SOLUTIONS MANUAL

Statistical Data Analytics

Foundations for Data Mining, Informatics, and Knowledge Discovery

Walter W. Piegorsch

WILEY

Statistical Data Analytics

Statistical Data Analytics

Foundations for Data Mining, Informatics, and Knowledge Discovery

Solutions Manual

Walter W. Piegorsch

University of Arizona, USA

WILEY

This edition first published 2016

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data applied for.

ISBN: 9781119030652

A catalogue record for this book is available from the British Library.

Typeset in 10/12pt TimesLTStd by SPi Global, Chennai, India

1 2016

Contents

Preface

This manual is designed to supplement the exercises and examples in *Statistical Data Analytics: Foundations for Data Mining, Informatics, and Knowledge Discovery* (John Wiley & Sons, Ltd) by providing solutions for end-of-chapter exercises.

In many cases, computer code using the **R** statistical environment (R Core Team 2014) is employed to complete part or all of the solution, mirroring use of this statistical package in the original text. The program operates on Windows®, Apple OS, and Linux systems. At the time of this writing, the version used is 3.1.0, employing 64-bit format. The bulk of the **R** material in this manual was prepared from that version. The **R** code is not intended to be the most efficient way to program the desired operations, but it will help illustrate a plausible approach for acquisition of the solution. For more on **R**, see Appendix B of the main text, the useful background sources by Dalgaard (2008) or Verzani (2005), the concise online guide by Owen (http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf), and of course the main Comprehensive **R** Archive Network (CRAN) website and its online manual at http://cran.r-project.org/doc/manuals/R-intro.html.

Instructors of classes adopting *Statistical Data Analytics* as a required text may employ materials herein for classroom use. Otherwise, no part of this material may be reproduced, stored in a retrieval system, or transcribed in any form or by any means – electronic, online, mechanical, photoreproduction, recording, or scanning – without the prior written consent of the author and John Wiley & Sons, Ltd.

Walter W. Piegorsch
Tucson, Arizona

1

Data analytics and data mining

Exercise Solutions

1.1 Besides the references to bioinformatics, medical informatics, ecoinformatics (which focuses on ecological informatics), geoinformatics, and socioinformatics made in the textbook, one can find mention of nursing informatics and healthcare informatics (offshoots of medical informatics), chemoinformatics (or cheminformatics), econoinformatics (also called business informatics), technoinformatics (which seems a bit ambiguous, actually), molecular informatics, etc.

1.2 For example, in environmental research, a study of pollution in Canadian streams, ponds, and lakes exposed to industrial effluents mined large numbers of tadpoles – of which there was no shortage – to examine their DNA for damage due to the toxic exposure (Ralph and Petras, 1997). Increases in DNA damage were seen as markers for potential ecological damage.

1.3 The database mentioned in Exercise 1.2 involved populations of tadpoles in streams, ponds, and lakes throughout southern Ontario. Measured were the average length-to-width ratios in DNA fragments from 25 of each tadpole's peripheral blood erythrocytes. The target population was all tadpoles living in these bodies of water who could possibly be exposed to the effluent. The sampling frame was all tadpoles living in the particular bodies of water actually sampled who could possibly be exposed to the effluent.

1.4 (a) Individual-level distortion occurs in all the sorts of situations mentioned: data collection errors, data entry errors with misplaced or missing decimal points, transposed digits, incorrect rounding errors, etc. Missing data is also a possibility, along with impossible classification combinations such as 'age=4/children=2,' etc.

Statistical Data Analytics: Foundations for Data Mining, Informatics, and Knowledge Discovery, Solution Manual, First Edition. Walter W. Piegorsch.

www.wiley.com/go/piegorsch/data_analytics

(b) Collective-level distortion may occur when the sampling frame and target population are confused, e.g. collect data for responses in laboratory mammals exposed to a toxin where inference is directed at another species immune to the toxin, or a study of magazine buying habits among newborn children.

1.5 Neither: *all* customers where queried so it's a complete census and there's no distortion (just poorly designed data acquisition).

1.6 As noted by Hand *et al*. (2000, p. 119) this is individual-level distortion.

1.7 This is convenience sampling: only students who happen to enter the Student Union or main cafeteria while the questioners are present are sampled. The sampling could be changed to sample only every *k*th student (after a random start; a form of 'systematic sampling') or to choose via a (wholly) random mechanism whether or not a student is sampled as s/he enters the building.

1.8 Yes, there is selection bias evident: whether or not a record is included in the database depends on the values of the variables.

1.9 The physician only summarized whether temporal patterns occurred in the patients' asthma onset when low-pressure weather fronts passed through the region. Thus this was an example of statistical description. No attempts were made to inferentially associate any connections between weather patterns and asthma onset.

1.10 As in Exercise 1.9, the physician only summarized proximity to construction sites and asthma onset. This remains an example of statistical description. No attempts were made to inferentially associate any patterns between construction sites and asthma onset.

1.11 Since the geographer determined via statistical inference if a difference existed in property loss due to the floods, this was a an example of statistical inference.

2

Basic probability and statistical distributions

Exercise Solutions

2.1 (a) The sample space is all possible non-negative integers: $\mathcal{S} = \{0, 1, 2, \ldots\}$. One could also say the space is all non-negative integers below some unknown upper bound, M, but this then will define the sample space in terms of an unknown parameter M.

(b) The sample space is all possible positive real numbers: $\mathcal{S} = [0, \infty)$. One could also say the space is all positive real numbers below some unknown upper bound, τ, but this then will define the sample space in terms of an unknown parameter τ.

(c) The sample space is all possible quantities of the form $I + J/100$, where I and J are non-negative integers: $\mathcal{S} = \{0.00, 0.01, 0.02, \ldots\}$.

(d) The sample space is all pairs of real numbers (x, y) with $-180 \leq x \leq 180$ and $-180 \leq y \leq 180$ and where negative longitudes indicate westerly direction from the Prime Meridian and negative latitudes indicate southernly direction from the Equator. Either x or y may also be represented in the traditional notation $\pm D^\circ M' S''$, i.e. in Degrees, Minutes, and Seconds of arc, where $0 \leq D \leq 90$, $0 \leq\ M \leq 60$, and $0 \leq\ S \leq 60$.

2.2 (a) Discrete.

(b) Continuous.

(c) Discrete.

(d) Continuous.

Statistical Data Analytics: Foundations for Data Mining, Informatics, and Knowledge Discovery, Solution Manual, First Edition. Walter W. Piegorsch.

www.wiley.com/go/piegorsch/data_analytics

(e) Discrete.

(f) Continuous.

(g) (Bivariate) continuous.

2.3 (a) No: violates $\sum_{m \in \mathcal{S}} f(m) = 1$. ($\sum_{m \in \mathcal{S}} f(m) = 14/12$).

(b) Yes: $0 \leq f(m) \leq 1$ and $\sum_{m \in \mathcal{S}} f(m) = 1$.

(c) No: violates $0 \leq f(m) \leq 1$ ($f(6) = -0.3 < 0$).

(d) Yes: clearly $0 \leq f(m) \leq 1$ for any $m \in \{1, 2, \ldots, N\}$. Also, recall that $\sum_{m=1}^{N} m^2 = \frac{1}{6}N(N+1)(2N+1)$, so that $\sum_{m=1}^{N} f(m) = 1$.

(e) Yes: this is known as the *logarithmic distribution* or sometimes the *log-series distribution* with parameter π. Since $\pi \in (0, 1)$, we see $\log(1 - \pi) < 0$ and so $0 \leq f(m) \leq 1$ for any $m \in \{1, 2, \ldots, \infty\}$. Also, recall the Maclaurin series expansion (a Taylor series expansion about zero) for $\sum_{m=1}^{\infty} \pi^m/m$ is $-\log(1 - \pi)$ for any $\pi \in (0, 1)$. This shows immediately that $\sum_{m=1}^{N} f(m) = 1$.

2.4 (a) No. Integral of p.d.f., $\int_1^{\infty} f(x)dx$, diverges.

(b) No. Integral of p.d.f. is 1, but $f(x) < 0$ over $x < 1$.

(c) Yes. This is $X \sim \text{Beta}(\alpha, 2)$.

(d) From the hint: the plot shows a positive-valued, continuous, linearly increasing-to-δ then linearly decreasing-to-ω function over $x \in (0, \omega)$. Simple integration shows $\int_0^{\omega} f(x)dx = 1$. So, this is a valid p.d.f., known as the *Triangular distribution*.

2.5 The Multiplication Rule (2c) gives $P[\mathcal{B} \text{ and } \mathcal{E}] = P[\mathcal{B}|\mathcal{E}]\,P[\mathcal{E}]$. But as noted, $P[\mathcal{B} \text{ and } \mathcal{E}] = P[\mathcal{E} \text{ and } \mathcal{B}]$. So, apply Rule (2c) to the latter term: $P[\mathcal{E} \text{ and } \mathcal{B}] = P[\mathcal{E}|\mathcal{B}]\,P[\mathcal{B}]$. Thus

$$P[\mathcal{B}|\mathcal{E}]\,P[\mathcal{E}] = P[\mathcal{E}|\mathcal{B}]\,P[\mathcal{B}]$$

Dividing by $P[\mathcal{B}]$ (and assuming $P[\mathcal{B}] \neq 0$) produces

$$P[\mathcal{E}|\mathcal{B}] = P[\mathcal{B}|\mathcal{E}]\frac{P[\mathcal{E}]}{P[\mathcal{B}]},$$

which is textbook Equation (2.1).

2.6 When X and Y are independent, we know $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Thus the conditional p.d.f. of $X|Y$ becomes

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x),$$

which is just the marginal p.d.f. of X.

2.7 Given: constants a and b and a continuous random variable $X \sim f_X(x)$. Assume all integrals are taken over the pertinent support of X.

(a) $E[a] = \int af_X(x)dx = a\int f_X(x)dx = (a)(1) = a$.

(b) $E[bX] = \int bxf_X(x)dx = b\int xf_X(x)dx = bE[X]$.

(c) $E[a + bX] = \int (a + bx)f_X(x)dx = \int af_X(x)dx + \int bxf_X(x)dx = a + bE[X]$, where the latter equality follows from the previous results in parts (2.7a) and (2.7b).

2.8 Given: $X \sim f_X(x)$ with finite population mean μ_X, and finite population variance σ^2_X.

(a) $\sigma^2_X = E[(X - \mu_X)^2] = E[X^2] - 2\mu_X E[X] + E[\mu^2_X] = E[X^2] - 2\mu_X\mu_X + \mu^2_X = E[X^2] - 2\mu^2_X + \mu^2_X = E[X^2] - \mu^2_X$, as desired.

(b) From part (2.8a), $\sigma^2_X = E[X^2] - \mu^2_X$ so add μ^2_X to both sides of the equation to find $E[X^2] = \mu^2_X + \sigma^2_X$.

2.9 Given: constants a and b and a continuous random variable $X \sim f_X(x)$ with finite population mean μ_X and finite population variance σ^2_X.

(a) From Exercise 2.7a we know $E[a] = a$, so $\mathrm{Var}[a] = E[(a - E[a])^2] = E(a - a)^2] = E[0] = 0$. Thus, in effect, a constant possesses no variation.

(b) Applying Exercise 2.7b we see $E[aX] = aE[X] = a\mu_X$, so $\mathrm{Var}[aX] = E[(aX - E[aX])^2] = E[(aX - a\mu_X)^2] = E[a^2(X - \mu_X)^2] = a^2E[(X - \mu_X)^2]$, which is clearly just $a^2\sigma^2_X$.

(c) From part (2.8a) we see, in effect, that adding a constant to a random variable adds no variability, so $\mathrm{Var}[a + bX] = \mathrm{Var}[bX]$. But then from part (2.9b) this is $\mathrm{Var}[a + bX] = b^2\mathrm{Var}[X]$

2.10 We have $f(m) = \frac{1}{6}$ for all $m = 1, \ldots, 6$. It is evident that $E[X] = 3.5$, so using Exercise 2.8a find the variance as $\mathrm{Var}[X] = E[X^2] - E^2[X] = \sum_{m=1}^{6} m^2 \left(\frac{1}{6}\right) - (3.5)^2 = \frac{1}{6}\sum_{m=1}^{6} m^2 - 12.25$. But recall as in Exercise 2.3d that $\sum_{m=1}^{6} m^2 = \left(\frac{1}{6}\right)(6)(7)(13) = 91$. This gives $\mathrm{Var}[X] = (91/6) - 12.25 = 35/12$, as desired.

2.11 Given: X and Y are independent. Assume the continuous case (the discrete case is similar) and that all pertinent integrals exist. We know $\mathrm{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = \int\int (x - \mu_X)(y - \mu_Y)f_{X,Y}(x, y)\,dx\,dy$. But under independence this is

$$\begin{aligned}
\mathrm{Cov}[X, Y] &= \int\int (x - \mu_X)(y - \mu_Y)f_X(x)f_Y(y)\,dx\,dy \\
&= \int (y - \mu_Y)f_Y(y) \int (x - \mu_X)f_X(x)\,dx\,dy \\
&= \int (y - \mu_Y)f_Y(y)\,E[X - \mu_X]\,dy \\
&= E[X - \mu_X] \int (y - \mu_Y)f_Y(y)\,dy\,.
\end{aligned}$$

But clearly $E[X - \mu_X] = E[X] - \mu_X = 0$ (the remaining integral with $f_Y(y)$ will be the same), so the covariance calculates to zero.

2.12 (a) For $c(b) = E[\{b(X - \mu_X) + (Y - \mu_Y)\}^2]$ find

$$\begin{aligned}
c(b) &= E[b^2(X - \mu_X)^2 + 2b(X - \mu_X)(Y - \mu_Y) + (Y - \mu_Y)^2] \\
&= b^2E[(X - \mu_X)^2] + 2bE[(X - \mu_X)(Y - \mu_Y)] + E[(Y - \mu_Y)^2] \\
&= b^2\sigma^2_X + 2b\sigma_{XY} + \sigma^2_Y
\end{aligned}$$

(b) Notice that $c(b) = E[\{b(X - \mu_X) + (Y - \mu_Y)\}^2]$ is the expected value of a squared quantity, so it must be nonnegative: $c(b) \geq 0$.

(c) Recognizing $c(b)$ as a quadratic polynomial in b, we see its leading coefficient, σ_X^2, is positive and so the parabola must be convex. But from part (2.12b) we know the function is nonnegative. Thus it will either have one real root if it can be zero at some b, or will have no real roots if it is strictly positive. In either case, it has no more than one real root.

(d) The discriminant of a parabola with at most one real root cannot be positive. Here, this translates to $4\sigma_{XY}^2 - 4\sigma_X^2\sigma_Y^2 = 4(\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2) \leq 0$.

(e) Divide by 4 across the inequality in part (2.12d) – note that the direction does not change – to find $\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2 \leq 0$. That is, $\sigma_{XY}^2 \leq \sigma_X^2\sigma_Y^2$. Now take square roots across the inequality: $-\sigma_X\sigma_Y \leq \sigma_{XY} \leq \sigma_X\sigma_Y$. Dividing by the product of the standard deviations (which are each defined to be positive, so again directions do not change) gives the desired result: $-1 \leq \rho_{XY} \leq 1$ (Casella and Berger, 2002, Section 4.5).

2.13 From the Hint:

$$\begin{aligned}
\text{Var}[X_i + X_j] &= E[\{(X_i + X_j) - (\mu_i + \mu_j)\}^2] \\
&= E[\{(X_i - \mu_i) + (X_j - \mu_j)\}^2] \\
&= E[\{(X_i - \mu_i)^2 + 2(X_i - \mu_i)(X_j - \mu_j) + (X_j - \mu_j)\}^2] \\
&= E[\{(X_i - \mu_i)^2] + 2E[(X_i - \mu_i)(X_j - \mu_j)] + E[(X_j - \mu_j)\}^2] \\
&= \text{Var}[X_i] + 2\text{Cov}[X_i, X_j] + \text{Var}[X_j].
\end{aligned}$$

2.14 $X \sim \text{Bin}(10, 0.2)$. Note the discrete nature of the binomial sample space here.

(a) $P[X = 6]$ via **R**: `dbinom(x=6, size=10, prob=.2)` gives 0.0055.

(b) $P[X \leq 2]$ via **R**: `pbinom(q=2, size=10, prob=.2)` gives 0.6778.

(c) $P[X \geq 1] = 1 - P[X < 1] = 1 - P[X = 0]$ from the Complement Rule (2d), so via **R**: `1-dbinom(x=0, size=10, prob=.2)` gives 0.8926.

(d) $P[2 \leq X \leq 6] = P[X \leq 6] - P[X < 2] = P[X \leq 6] - P[X \leq 1]$, so via **R**:

```
pbinom(q=6,size=10,prob=.2) - pbinom(q=1,size=10,prob=.2)
```

gives 0.6233.

(e) $P[2 < X < 6] = P[3 \leq X \leq 5] = P[X \leq 5] - P[X \leq 2]$, so via **R**:

```
pbinom(q=5,size=10,prob=.2) - pbinom(q=2,size=10,prob=.2)
```

gives 0.3158.

[One could also use `sum(dbinom(x=3:5, size=10, prob=.2))`.]

2.15 $X \sim \text{Poisson}(\lambda)$. Note the discrete nature of the Poisson sample space here.

(a) $\lambda = 4.95$: $P[X = 3]$ via **R** is `dpois(x=3, lambda=4.95)` to find 0.1432.

(b) $\lambda = 4.95$: $P[X > 0] = 1 - P[X \leq 0] = 1 - P[X = 0]$ from the Complement Rule (2d), so via **R** use `1-dpois(x=0, lambda=4.95)` to find 0.9929.

(c) $\lambda = 13.65$: $P[4 < X \leq 11] = P[5 \leq X \leq 11] = P[X \leq 11] - P[X \leq 4]$. In **R** use
`ppois(q=11,lambda=13.65) - ppois(q=4,lambda=13.65)`
to find 0.2883.

(d) $\lambda = 13.65$: $P[X \geq 8.05] = 1 - P[X < 8.05] = 1 - P[X \leq 8]$. In **R** use
`1 - ppois(q=8,lambda=13.65)` to find 0.9265.

(e) $\lambda = 0.55$: $P[X \leq 4]$ via **R** is `ppois(q=4, lambda=.55)` to find 0.9997.

2.16 $X \sim \text{Geom}(\pi)$. Assume that t and u are positive integers such that $t > u$. Similar to that seen in Section 2.3.7, $P[X \geq t|X \geq u] = P[X \geq t]/P[X \geq u] = \{1 - F_X(t-1)\}/\{1 - F_X(u-1)\}$. Now, the c.d.f. of X was given in Section 2.3.3 as $F_X(m) = 1 - (1-\pi)^{m+1}$. Thus $P[X \geq t|X \geq u] = \{1 - 1 + (1-\pi)^{t-1+1}\}/\{1 - 1 + (1-\pi)^{u-1+1}\} = (1-\pi)^{t-u}$. But this is clearly $P[X \geq t-u] = 1 - F_X(t-u-1) = (1-\pi)^{t-u}$, which is again a function only of $t-u$ and illustrates the memoryless property of the Geometric distribution.

2.17 $Z \sim \text{N}(0,1)$

(a) $P[Z \leq 2.63]$ via **R** is `pnorm(q=2.63)`, or 0.9957.

(b) $P[Z > 2.63] = 1 - P[Z \leq 2.63]$, using the Complement Rule (2d), via **R** is `1-pnorm(q=2.63)`, i.e. 0.0043. Or, just subtract 1 from the answer in part (2.17a).

(c) $P[\,|Z| \leq 2.63] = 1 - 2P[Z > 2.63]$ using the symmetry of the standard normal. From part (2.17b) this is $1 - (2)(0.0043) = 0.9914$. (**R** gives it more precisely as 0.9914615.)

(d) $P[\,|Z| \geq 2.63] = 1 - P[\,|Z| < 2.63]$, using the Complement Rule (2d). From part (2.17c) this is $1 - 0.9914615 = 0.0085385$.

(e) The upper-2.5% standard normal critical point is $z_{0.025} = 1.9600$, via, e.g. `qnorm(p=0.025,lower=F)` in **R**.

(f) The upper-5% standard normal critical point is $z_{0.05} = 1.6449$, via, e.g. `qnorm(p=0.05,lower=F)` in **R**.

(g) The upper-0.5% standard normal critical point is $z_{0.005} = 2.5758$, via, e.g. `qnorm(p=0.005,lower=F)` in **R**.

(h) The upper-1% standard normal critical point is $z_{0.01} = 2.3263$, via, e.g. `qnorm(p=0.01,lower=F)` in **R**.

2.18 $X \sim \text{N}(\mu, \sigma^2)$. In all cases, we can standardize to $Z = (X - \mu)/\sigma \sim \text{N}(0,1)$ and find the probabilities as in Exercise 2.17. In **R**, however, we can also enter μ and σ directly into the `qnorm()` function.

(a) $\mu = 1.3$ and $\sigma^2 = 16$: $P[X \leq 11.82]$ is found in **R** via
`pnorm(q=11.82, mean=1.3, sd=4)`,
or 0.9957. (By the way, notice $z = [11.82 - 1.3]/4 = 2.63$.)

(b) $\mu = -2.5$ and $\sigma^2 = 9$. Find $P[X > 5.39]$ a number of different ways. Fastest is to notice that $z = [5.39 - (-2.5)]/3 = 2.63$. So $P[X > 5.39] = P[Z > 2.63] = 1 - 0.9957 = 0.0043$ from Exercise 2.17b.

(c) $\mu = 1.3$ and $\sigma^2 = 16$: notice $z = [11.82 - 1.3]/4 = 2.63$, so $P[\,|X| \leq 11.82] = P[\,|Z| \leq 2.63] = 0.9914615$ from Exercise 2.17c.

(d) $\mu = -2.5$ and $\sigma^2 = 9$. Find $P[\,|X| \geq 5.39] = P[\,|Z| \geq 2.63]$, which from Exercise 2.17d is $1 - 0.9914615 = 0.0085385$.

2.19 $X_i \sim$ i.i.d. Poisson(2.7), $i = 1, \ldots, 100$. Calculate $P[\bar{X} > 2]$.

(a) From the Hint: $P[\bar{X} > a] = P\left[\sum_{i=1}^{n} X_i > na\right]$ for $n = 100$. But, the closure of the Poisson distribution tells us that $\sum_{i=1}^{n} X_i \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right) = \text{Poisson}(270)$. Thus $P[\bar{X} > 2] = P\left[\sum_{i=1}^{n} X_i > 200\right] = 1 - P\left[\sum_{i=1}^{n} X_i \leq 199\right]$ can be found in **R** from `ppois(q=199,lambda=270)`, which calculates to 3.5544×10^{-6}. Subtract this from 1.0 to reach the final answer.

(b) From the Central Limit Theorem, $Z = (\bar{X} - \lambda)/\sqrt{\lambda/n} \overset{\cdot}{\sim} \text{N}(0,1)$. So, $P[\bar{X} > 2] = P[Z > (2 - 2.7)/\sqrt{0.027}] = P[Z > -4.2601] = 1 - P[Z \leq -4.2601] \approx 1 - \Phi(-4.2601)$. In **R**, this requires `pnorm(q=-4.2601)` or 1.0217×10^{-5}. Subtract this from 1.0 to reach the final answer.

2.20 Entropy is defined as $H(f_X) = -E[\log\{f_X(X)\}]$. Use this in the following:

(a) $X \sim \text{Bin}(1, \pi)$. (Use $\log_2$ in place of the natural logarithm.)

$$\begin{aligned} H(f_X) &= -E[\log_2\{f_X(X)\}] \\ &= -E\left[\log_2\left\{\pi^X(1-\pi)^{1-X}\right\}\right] \\ &= -E[X\log_2(\pi)] - E[(1-X)\log_2(1-\pi)] \\ &= -\log_2(\pi)E[X] - \log_2(1-\pi)E[(1-X)] \\ &= -\pi\log_2(\pi) - (1-\pi)\log_2(1-\pi). \end{aligned}$$

(b) $X \sim U(0, \theta)$. Being careful with the notation, here $\log f_X(X) = \log\{\theta^{-1}I_{(0,\theta)}(X)\}$. The expected value is an integral taken over the range of the data, where the indicator function will always be 1, thus

$$\begin{aligned} H(f_X) &= -E[\log\{\theta^{-1}I_{(0,\theta)}(X)\}] \\ &= -\int_0^\theta \theta^{-1}\log(\theta^{-1})dx = -\theta^{-1}\log(\theta^{-1})\int_0^\theta dx \\ &= -\frac{\theta}{\theta}\log(\theta^{-1}) \;=\; -\log(\theta^{-1}) = \log(\theta). \end{aligned}$$

In particular, at $\theta = 1$, the entropy is zero. [In fact, one needs to be careful: if $\theta \in (0,1)$, the entropy is negative.]

(c) $X \sim \text{Exp}(\beta)$ for $\beta > 0$.

$$\begin{aligned} H(f_X) &= -E[\log\{\beta^{-1}\exp(-X/\beta)\}] \\ &= -E[\log\{\beta^{-1}\} - X/\beta] \\ &= -\log\{\beta^{-1}\} + \beta^{-1}E[X] \\ &= \log(\beta) + \beta^{-1}\beta \\ &= 1 + \log(\beta). \end{aligned}$$

A plot follows in Figure 2.1. Notice the increase as β rises.

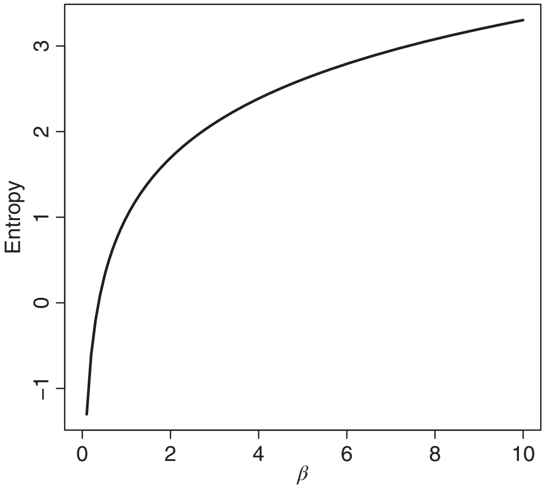

Figure 2.1 Entropy function for $X \sim \mathrm{Exp}(\beta)$ as a function of β.

(d) $X \sim \mathrm{N}(0, \sigma^2)$.

$$\begin{aligned} H(f_X) &= -E[\log\{(2\pi\sigma^2)^{-1/2}\exp(-\tfrac{1}{2}X/\sigma^2)\}] \\ &= -E[-\tfrac{1}{2}\log(2\pi\sigma^2) - \tfrac{1}{2}(X/\sigma^2)] \\ &= \tfrac{1}{2}\log(2\pi\sigma^2) + \tfrac{1}{2}E[X]/\sigma^2 \\ &= \tfrac{1}{2}\log(2\pi\sigma^2) + 0 \\ &= \tfrac{1}{2}\log(2\pi\sigma^2). \end{aligned}$$

A plot follows in Figure 2.2. Notice the increase as the standard deviation rises.

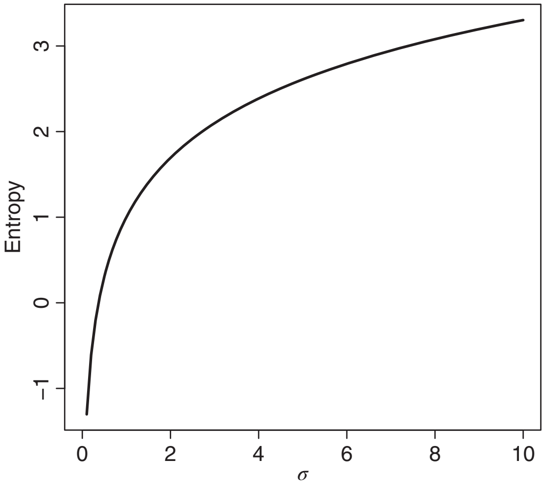

Figure 2.2 Entropy function for $X \sim \mathrm{N}(0, \sigma^2)$ as a function of the standard deviation σ.

2.21 $X \sim \chi^2(\nu)$

(a) $\nu = 8$: $P[X > 16.38] = 1 - P[X \leq 16.38]$ via **R** is `1-pchisq(q=16.38,df=8)`, or 0.0373.

(b) $\nu = 10$: $P[X \leq 1.8]$ via **R** is `pchisq(q=1.8,df=10)`, or 0.0023.

(c) $\nu = 10$: $P[X > 17.1] = 1 - P[X \leq 17.1]$ via **R** is `1-pchisq(q=17.1,df=10)`, or 0.0722.

(d) $\nu = 10$: $P[1.8 \leq X \leq 17.1] = P[X \leq 17.1] - P[X < 1.8]$ via **R** is `pchisq(q=17.1,df=10)-pchisq(q=1.8,df=10)`, or 0.9255.

(e) The upper-1% $\chi^2(5)$ critical point is $\chi^2_{0.01}(5) = 15.0863$, via, e.g. `qchisq(p=0.01,df=5,lower=F)` in **R**.

(f) The upper-5% $\chi^2(5)$ critical point is $\chi^2_{0.05}(5) = 11.0705$, via, e.g. `qchisq(p=0.05,df=5,lower=F)` in **R**.

(g) The upper-5% $\chi^2(15)$ critical point is $\chi^2_{0.05}(15) = 24.9958$, via, e.g. `qchisq(p=0.05,df=15,lower=F)` in **R**.

(h) The upper-5% $\chi^2(25)$ critical point is $\chi^2_{0.05}(25) = 37.6525$, via, e.g. `qchisq(p=0.05,df=25,lower=F)` in **R**.

2.22 $T \sim t(\nu)$.

(a) $\nu = 4$: $P[T \leq 2.63]$ via **R** is `pt(q=2.63,df=4)`, or 0.9709.

(b) $\nu = 4$: $P[T > 2.63] = 1 - P[T \leq 2.63]$, using the Complement Rule (2d), via **R** is `1-pt(q=2.63,df=4)`, or 0.0291. Or, just subtract 1 from the answer in part (2.22a).

(c) $\nu = 13$: $P[\,|T| \leq 2.63] = 1 - 2P[T > 2.63]$ using the symmetry of the t-distribution. Use **R** to find $P[T > 2.63] = 0.0104$ via `1-pt(q=2.63,df=13)`. Then $P[\,|T| \leq 2.63] = 1 - (2)(0.0104) = 0.9792$.

(d) $\nu = 13$: $P[\,|T| \geq 2.63] = 1 - P[\,|T| < 2.63]$, using the Complement Rule (2d). Using part (2.22c) this is $1 - 0.9792 = 0.0208$.

(e) The upper-5% $t(4)$ critical point is $t_{0.05}(4) = 2.1318$, via, e.g. `qt(p=0.05,df=4,lower=F)` in **R**.

(f) The upper-5% $t(11)$ critical point is $t_{0.05}(11) = 1.7959$, via, e.g. `qt(p=0.05,df=11,lower=F)` in **R**.

(g) The upper-5% $t(33)$ critical point is $t_{0.05}(33) = 1.6924$, via, e.g. `qt(p=0.05,df=33,lower=F)` in **R**.

(h) The upper-5% $t(88)$ critical point is $t_{0.05}(88) = 1.6624$, via, e.g. `qt(p=0.05,df=88,lower=F)` in **R**.

2.23 Recall that $t^2(\nu) = F(1, \nu)$. Thus if we desire the upper critical point $t_\alpha(\nu)$, we could find it as $t_\alpha(\nu) = \sqrt{F_{2\alpha}(1, \nu)}$. (After squaring, the upper tail area in the F gains contributions from both the lower and upper tails in the t. So, for a single upper tail point in the t, we check the upper-2α critical point from the F.)

2.24 $F \sim F(\nu_1, \nu_2)$.

(a) View F as the ratio of two indep. χ^2 random variables over their d.f.: $U_i \sim$ indep. $\chi^2(\nu_i)$ $(i = 1, 2)$. Thus $F = (U_1/\nu_1)/(U_2/\nu_2)$. But then clearly $1/F = (U_2/\nu_2)/(U_1/\nu_1)$ is also a ratio of indep. χ^2 random variables over their d.f., so we can write $1/F \sim F(\nu_2, \nu_1)$.

(b) By definition $F_\alpha(\nu_1, \nu_2)$ satisfies $P[F > F_\alpha(\nu_1, \nu_2)] = \alpha$. By taking reciprocals (and reversing direction of the inequality) this is then $P[1/F < 1/F_\alpha(\nu_1, \nu_2)] = \alpha$. But from part (2.24a) we know $V = 1/F \sim F(\nu_2, \nu_1)$. So, write $P[V < 1/F_\alpha(\nu_1, \nu_2)] = \alpha$. Now apply the Complement Rule (2d): $P[V \geq 1/F_\alpha(\nu_1, \nu_2)] = 1 - \alpha$. This says that $1/F_\alpha(\nu_1, \nu_2)$ satisfies the definition of the upper-$(1 - \alpha)$ critical point of V. But $V \sim F(\nu_2, \nu_1)$, so its upper-$(1 - \alpha)$ critical point is more conventionally denoted as $F_{1-\alpha}(\nu_2, \nu_1)$. This establishes the identity $1/F_\alpha(\nu_1, \nu_2) = F_{1-\alpha}(\nu_2, \nu_1)$. Lastly, take reciprocals to achieve the desired result.

2.25 $F \sim F(\nu_1, \nu_2)$.

(a) $\nu_1 = 13$, $\nu_2 = 28$: $P[F \leq 1.9]$ via **R** is `pf(q=1.9,df1=13,df2=28)`, or 0.9245.

(b) $\nu_1 = 21$, $\nu_2 = 9$: $P[F > 3.4] = 1 - P[F \leq 3.4]$, using the Complement Rule (2d), via **R** is `1-pf(q=3.4,df1=21,df2=9)`, or 0.0315. Or, one can also use `pf(q=3.4,df1=21,df2=9,lower=F)`.

(c) $\nu_1 = 1$, $\nu_2 = 4$: $P[F \geq 6.2] = 1 - P[F < 6.2]$, using the Complement Rule (2d), via **R** is `1-pf(q=6.2,df1=1,df2=4)`, or 0.0315. [One could also appeal to the relationship $t^2(4) = F(1, 4)$.]

(d) The upper-2% $F(1, 4)$ critical point is $F_{0.02}(1, 4) = 14.0396$, via, e.g. `qf(p=0.02,df1=1,df2=4,lower=F)` in **R**.

(e) The upper-5% $F(1, 4)$ critical point is $F_{0.05}(1, 4) = 7.7087$, via, e.g. `qf(p=0.05,df1=1,df2=4,lower=F)` in **R**.

(f) The upper-5% $F(8, 7)$ critical point is $F_{0.05}(8, 7) = 3.7257$, via, e.g. `qf(p=0.05,df1=8,df2=7,lower=F)` in **R**.

(g) The upper-1% $F(3, 49)$ critical point is $F_{0.01}(3, 49) = 4.2084$, via, e.g. `qf(p=0.01,df1=3,df2=49,lower=F)` in **R**.

2.26 $X \sim$ Poisson(λ). Write the p.m.f. as

$$\begin{aligned} f_X(m) &= \frac{\lambda^m \, \mathrm{e}^{-\lambda}}{m!} I_{\{0,1,\ldots\}}(m) \\ &= \exp\left\{ m \log(\lambda) - \lambda + \log\left[\frac{I_{\{0,1,\ldots\}}(m)}{m!}\right] \right\}. \end{aligned}$$

Decomposed into this form, the natural parameter is $\theta = \log(\lambda)$, and the dispersion parameter is (trivially) fixed at $\varphi = 1$. Then, $f_X(m)$ does satisfy the exponential family characterization, with $a(\varphi) = 1$, $b(\theta) = \mathrm{e}^\theta$, and

$$c(m, 1) = \log\left[\frac{I_{\{0,1,\ldots\}}(m)}{m!}\right].$$

2.27 $X \sim \text{NB}(r, \pi)$.

(a) Complete part (2.27b) first and then simply fix $r = 4$.

(b) For known positive integer r, write the p.m.f. as

$$\begin{aligned} f_X(m) &= \binom{r+m-1}{m} \pi^r (1-\pi)^m I_{\{0,1,\ldots\}}(m) \\ &= \exp\left\{ m\log(1-\pi) + r\log(\pi) + \log\left[\binom{r+m-1}{m} I_{\{0,1,\ldots\}}(m)\right]\right\}. \end{aligned}$$

Decomposed into this form, the natural parameter is $\theta = \log(1-\pi)$, and the dispersion parameter is (trivially) fixed at $\varphi = 1$. Then, $f_X(m)$ does satisfy the exponential family characterization, with $a(\varphi) = 1$, $b(\theta) = -r\log\left(1 - \mathrm{e}^{\theta}\right)$, and

$$c(m, 1) = \log\left[\binom{r+m-1}{m} I_{\{0,1,\ldots\}}(m)\right].$$

(c) Since we know $X \sim \text{Geom}(\pi) = \text{NB}(1, \pi)$, part (2.27b) tells us that the Geometric p.m.f. must also be a member of the exponential family.

(d) For the redefined parameterization in textbook Equation (2.25), set $\delta = 2$. Then, write the p.m.f. as

$$\begin{aligned} f_X(m) &= \frac{\Gamma(m+\frac{1}{2})}{m!\sqrt{\pi}}\left(\frac{2\mu}{1+2\mu}\right)^m \frac{1}{\sqrt{1+2\mu}} I_{\{0,1,\ldots\}}(m) \\ &= \exp\left\{ m\log\left(\frac{2\mu}{1+2\mu}\right) - \tfrac{1}{2}\log(1+2\mu) \right. \\ &\qquad \left. + \log\left[\frac{\Gamma(m+\frac{1}{2})}{m!\sqrt{\pi}} I_{\{0,1,\ldots\}}(m)\right]\right\}. \end{aligned}$$

Decomposed into this form, the natural parameter is $\theta = \log\{2\mu/(1+2\mu)\}$, and the dispersion parameter is (trivially) fixed at $\varphi = 1$. Then, $f_X(m)$ does satisfy the exponential family characterization, with $a(\varphi) = 1$, $b(\theta) = \frac{1}{2}\log(1 - \mathrm{e}^{\theta})$, and

$$c(m, 1) = \log\left[\frac{\Gamma(m+\frac{1}{2})}{m!\sqrt{\pi}} I_{\{0,1,\ldots\}}(m)\right].$$

2.28 $X \sim \text{Exp}(\beta)$. Write the p.m.f. as

$$\begin{aligned} f_X(x) &= \frac{1}{\beta}\,\mathrm{e}^{-x/\beta} I_{(0,\infty)}(x) \\ &= \exp\left\{ -\frac{x}{\beta} - \log(\beta) + \log[I_{(0,\infty)}(x)]\right\}. \end{aligned}$$

Decomposed into this form, the natural parameter is $\theta = -1/\beta$ (which, notice, is negative), and the dispersion parameter is (trivially) fixed at $\varphi = 1$. Then, the simple exponential p.d.f. does satisfy the larger exponential family characterization, with $a(\varphi) = 1$, $b(\theta) = -\log(-\theta)$ (again, recall that $\theta < 0$), and $c(x, 1) = \log[I_{(0,\infty)}(x)]$.

2.29 The Pareto distribution has p.d.f.

$$f_X(x) = \frac{\beta\gamma^\beta}{x^{\beta+1}} I_{(\gamma,\infty)}(x)$$

where $\beta > 0$ and $\gamma > 0$.

(a) $E[X] = \int_\gamma^\infty \beta\gamma^\beta (x)(x^{-\beta-1})\,dx$. Straightforward integration gives

$$\begin{aligned} E[X] &= \beta\gamma^\beta \int_\gamma^\infty x^{-\beta}\,dx \\ &= \beta\gamma^\beta \left[\frac{x^{-\beta+1}}{1-\beta}\right]_\gamma^\infty dx \\ &= \frac{\beta\gamma^\beta}{1-\beta}\left(\lim_{x\to\infty} x^{1-\beta} - \gamma^{1-\beta}\right) \end{aligned}$$

in which the limit diverges unless we restrict $\beta > 1$. Doing so produces $E[X] = \gamma\beta/(\beta-1)$.

(b) Recall that $\text{Var}[X] = E[X^2] - E^2[X]$. So, consider

$$E[X^2] = \int_\gamma^\infty \beta\gamma^\beta (x^2)(x^{-\beta-1})\,dx.$$

Similar integration calculations as in part (2.29a) lead to $E[X^2] = (2-\beta)^{-1}\beta\gamma^\beta\left(\lim_{x\to\infty} x^{2-\beta} - \gamma^{2-\beta}\right)$. Here the limit diverges unless we restrict $\beta > 2$. Doing so gives $E[X] = \gamma^2\beta/(\beta-2)$. Combining this with $E[X]$ in part (2.29a) yields $\text{Var}[X] = \gamma^2\beta/(\beta-2) - \gamma^2\beta^2/(\beta-1)^2$. Simplifying produces

$$\text{Var}[X] = \frac{\gamma^2\beta}{(\beta-1)^2(\beta-2)}.$$

(c) Fix $\gamma > 0$. Write the p.d.f. as

$$\begin{aligned} f_X(x) &= \frac{\beta\gamma^\beta}{x^{\beta+1}} I_{(\gamma,\infty)}(x) \\ &= \exp\left\{-\beta\log(x) + \log(\beta\gamma^\beta) - \log(x) + \log[I_{(\gamma,\infty)}(x)]\right\}. \end{aligned}$$

Decomposed into this form, the Pareto p.d.f. does not correspond to the simple exponential family in textbook Equation (2.42), but it does correspond to the extended exponential family mentioned immediately thereafter:

$$f_X(x) = \exp\left\{\frac{t(x)\theta - b(\theta)}{a(\varphi)} + c(x,\varphi)\right\}.$$

Then, with $t(x) = \log(x)$ the natural parameter is $\theta = -\beta$, and the dispersion parameter is (trivially) fixed at $\varphi = 1$. This satisfies the extended exponential family characterization, with $a(\varphi) = 1$, $b(\theta) = -\log(-\theta\gamma^{-\theta})$, and

$$c(x,1) = -\log(x) + \log[I_{(\gamma,\infty)}(x)].$$