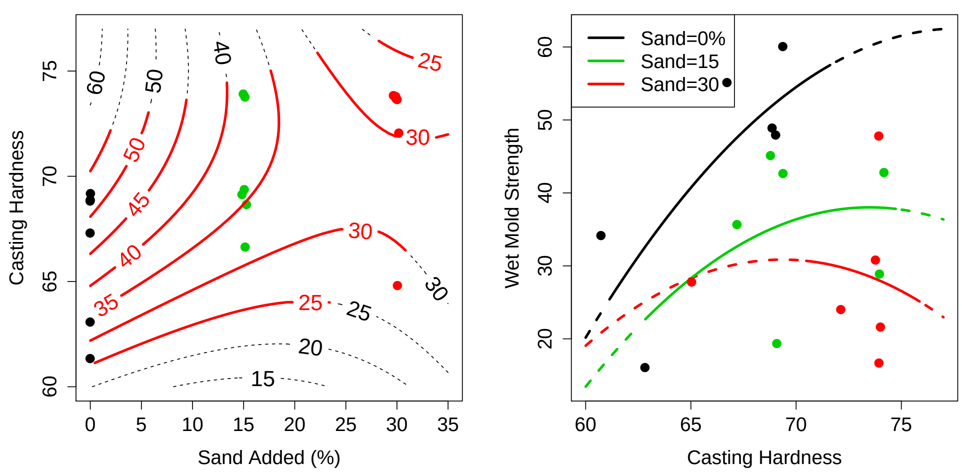


Figure 9.7: Surface of predicted average mold strength as a function of the two predictor variables. The contours and curves are shown as dotted lines where the standard deviation of the prediction exceeds 2.5.

 © 2019 D. W. Scott

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

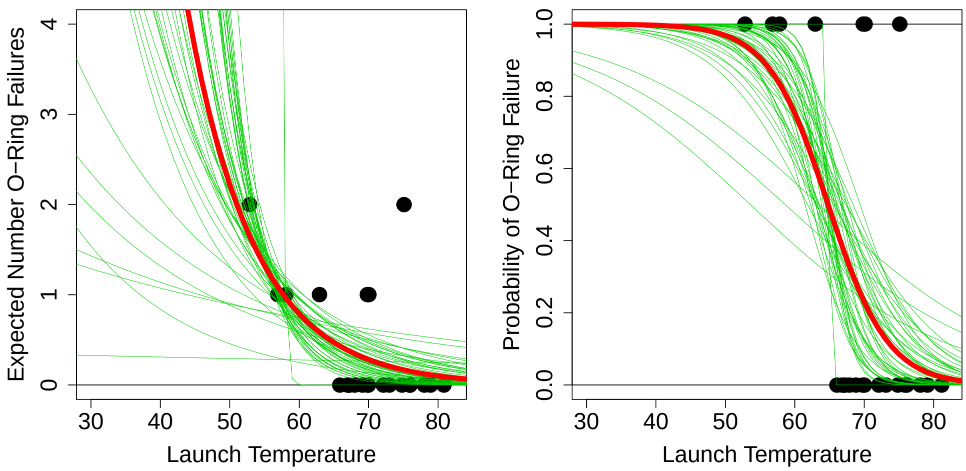


Figure 9.8: (L) Poisson and (R) Logistic regression models fitted by the **R** function `glm` to the space shuttle data (red line). Fifty bootstrap fits are superimposed (green lines).

9.8 Robustness

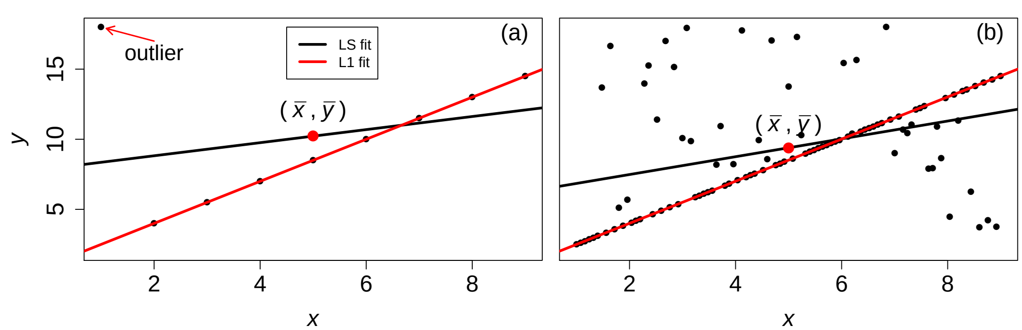


Figure 9.9: (a) $n = 9$ points satisfying $y = 1 + 1.5x$ with one outlier at $x = 1$; (b) $n = 101$ with 40 randomly selected outliers (with negative slope).

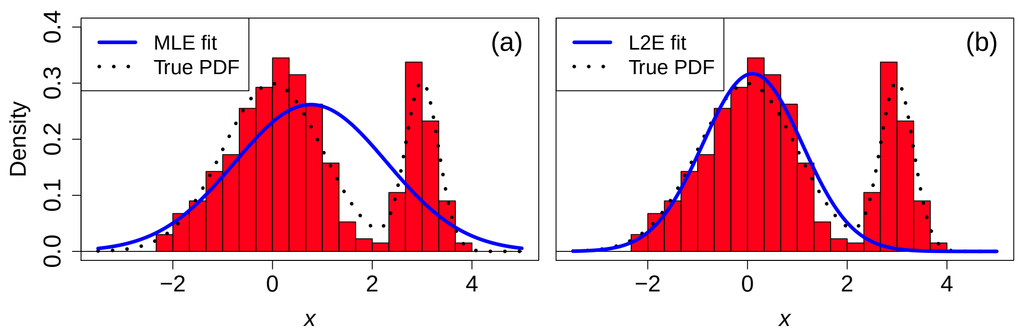


Figure 9.10: (a) MLE $N(\bar{x}, s^2)$ and (b) L_2E Normal fits to a random sample of 400 points from the mixture $0.75 \times N(0,1) + 0.25 \times N(3, 1/3^2)$; see text.

9.9 Conclusions

© 2019 D. W. Scott