Errata and revisions for the first printing

Title: Bayesian Logical Data Analysis

for the Physical Sciences:

A Comparative Approach with

Mathematica Support

P. C. Gregory
Department of Physics and Astronomy
University of British Columbia

June 13, 2012

Many of the errata were corrected in the second and third printing but occasionally new ones are found. The latest concerns changes to equations 4.53 to 4.61 and added footnote. The next most recent errata concerns a missing factor of 2 in equ. 11.26 and related factor of 1/2 in equations 11.28, 11.29, and the equation defining Ω_k on page 300. Before that, a change to the formula for the variance of the Beta distribution on page 117 and the free software accompanying the book entitled "Additional book examples for *Mathematica*." The latter correction concerns a missing bracket in one line of the Chapter 9 code concerning a Bayesian analysis of two independent samples. The author thanks a number of dedicated readers for pointing out some of these corrections.

Preliminaries errata

Credit for book cover photo missing

Cover photo: In the foreground, the Robert C. Byrd Green Bank Telescope (GBT), the world's largest fully steerable radio telescope. Image courtesy the National Radio Astronomy Observatory/AUI/NSF.

Library of Congress Cataloging data, Authors name misspelled

Change:
Gregroy, P. C. (Philip Christopher), 1941-

To:
Gregory, P. C. (Philip Christopher), 1941-

Chapter 1 errata

P. 6, change sentence 2 lines above Equ. (1.8)

Change:
The truth of the proposition can be represented by $p(H_0|D,I)dH$, where $p(H_0|D,I)$ is a *probability density function* (PDF).

To:
The truth of the proposition can be represented by $p(H_0|D,I)dH_0$, where $p(H_0|D,I)$ is a *probability density function* (PDF).

P. 7, change Equ. (1.12)

Change:

$$p(H_0|D_1,I) \propto p(H_0|I_0)\; p(D_1|H_0,I). \tag{1.12}$$

To:

$$p(H_0|D_1,I) \propto p(H_0|I)\; p(D_1|H_0,I). \tag{1.12}$$

Chapter 3 errata and revisions

P. 47, Equation (3.18, change i index to j in denominator.)

Change:

$$p(M_i \mid D,I) = \frac{O_{i1}}{\sum_{i=1}^{N_{\rm mod}} O_{i1}}\ , \tag{3.18}$$

To:

$$p(M_i \mid D,I) = \frac{O_{i1}}{\sum_{j=1}^{N_{\rm mod}} O_{j1}}\ , \tag{3.18}$$

P. 48, replacement Figure 3.1
Change the symbol L that appears twice in the figure to $\mathcal{L}$.

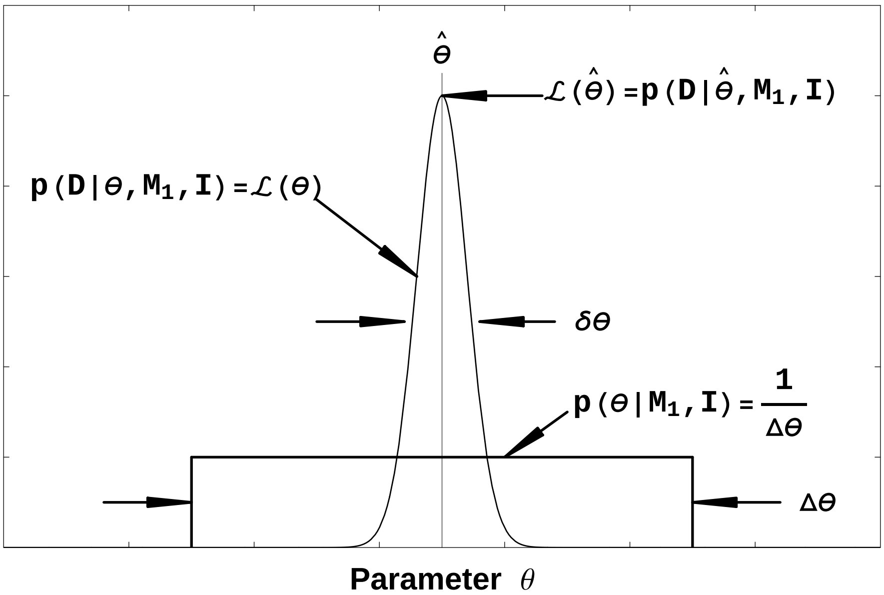


Figure 3.1: The characteristic width $\delta\theta$ of the likelihood peak and $\Delta\theta$ of the prior.

P. 57, corrected Equation (3.45)

$$\begin{aligned} p(D|M_1, I) &= \frac{(2\pi)^{-N/2}\sigma^{-N}}{\Delta T} \exp\left\{\frac{-\sum d_i^2}{2\sigma^2}\right\} \\ &\quad \times \int_{T_{\min}}^{T_{\max}} dT \ \exp\left\{\frac{T\sum d_i f_i}{\sigma^2}\right\} \exp\left\{-\frac{T^2\sum f_i^2}{2\sigma^2}\right\} \quad (3.45) \\ &= 1.131 \times 10^{-38} \end{aligned}$$

P. 57, corrected Equation (3.47)

$$
\begin{aligned}
p(D|M_1, I) &= \frac{(2\pi)^{-\frac{N}{2}}\sigma^{-N}}{\ln\left(\frac{T_{\max}}{T_{\min}}\right)} \exp\left\{\frac{-\sum d_i^2}{2\sigma^2}\right\} \\
&\times \int_{T_{\min}}^{T_{\max}} dT \, \frac{\exp\left\{\frac{T\sum d_i f_i}{\sigma^2}\right\}\exp\left\{-\frac{T^2\sum f_i^2}{2\sigma^2}\right\}}{T} \\
&= 1.239 \times 10^{-37}
\end{aligned}
\tag{3.47}
$$

P. 59, add two clarifying words to the sentence just preceding Section 3.8.1

Change:

The increase in line strength has a dramatic effect on the odds which rises to 1.6×10^{12} - -

To:

The increase in line strength has a dramatic effect on the odds which rises to a whopping 1.6×10^{12} - -

P. 67, corrected 2nd sentence following Equ. (3.67)

(change $\sigma = 5$ km s^{-1} to $\sigma = 5 \times 10^3$ km s^{-1}.)

Change existing sentence:

Assume that the probability density function for e can be described by a Gaussian with mean 0 and $\sigma = 5$ km s^{-1}.

To:

Assume that the probability density function for e can be described by a Gaussian with mean 0 and $\sigma = 5 \times 10^3$ km s^{-1}.

P. 68, corrected Figure 3.10

The scale of the vertical axis was not correctly normalized for a PDF in the original Figure. Below is the corrected version.

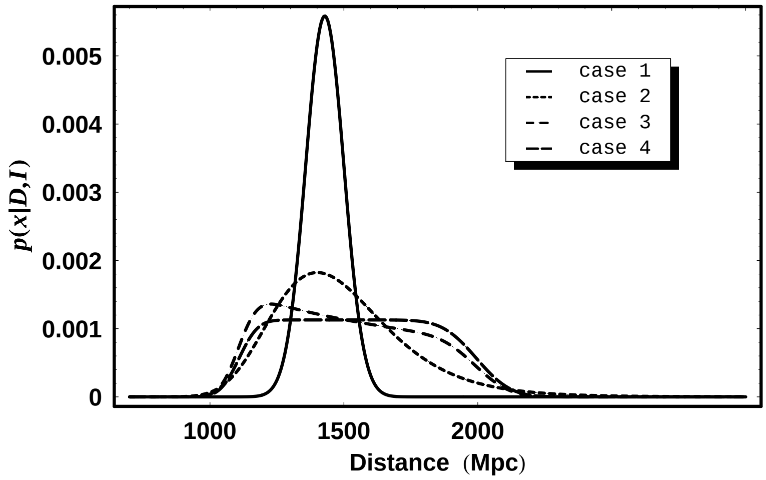


Figure 3.10: Posterior PDF for the galaxy distance, x: 1) assuming a fixed value of the Hubble constant (H_0), 2) incorporating a Gaussian prior uncertainty for H_0, 3) incorporating a uniform prior uncertainty for H_0, and 4) incorporating a Jeffreys prior uncertainty for H_0.

P. 63, ADDITION to Section 3.9, "Lessons"

6. The robustness of our model selection conclusion depends both on the choice of prior and evidence provided by the data. The spectrum of Figure 3.3 provided only very weak support for the more complicated model M_1. Thus, our conclusion depended strongly on the choice of prior. When the data provided stronger support for M_1 (e.g., spectrum of Figure 3.7), the conclusion did not depend in an important way on whether we used a uniform or Jeffreys prior. For the Figure 3.7 spectrum, the odds ranged from 1.6×10^{12} (uniform prior) to 5.3×10^{12} (Jeffreys prior). The factor of 10^{12} is so large that we are not terribly interested in whether the factor in front is 1.6 or 5.3. The lower left

panel in Figure 3.7 shows the dependence of the Log_{10} odds on the assumed value of $T_{\max}$ for this case.

P. 60, MODIFICATION to Figure 3.7

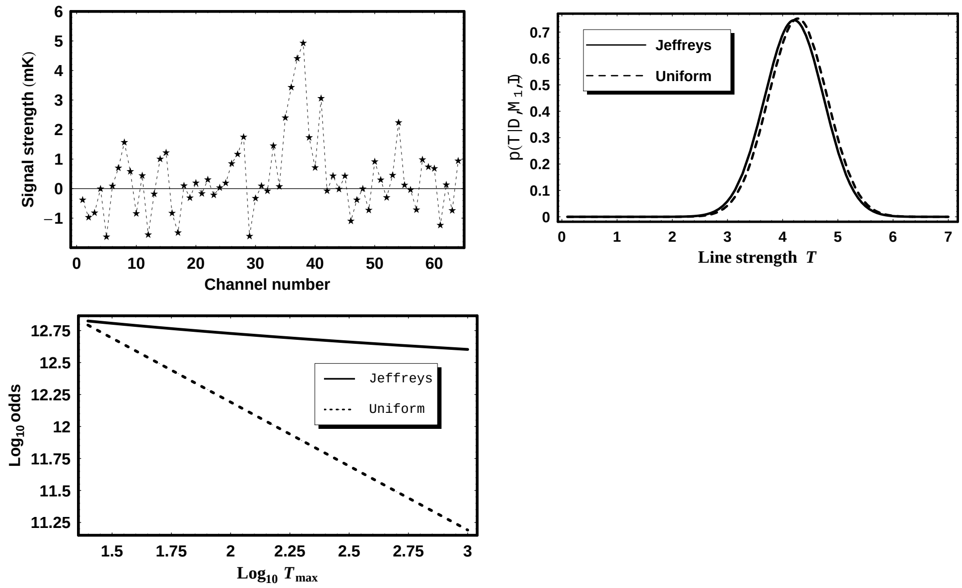


Figure 3.7: The top left panel shows a spectrum with a stronger spectral line. The top right panel shows the computed posterior PDF for the line strength. The bottom left panel shows the Log_{10} odds versus $\text{Log}_{10}\ T_{\max}$.

Chapter 4 errata and revisions

P. 73, Equations (4.5) and (4.6), LHS needs conditional on I

$$\begin{aligned} p(Q_1, \overline{Q_2}, \overline{Q_3}|I) &= p(Q_1|I)\, p(\overline{Q_2},\ \overline{Q_3}|Q_1, I) \\ &= p(Q_1|I)\, p(\overline{Q_2}|Q_1, I)\, p(\overline{Q_3}|Q_1, \overline{Q_2}, I). \end{aligned} \tag{4.4}$$

$$\begin{aligned} p(Q_1, \overline{Q_2}, \overline{Q_3}|I) &= p(Q_1|I)\, p(\overline{Q_2}|I)\, p(\overline{Q_3}|I) \\ &= p(Q|I)\, p(\overline{Q}|I)\, p(\overline{Q}|I) \\ &= p(Q|I)\, p(\overline{Q}|I)^2. \end{aligned} \tag{4.5}$$

P. 75, Corrections to Box 4.1

Change:
Needs[“Statistics ‘DiscreteDistributions’”]
To:
Needs[“Statistics`DiscreteDistributions`”]

Change:
PDF[Binomial Distribution[*n*, *p*], *r*]
To:
PDF[BinomialDistribution[*n*, *p*], *r*]
i.e., no space after Binomial.

P. 83 equations in examples 1 & 2 change $<$ to $>$.

Correct equation is:

$$p(A|B, I) > p(A|I)$$

P. 85, Section 4.7 first line

Remove the fourth word “and.”

p. 85, mistake in footnote and missing reference

In footnote change: E. T. Jaynes (1989) to E. T. Jaynes (1990)

Add reference:

Jaynes, E. T. (1990). Probability Theory as Logic. *In Maximum Entropy and Bayesian Methods*, P. F. Fougre (ed.), Dordrecht, Kluwer Academic Publishers, p. 1.

P. 91, equation 4.49

The sigma in the exponential should be changed from σ_i^2 to σ_{mi}^2. Corrected equation:

$$\begin{aligned} p(Z_i|M,\theta,I) &= \frac{1}{\sqrt{2\pi}\,\sigma_{mi}} \exp\left\{\frac{-(z_i - m(x_i|\theta))^2}{2\sigma_{mi}^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}\,\sigma_{mi}} \exp\left\{\frac{-\epsilon_i^2}{2\sigma_{mi}^2}\right\} = f_Z(z_i). \end{aligned} \tag{4.49}$$

P. 88, IMPROVEMENT to Figure 4.4

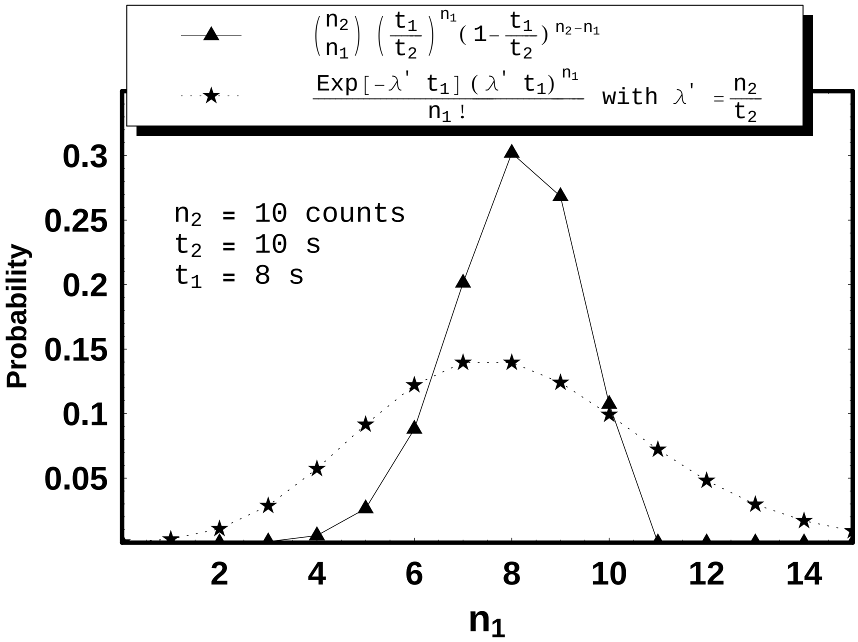


Figure 4.4: A comparison of the predictions for $p(n_1|n_2, r, t_1, t_2, I)$ based on equations (4.36) and (4.37) where we set $r = n_2/t_2$. The assumed values are $t_1 = 8$ s, $t_2 = 10$ s and $n_2 = 10$ counts.

P. 92, replace two sentences starting three lines above Equ. 4.53

Replace:
Let x_{i0} be the nominal value of the independent variable and x_i the true value. Then $\delta x_i = x_i - x_{i0}$ is the uncertainty in x_i.

By:
Let x_{ti} represent the true value of the independent variable and x_i the measured value. Then $\delta x_i = x_i - x_{ti}$ is the uncertainty in x_i.

P. 92, make changes to Equ. 4.53, 4.56, 4.58, 4.59, 4.60, and 4.61

Replace the symbol x_{i0} in all these equations by x_{ti}.

Replace the symbol z_{i0} in equations 4.56 and 4.58 by z_{ti}.

P. 93, replace sentence immediately following Equ. 4.60

Replace:
The reader is directed to Section 11.7 for a worked problem of this kind.

By:
The true value, x_{ti}, is unknown and frequently referred to as hidden [1] data.

[1] To deal with hidden data, extend the conversation to include the unknown x_{ti} by writing down the joint probability distribution for y_i, x_{ti}, specify a prior for x_{ti}, and then integrate over x_{ti}. This gets us into the realm of hierarchical Bayesian models. For a good discussion of these issues see "Some Aspects of Measurement Error in Linear Regression of Astronomical Data," by Brandon Kelly, Astrophysical Journal, 666, pp. 1489-1506, 2007.

Chapter 5 errata and revisions

P. 102, Figure 5.3, the correct labeling of the 5 diagrams is

$\alpha_3 > 0 \equiv$ positively skewed $\rightarrow$

$\alpha_3 < 0 \equiv$ negatively skewed $\rightarrow$

$\alpha_3 = 0 \equiv$ symmetric $\rightarrow$

$\alpha_4 > 3$ leptokurtic $\equiv$ highly-peaked $\rightarrow$

$\alpha_4 < 3$ platykurtic $\equiv$ flat-topped $\rightarrow$

P. 110, three lines below equation (5.29)

remove "the" before last word in the sentence.

P. 108, corrected unnumbered equation at foot of page

$$\begin{aligned}\mu_2' &= \left.\frac{d^2 m_x(t)}{dt^2}\right|_{t=0} \\ &= \left[n(n-1)(1-p+e^t p)^{n-2}(e^t p)^2 + n(1-p+e^t p)^{n-1} e^t p\right]\Big|_{t=0} \\ &= n(n-1)p^2 + np.\end{aligned}$$

P. 108, Corrections to Box 5.2

Change:

Needs[“Statistics ‘DiscreteDistributions’ ”]

To:

Needs[“Statistics`DiscreteDistributions` ”]

P. 113, Corrections to Box 5.4

Change:

Needs[“Statistics ‘ContinuousDistributions’ ”]

To:

Needs[“Statistics`ContinuousDistributions` ”]

P. 117, Correction to variance following equation (5.44)

Change:

$$\text{variance} = \frac{\alpha\beta}{(1+\beta)^2(\alpha+\beta+1)}$$

to:

$$\text{variance} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

P. 130, correct equation (5.61)

$$f(x) = \begin{cases} \frac{1}{\theta}\exp(-x/\theta), & \text{for } x > 0, \theta > 0 \\ 0, & \text{elsewhere.} \end{cases} \tag{5.61}$$

P. 132, Corrections to Box 5.6

Change:

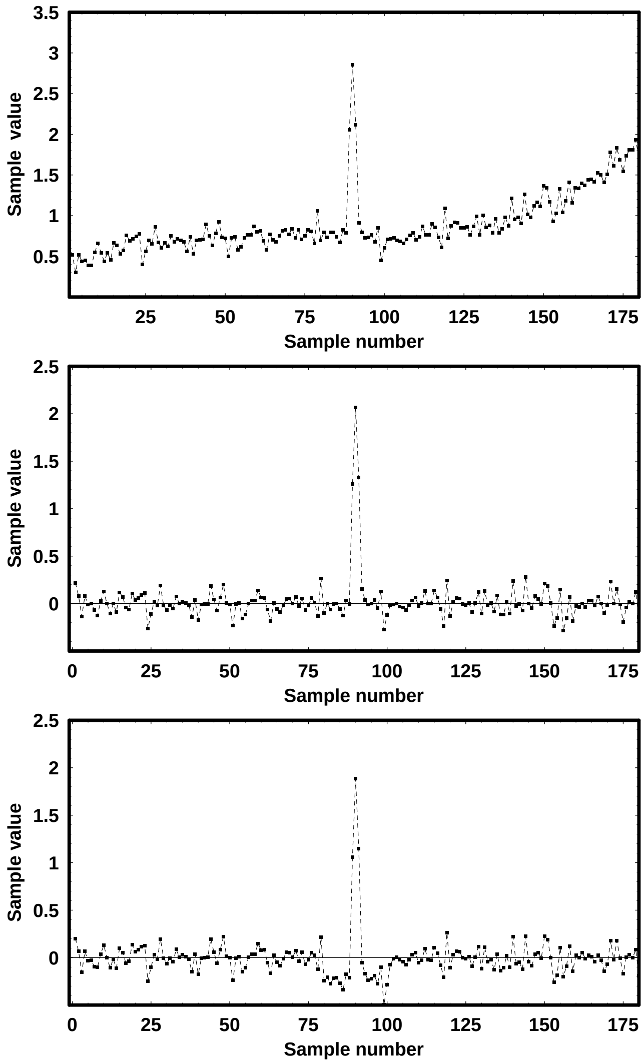
Needs[“Statistics ‘ContinuousDistributions’ ”]

To:

Needs[“Statistics`ContinuousDistributions` ”]

P. 104, IMPROVEMENT to Figure 5.5, change ordinate label

From: ‘Signal strength’ to ‘Sample value’



Chapter 6 errata and revisions

P. 146, second sentence below equation (6.28)

Change current sentence:

Thus, the probability that the random variable $\sigma^2 < 0.49S^2 = 1\%$.

To:

Thus, the probability that $\sigma^2 < 0.49S^2 = 1\%$.

P. 147, Corrections to Box 6.1

Change:

Needs["Statistics 'ContinuousDistributions' "]

To:

Needs["Statistics\`ContinuousDistributions\` "]

P. 149, Corrections to Box 6.2

Change:

Needs["Statistics 'ContinuousDistributions' "]

To:

Needs["Statistics\`ContinuousDistributions\` "]

P. 151, Corrections to Box 6.3

Change:

Needs["Statistics 'ContinuousDistributions' "]

To:

Needs["Statistics\`ContinuousDistributions\` "]

P. 156, Corrections to Box 6.4

Change:

Needs["Statistics 'ConfidenceIntervals' "]

To:

Needs["Statistics\`ConfidenceIntervals\` "]

P. 158, change equation (6.62)

Change existing equation:

$$\bar{x} - \bar{y} \pm s_p \ t_{1-\frac{\alpha}{2},(n_x+n_y-2)}. \tag{6.62}$$

To:

$$\bar{x} - \bar{y} \pm s_D \ t_{1-\frac{\alpha}{2},(n_x+n_y-2)}. \tag{6.62}$$

P. 159, change first equation with no number

Change existing equation:

$$T = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S_D}$$

To:

$$T = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S_p}$$

P.159, change equation (6.63)

Change existing equation:

$$S_D^2 = \frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}, \tag{6.63}$$

To:

$$S_p^2 = \frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}, \tag{6.63}$$

Chapter 7 errata and revisions

P. 166, Corrections to Box 7.1

Change:

Needs[“Statistics ‘HypothesisTests’ ”]

To:

Needs[“Statistics`HypothesisTests` ”]

P. 168, change equation (7.9)

Change existing equation:

$$H_0 \equiv \ \mu_1 - \ = 0. \tag{7.9}$$

To:

$$H_0 \equiv \ \mu_1 - \mu_2 = 0. \tag{7.9}$$

P. 169, Corrections to Box

Change:

Needs[“Statistics ‘HypothesisTests’ ”]

To:

Needs[“Statistics`HypothesisTests` ”]

P. 170, Corrections to Box

Change:

Needs[“Statistics ‘HypothesisTests’ ”]

To:

Needs[“Statistics`HypothesisTests` ”]

P. 181, Corrections to Box

Change:

Needs[“Statistics ‘HypothesisTests’ ”]

To:

Needs[“Statistics`HypothesisTests` ”]

P. 173, change sentence 2 lines below Example 3

Change:

Here, the grades are discrete or binned and the distribution we to is continuous.

To:

Here, the grades are discrete or binned and the distribution we are comparing to is continuous.

P. 175, change subscript on the term on LHS of the equation

Change:

$$\chi^2_{\nu-1} = \sum_{i=1}^{2} \frac{(N_i - np_i)^2}{np_i} = \frac{(54-63)^2}{63} + \frac{(36-27)^2}{27} = 4.29$$

To:

$$\chi^2_{\nu} = \sum_{i=1}^{2} \frac{(N_i - np_i)^2}{np_i} = \frac{(54-63)^2}{63} + \frac{(36-27)^2}{27} = 4.29$$

P. 176, first equation in Example 2

Change existing equation:

$$p(n) = \frac{(\lambda)^n e^{-\lambda}}{n!},$$

To:

$$p(x) = \frac{(\lambda)^x e^{-\lambda}}{x!},$$

P. 181, 2nd line in Mathematica box

Change:

VarianceRatioTest[data1,data2,ratio,FullReport − <True]

To:

VarianceRatioTest[data1,data2,ratio,FullReport − >True]

P. 165, replace sentence starting 4 lines above equation (7.4)

Change:

The question of how unlikely is this value of χ^2 is usually interpreted in terms of the area in the tail of the χ^2 distribution to the right of this line which is called the *P-value* or *significance*.

To:

The question of how unlikely is this value of χ^2, is by convention interpreted in terms of the area in the tail of the χ^2 distribution beyond this line. This area is called the *P-value* or *significance*. We need to specify an area because there is no probability associated with a point.

Chapter 8 errata and revisions

P. 192, 3 lines above Section 8.7

Change:
This yields a set of M equations which can be solved for the $\{p_i\}$.
To:
This set of M equations, together with the constraint equation, can be solved for the $\{p_i\}$ and the Lagrange multiplier λ.

P. 203, second-to-last equation on the page

Change:

$$S(f, m) = -\int dy \left[f(x, y) - m(x, y) - f(x, y) \ln \left(\frac{f(x, y)}{m(x, y)} \right) \right],$$

To:

$$S(f, m) = -\int \int \left[f(x, y) - m(x, y) - f(x, y) \ln \left(\frac{f(x, y)}{m(x, y)} \right) \right] dxdy,$$

P. 206, equation (8.64)

Change:

$$p(d_j|I_{ij},B) = p(e_j|I_{ij},B) \propto \exp\left[-\frac{e_j^2}{2\sigma_j^2}\right] = \exp\left[-\left(\frac{d_j - I_{ij}}{2\sigma_j}\right)^2\right] \tag{8.64}$$

To:

$$p(d_j|I_{ij},B) = p(e_j|I_{ij},B) \propto \exp\left[-\frac{e_j^2}{2\sigma_j^2}\right] = \exp\left[-\frac{1}{2}\left(\frac{d_j - I_{ij}}{\sigma_j}\right)^2\right] \tag{8.64}$$

P. 206, equation (8.65)

Change:

$$\begin{aligned} p(D|I_i,B) &\propto \prod_{j=1}^{m} \exp\left[-\left(\frac{d_j - I_{ij}}{2\sigma_j}\right)^2\right] \\ &= \exp\left[-\frac{1}{2}\sum_{j=1}^{m}\left(\frac{d_j - I_{ij}}{\sigma_j}\right)^2\right] = \exp\left[-\frac{\chi^2}{2}\right]. \end{aligned} \tag{8.65}$$

To:

$$\begin{aligned} p(D|I_i,B) &\propto \prod_{j=1}^{m} \exp\left[-\frac{1}{2}\left(\frac{d_j - I_{ij}}{\sigma_j}\right)^2\right] \\ &= \exp\left[-\frac{1}{2}\sum_{j=1}^{m}\left(\frac{d_j - I_{ij}}{\sigma_j}\right)^2\right] = \exp\left[-\frac{\chi^2}{2}\right]. \end{aligned} \tag{8.65}$$

IMPROVEMENT, Add a chapter 8 Summary section

8.10 Summary

We started this chapter by introducing the concept of testable information. We learned how to measure the uncertainty of a probability distribution using Shannon's entropy measure. This led to our use of the maximum entropy principle (MaxEnt) to convert the testable information into a unique probability distribution.

We examined three simple constraint problems and derived their corresponding probability distributions. In the course of this examination, we gained further insight into the special properties of a Gaussian distribution. It says that unless we have some additional prior information which justifies the use of some other sampling distribution, then use a Gaussian sampling distribution. It makes the fewest assumptions about the information you don't have and will lead to the most conservative estimates (i.e., greater uncertainty than you would get from choosing a more appropriate distribution based on more information).

We derived the multivariate Gaussian distribution from the MaxEnt principle, given constraint information on the variances and covariances of multiple variables. This provides a means for incorporating relevant prior information about correlations between data points, for example.

We also explored the application of MaxEnt to situations where the constraints were uncertain and considered an application to image reconstruction. Finally we considered another promising Bayesian image reconstruction/compression technique called the Pixon method.

Chapter 9 errata and revisions

P. 225, equation (9.43)
to eliminate the double negative in the exponent and make the single negative sign in the other exponents more visible by adding a separator.

Change:

$$
\begin{aligned}
p(\sigma|D,I) &= \frac{(2\pi)^{-\frac{N}{2}} \frac{1}{R_\mu \ln \frac{\sigma_{\rm H}}{\sigma_{\rm L}}} \sigma^{-(N+1)} e^{-\frac{-Nr^2}{2\sigma^2}} \int_{\mu_{\rm L}}^{\mu_{\rm H}} e^{-\frac{N(\mu-\overline{d})^2}{2\sigma^2}}\, d\mu}{(2\pi)^{-\frac{N}{2}} \frac{1}{R_\mu \ln \frac{\sigma_{\rm H}}{\sigma_{\rm L}}} \int_{\sigma_{\rm L}}^{\sigma_{\rm H}} \sigma^{-(N+1)} e^{-\frac{-Nr^2}{2\sigma^2}} \int_{\mu_{\rm L}}^{\mu_{\rm H}} e^{-\frac{N(\mu-\overline{d})^2}{2\sigma^2}}\, d\mu d\sigma} \\
p(\sigma|D,I) &= \frac{\sigma^{-(N+1)} e^{-\frac{-Nr^2}{2\sigma^2}} \int_{\mu_{\rm L}}^{\mu_{\rm H}} e^{-\frac{N(\mu-\overline{d})^2}{2\sigma^2}}\, d\mu}{\int_{\sigma_{\rm L}}^{\sigma_{\rm H}} \sigma^{-(N+1)} e^{-\frac{-Nr^2}{2\sigma^2}} \int_{\mu_{\rm L}}^{\mu_{\rm H}} e^{-\frac{N(\mu-\overline{d})^2}{2\sigma^2}}\, d\mu d\sigma} \\
&= \frac{\sigma^{-(N+1)} e^{-\frac{-Nr^2}{2\sigma^2}} \sqrt{2\pi}\frac{\sigma}{\sqrt{N}}}{\int_{\sigma_{\rm L}}^{\sigma_{\rm H}} \sigma^{-(N+1)} e^{-\frac{-Nr^2}{2\sigma^2}} \sqrt{2\pi}\frac{\sigma}{\sqrt{N}}\, d\sigma} \\
&= C\sigma^{-N} e^{-\frac{-Nr^2}{2\sigma^2}}. \qquad (9.43)
\end{aligned}
$$

To:

$$
\begin{aligned}
p(\sigma|D,I) &= \frac{(2\pi)^{-\frac{N}{2}} \frac{1}{R_\mu \ln \frac{\sigma_{\rm H}}{\sigma_{\rm L}}} \sigma^{-(N+1)} e^{-\left(\frac{Nr^2}{2\sigma^2}\right)} \int_{\mu_{\rm L}}^{\mu_{\rm H}} e^{-\ \frac{N(\mu-\overline{d})^2}{2\sigma^2}}\, d\mu}{(2\pi)^{-\frac{N}{2}} \frac{1}{R_\mu \ln \frac{\sigma_{\rm H}}{\sigma_{\rm L}}} \int_{\sigma_{\rm L}}^{\sigma_{\rm H}} \sigma^{-(N+1)} e^{-\left(\frac{Nr^2}{2\sigma^2}\right)} \int_{\mu_{\rm L}}^{\mu_{\rm H}} e^{-\ \frac{N(\mu-\overline{d})^2}{2\sigma^2}}\, d\mu d\sigma} \\
p(\sigma|D,I) &= \frac{\sigma^{-(N+1)} e^{-\left(\frac{Nr^2}{2\sigma^2}\right)} \int_{\mu_{\rm L}}^{\mu_{\rm H}} e^{-\ \frac{N(\mu-\overline{d})^2}{2\sigma^2}}\, d\mu}{\int_{\sigma_{\rm L}}^{\sigma_{\rm H}} \sigma^{-(N+1)} e^{-\left(\frac{Nr^2}{2\sigma^2}\right)} \int_{\mu_{\rm L}}^{\mu_{\rm H}} e^{-\ \frac{N(\mu-\overline{d})^2}{2\sigma^2}}\, d\mu d\sigma} \\
&= \frac{\sigma^{-(N+1)} e^{-\left(\frac{Nr^2}{2\sigma^2}\right)} \sqrt{2\pi}\frac{\sigma}{\sqrt{N}}}{\int_{\sigma_{\rm L}}^{\sigma_{\rm H}} \sigma^{-(N+1)} e^{-\left(\frac{Nr^2}{2\sigma^2}\right)} \sqrt{2\pi}\frac{\sigma}{\sqrt{N}}\, d\sigma} \\
&= C\sigma^{-N} e^{-\left(\frac{Nr^2}{2\sigma^2}\right)}. \qquad (9.43)
\end{aligned}
$$

P. 221 , 7 lines below equation (9.34)

Change:
However, at $|x - \bar{d}| = 2.3$
To:
However, at $|\mu - \bar{d}| = 2.3$

P. 238, replacement for Figure 9.7

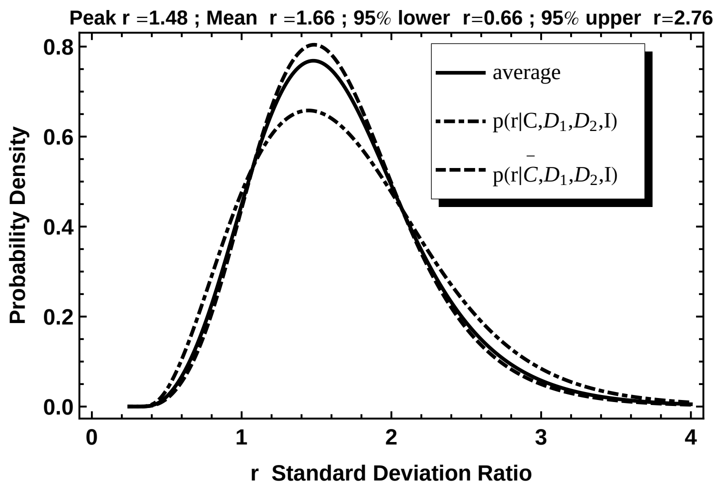


Figure 9.7: Probability density for the ratio of the standard deviations. Three probability density functions are shown: (1) the probability for the ratio of standard deviations given that the means are the same (dotted line), (2) the probability for the ratio of standard deviations given that the means are different (dashed line), (3) the probability for the ratio of standard deviations independent whether or not the means are the same (solid line).

P. 239, replacement for Figure 9.8

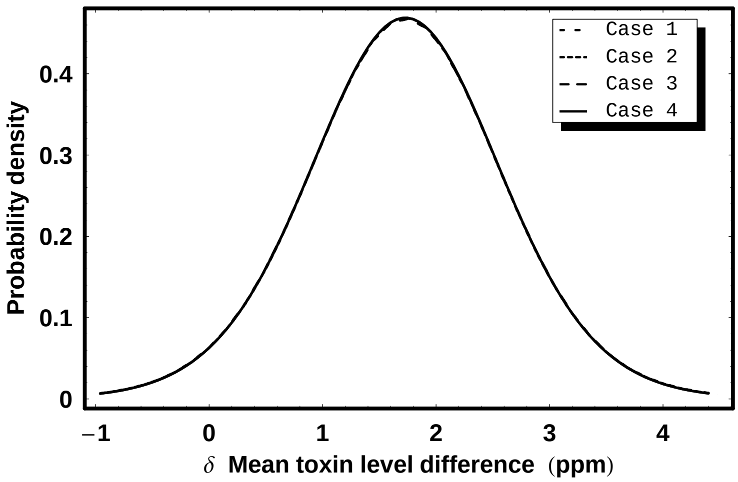


Figure 9.8: Posterior probability of the differences in the mean river sediment toxin concentration for the four different choices of prior boundaries given in Table 9.2. The effects of different choices of prior boundaries are barely discernible near the peak.

P. 240, replacement for Figure 9.9

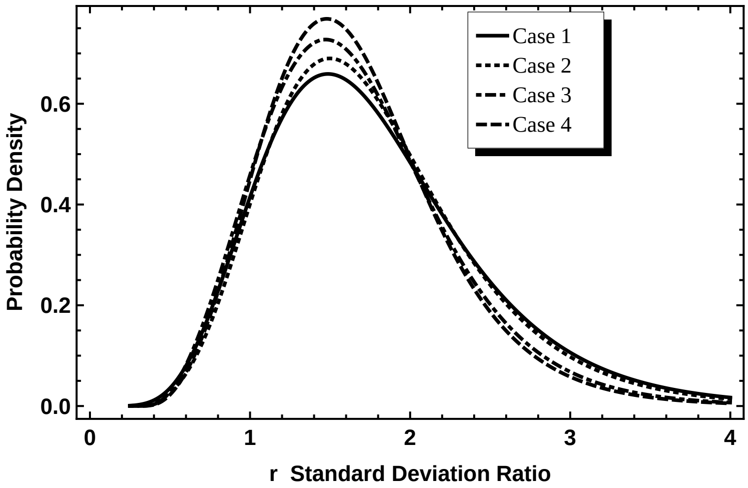


Figure 9.9: Posterior probability for the ratio of standard deviations of the river sediment toxin concentration for the four different choices of prior boundaries given in Table 9.2.

Chapter 10 errata and revisions

P. 258, starting with Equation (10.64)
To help clarify the point of this section the material between Equations (10.64) and (10.66) has been rewritten. This includes modifying Equation (10.64)

Replace Equation (10.64) and modify the text leading up to Equation (10.66).

Replacement:

$$\begin{aligned} p(\{A_\alpha\} \mid D, I) &= C' \exp\left\{-\frac{\Delta Q}{2\sigma^2}\right\} \\ &= C' \exp\left\{-\frac{1}{2}\sum_{\alpha\beta} \delta A_\alpha \left(\frac{\psi_{\alpha\beta}}{\sigma^2}\right) \delta A_\beta\right\} \\ &= C' \exp\left\{-\frac{1}{2}\sum_{\alpha\beta} \delta A_\alpha \Psi_{\alpha\beta}\, \delta A_\beta\right\}. \end{aligned} \tag{10.64}$$

where

$$C' = C\ \exp\left\{-\frac{Q_{\min}}{2\sigma^2}\right\}, \tag{10.65}$$

is an adjusted normalization constant.

Apart from the normalization constant, Equation (10.64) has the same form as Equation (10.41), which is the equation of a multivariate Gaussian. Both $\Psi_{\alpha\beta}$ and $[\mathbf{E}^{-1}]_{ij}$ are symmetric matrices. From this we conclude that the posterior $p(\{A_\alpha\} \mid D, I)$ is also a multivariate Gaussian. That means there is a single peak in the joint posterior. We will investigate further parallels between Equations (10.64) and (10.41) in Section 10.5.3.

We now write Equation (10.64) in two different matrix forms, and examine the cross section of the posterior at constant $p(\{A_\alpha\} \mid D, I)$, for the $M = 2$ case. For a two dimensional Gaussian posterior we expect this cross section to be an ellipse. If we let $\delta\mathbf{A}$ be a column matrix of δA_α values, then Equation (10.64) can be rewritten as

$$p(\{A_\alpha\} \mid D, I) = C' \exp\left\{-\frac{\delta\mathbf{A}^T \boldsymbol{\psi}\, \delta\mathbf{A}}{2\sigma^2}\right\}, \tag{10.66}$$

P. 262 , Remove explanation mark in answer to 2nd question
Could be confused with a factorial sign.

Change:
Answer: 2!
To:
Answer: 2

P. 268, replace paragraph following Equation (10.102)
Note: there is a new Appendix F dealing with highly correlated parameters.

Replacement Paragraph:
In the extreme case of $\rho = \pm 1$, the elliptical contours will be infinitely wide in one direction. In this case, the parameter error bars σ_α and σ_β will be infinite as well, saying that our individual estimates of A_α and A_β are completely unreliable, but we can still infer a linear combination of the parameters quite well. See Appendix F for a discussion of this topic.

P. 269, change second sentence of Problem

From:
We are now in a position to evaluate the errors for the marginal posterior density functions for the intercept, A_1, and slope, A_2, from the diagonal elements of $\mathbf{V} = \mathbf{\Psi}^{-1} = (\mathbf{G}^T\mathbf{E}^{-1}\mathbf{G})^{-1}$ which is given by
To:
Evaluate the errors for the joint and marginal posterior density functions for the intercept, A_1, and slope, A_2, from the diagonal elements of $\mathbf{V} = \mathbf{\Psi}^{-1} = (\mathbf{G}^T\mathbf{E}^{-1}\mathbf{G})^{-1}$ which is given by

P. 276, interchange A and A' in Equation (10.126)

Change:

$$
\begin{aligned}
O_{12} &= \frac{p(M_1|I)}{p(M_2|I)} \times \frac{p(D \mid M_1, I)}{p(D \mid M_2, I)} = 1 \times \frac{p(D \mid M_1, I)}{p(D \mid M_2, I)} \\
&= e^{\Delta\chi^2_{\min}/2}\,(2\pi)^{(M_1-M_2)/2}\sqrt{\frac{\det \mathbf{V}_1}{\det \mathbf{V}_2}}\,\frac{\prod_{\alpha=1}^{M_2}\Delta A_\alpha}{\prod_{\alpha=1}^{M_1}\Delta A'_\alpha},
\end{aligned}
\tag{10.126}
$$

To:

$$
\begin{aligned}
O_{12} &= \frac{p(M_1|I)}{p(M_2|I)} \times \frac{p(D \mid M_1, I)}{p(D \mid M_2, I)} = 1 \times \frac{p(D \mid M_1, I)}{p(D \mid M_2, I)} \\
&= e^{\Delta\chi^2_{\min}/2}\,(2\pi)^{(M_1-M_2)/2}\sqrt{\frac{\det \mathbf{V}_1}{\det \mathbf{V}_2}}\,\frac{\prod_{\alpha=1}^{M_2}\Delta A'_\alpha}{\prod_{\alpha=1}^{M_1}\Delta A_\alpha},
\end{aligned}
\tag{10.126}
$$

P. 277, change last word on the page from ‘is’ to ‘are’

P. 277, replace 3rd line, the prime needs to be a superscript.

Replace:

the ratio of the prior ranges ΔA_α and $\Delta A\prime_\alpha$ for these parameters will cancel.

By:

the ratio of the prior ranges ΔA_α and $\Delta A'_\alpha$ for these parameters will cancel.

P. 277, change Equation (10.130)

Change:

$$\mathbf{E}^{-1} = \frac{1}{\sigma^2} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \tag{10.130}$$

To:

$$\mathbf{E}^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \tag{10.130}$$

P. 282, change Equation (10.148)

Replace:

$$F = \frac{\sum_{i=1}^{N} (d_i - f_i)^2 - \sum_{i=1}^{N} (d_i - f_{+i})^2}{\sum_{i=1}^{N} (d_i - f_{+i})^2}, \tag{10.148}$$

By:

$$F = (\nu - 1) \frac{\sum_{i=1}^{N} (d_i - f_i)^2 - \sum_{i=1}^{N} (d_i - f_{+i})^2}{\sum_{i=1}^{N} (d_i - f_{+i})^2}, \tag{10.148}$$

P. 283, change last 2 sentences above Section 10.9 Summary

Replace:

Substituting these values into Equation (10.147) yields $f = 10.5$. This corresponds to a P-value = 0.2%. On the basis of this F-test, we can reject the simpler model M_2 at a 99.8% confidence level.

By:

Substituting these values into Equation (10.147) yields $f = 11.23$. This corresponds to a P-value = 0.14%. On the basis of this F-test, we can reject the simpler model M_2 at a 99.86% confidence level.

P. 284, Change first sentence of question 2

Change:

Compute and plot the ellipse that defines the 68.3% and 95.4% joint credible region for the slope and intercept, for the data given in Table 10.3.

To:

Compute and plot the ellipse that defines the 68.3% and 95.4% joint credible region for the slope and intercept, for the data given in Table 10.3 (see question 1).

P. 280, useful ADDITION, insert new paragraph above sentence
"Note: *Mathematica* provides a command called"

Insert paragraph:

As an example, we use the χ^2 hypothesis test to see if we can reject M_2 (no line exists) in the spectral line problem of Section 3.6. The best fit for M_2, with 64 degrees of freedom, $\chi^2_{\nu=64} = 57.13$. The computed significance (P-value) of this test is **GammaRegularized**$[\frac{\mathbf{64}}{\mathbf{2}}, \frac{\mathbf{57.13}}{\mathbf{2}}] = 0.72$. On the basis of this test, our confidence in rejecting the simpler model M_2 is only 28%.

Chapter 11 errata and revisions

P. 295, Equation 11.20, put bracket around product

Replace:

$$\begin{aligned} p(D|\theta, M, I) &= (2\pi)^{-N/2} \prod_{i=1}^{N} \sigma_i^{-1} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{(d_i - f_i)^2}{\sigma_i^2}\right] \\ &= C \exp\left[-\frac{\chi^2(\theta)}{2}\right]. \end{aligned} \tag{11.20}$$

By:

$$\begin{aligned} p(D|\theta, M, I) &= (2\pi)^{-N/2} \left\{\prod_{i=1}^{N} \sigma_i^{-1}\right\} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{(d_i - f_i)^2}{\sigma_i^2}\right] \\ &= C \exp\left[-\frac{\chi^2(\theta)}{2}\right]. \end{aligned} \tag{11.20}$$

P. 299 , replace equation 11.26 (missing factor of 2)

and equations 11.28 and 11.29 (missing factor of 1/2)

Replace:

$$\nabla\chi^2(\theta) \approx \nabla\chi^2(\theta_c) + \boldsymbol{\kappa}\ \delta\boldsymbol{\theta}. \tag{11.26}$$

By:

$$\nabla\chi^2(\theta) \approx \nabla\chi^2(\theta_c) + 2\boldsymbol{\kappa}\ \delta\boldsymbol{\theta}. \tag{11.26}$$

Replace:

$$\boldsymbol{\kappa}\ \delta\boldsymbol{\theta} = -\nabla\chi^2(\theta_c) \tag{11.28}$$

By:

$$\boldsymbol{\kappa}\ \delta\boldsymbol{\theta} = -\frac{1}{2}\nabla\chi^2(\theta_c) \tag{11.28}$$

Replace:

$$\hat{\theta} = \theta_c - \boldsymbol{\kappa}^{-1} \nabla \chi^2(\theta_c) \tag{11.29}$$

By:

$$\hat{\theta} = \theta_c - \boldsymbol{\kappa}^{-1} \frac{1}{2} \nabla \chi^2(\theta_c) \tag{11.29}$$

P. 300 , definition of Ω_k following equation 11.32 (missing factor of 1/2),

and again 6 lines below equation 11.33

Replace:

$\Omega_k = -\partial\chi^2(\theta_c)/\partial\theta_k$

By:

$\Omega_k = -\frac{1}{2}\partial\chi^2(\theta_c)/\partial\theta_k$

P. 304 , first equation, missing an f

Replace:

model[*f*, *f*_] := *a*0 + *a*1 line[*f*1] + *a*2 line[*f*, *f*2].

By:

model[*f*, *f*_] := *a*0 + *a*1 line[*f*, *f*1] + *a*2 line[*f*, *f*2].

P. 306, missing footnotes 4 and 5

[4] The quantity **result**[[**3,2**]] is $\mathbf{V}^*$ expressed in *Mathematica's* **MatrixForm**. To compute the determinant of this matrix we need to extract the argument of **MatrixForm** which is given by **result**[[**3,2**]][[**1**]].

[5] A fully Bayesian way of handling this would be to treat k as a parameter and marginalize over a prior range for k.

P. 307, please ignore Section 11.7 entirely,
for an explanation refer to ch. 4, p. 93 revisions.

Chapter 12 errata and revisions

P. 316, Corrections to Box

Change:

Needs[“Statistics ‘MultinormalDistribution’ ”]

To:

Needs[“Statistics`MultinormalDistribution` ”]

P. 316, *Mathematica* box, remove extra opening curly bracket on 5th line.

Replace:

dist2 = MultinormalDistribution[{{4, 0}, {{2, 0.8}, {0.8, 2}}]

By:

dist2 = MultinormalDistribution[{4, 0}, {{2, 0.8}, {0.8, 2}}]

P. 349, Section 12.12, replace existing problems with the following.

1. In Section 12.6, we used both the Metropolis-Hastings and parallel tempering (PT) versions of MCMC to re-analyze the toy spectral line problem of Section 3.6. A program to perform automatic PT calculations is given in the Markov Chain Monte Carlo section of the *Mathematica* tutorial. Use this program to analyze the spectrum given in Table 12.4, for $n = 20,000$ to $50,000$ iterations, depending on the speed of your computer (try it out first with only two tempering levels). As part of your solution, include figures like 12.7 and 12.8, and compute the quasi-Monte Carlo estimate of the Bayes factor, used to compare the two competing models. Explain how you arrived at your choice for the number of burn-in samples.

   The prior information is the same as that assumed in Section 12.6. Theory predicts the spectral line has a Gaussian shape with a line width $\sigma_L = 2$ frequency channels, and line center between channels 1 and 44. The noise in each channel is known to be Gaussian with a $\sigma = 1.0$ mK and the spectrometer output is in units of mK.

2. Repeat the analysis of problem 1 only this time assume the line width is also uncertain. Adopt a uniform prior for the line width (σ_l), with upper and lower bounds of 0.5 and 4 frequency channels, respectively.

You will need to modify the parallel tempering MCMC program to allow for the addition of the line width parameter. You will also need to specify prior bounds on σ_l and a starting σ for the line width Gaussian proposal distribution (typically 10% of the prior range). Your solution should include a plot of the marginal probability distribution for each of the three parameters and a calculation of the Bayes factor for comparing the two models. Justify your choice for the number of burn-in samples.

3. (Difficult problem) In Section 11.6, we illustrated the solution of a simple nonlinear model fitting problem using *Mathematica's* **NonlinearRegress**, which implements the Levenberg-Marquardt method. In this problem we want to analyze the same spectral line data (Table 12.5) using the experimental APT MCMC software given in the *Mathematica* tutorial and discussed in Section 12.8. It will yield a fully Bayesian solution to the problem without the need to assume the asymptotic normal approximation, or, assume the Laplacian approximations for computing the Bayes factor and marginals. In general, MCMC solutions come into their own for higher dimensional problems but it is desirable to gain experience working with simpler problems.

   Modify the APT MCMC software to analyze this data for the two models described in Section 11.6.

   In *Mathematica*, model 1 has the form:
   $\mathbf{model}[\boldsymbol{a0\_}, \boldsymbol{a1\_}, \boldsymbol{f1\_}] := \boldsymbol{a0} + \boldsymbol{a1}\ \mathbf{line}[\boldsymbol{f1}]$.
   where
   $\mathbf{line}[\boldsymbol{f1\_}] := \dfrac{\mathbf{Sin}[2\pi(\boldsymbol{f} - \boldsymbol{f1})/\boldsymbol{\Delta f}]}{2\pi(\boldsymbol{f} - \boldsymbol{f1})/\boldsymbol{\Delta f}}$ and $\boldsymbol{\Delta f} = \mathbf{1.5}$.

   Model 2 has the form:
   $\mathbf{model}[\boldsymbol{a0\_}, \boldsymbol{a1\_}, \boldsymbol{a2\_}, \boldsymbol{f1\_}, \boldsymbol{f2\_}] := \boldsymbol{a0} + \boldsymbol{a1}\ \mathbf{line}[\boldsymbol{f1}] + \boldsymbol{a2}\ \mathbf{line}[\boldsymbol{f2}]$,
   where $f2$ is assumed to be the higher frequency line.

   Adopt uniform priors for all parameters and assume a prior range from 0 to 10 for $a0, a1$ and $a2$. For the frequency parameters use a prior range from 1.0 to 5.0. Note: for a multi-spectral line model, each spectral peak is free to move through the full frequency range which can result in the occurrence of degenerate peaks in the joint posterior. It is therefore necessary to redefine the parameters after the MCMC iterations are terminated, in such a way that $a1$ and $f1$ correspond to the lowest frequency spectral line parameters, etc. Please refer to section 5 of a recent paper that discusses a suitable multi-frequency prior (P. C. Gregory, Monthly Notices of the Royal Astronomical Society **374**, p. 1321, 2007).

Chapter 13 errata and revisions

P. 355, change the line below Equ. 13.5. Replace j subscript by i.

Replace:

where N is the number of data values and $\overline{d^2} = \frac{1}{N}\sum_j \ d_i^2$ is the mean

By:

where N is the number of data values and $\overline{d^2} = \frac{1}{N}\sum_i \ d_i^2$ is the mean

P. 356, change Equ. 13.6.

Replace:

$$\begin{aligned} C(f_n) &= \frac{1}{N}\left|\sum_{j=1}^{N} d_j e^{i2\pi f_n t_j}\right|^2 \\ &= \frac{1}{N}\ |\mathrm{FFT}|^2 \\ \text{or}\quad C(n) &= \frac{1}{N}\left|\sum_{j=1}^{N} d_j e^{i2\pi \frac{nj}{N}}\right|^2 \\ &= \frac{|H_n|^2}{N}, \end{aligned} \tag{13.6}$$

By:

$$\begin{aligned} C(f_n) &= \frac{1}{N}\left|\sum_{j=1}^{N} d_j e^{i2\pi f_n t_j}\right|^2 \\ &= \frac{1}{N}\ |\mathrm{FFT}|^2 \\ \text{or}\quad C(n) &= \frac{1}{N}\left|\sum_{j=1}^{N} d_j e^{i2\pi \frac{(n-1)(j-1)}{N}}\right|^2 \\ &= \frac{|H_n|^2}{N}, \end{aligned} \tag{13.6}$$

P. 358, change material in Box 13.1.

Replace:

> Note: *Mathematica* uses a slightly different definition
> of H_n to that given in equation (13.6),
> which we designate by $[H_n]_{\text{Math}}$.
>
> $[H_n]_{\text{Math}} = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} d_j e^{i2\pi \frac{nj}{N}}$
>
> The modified version of equation (13.9) is
>
> $C(n) = \frac{N_{\text{zp}}}{N_{\text{orig}}} |[H_n]_{\text{Math}}|^2 \quad \text{for } n = 0, 1, \cdots, \frac{N_{\text{zp}}}{2},$
>
> where $[H_n]_{\text{Math}} =$ **Fourier[data]**, and **data** is a list of d_j values.

By:

> Note: *Mathematica* uses a slightly different definition of H_n to
> that given in equation (13.6), which we designate by $[H_n]_{\text{Math}}$.
>
> $[H_n]_{\text{Math}} = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} d_j e^{i2\pi \frac{(n-1)(j-1)}{N}}$
>
> The modified version of equation (13.9) is
>
> $C(n) = \frac{N_{\text{zp}}}{N_{\text{orig}}} |[H_n]_{\text{Math}}|^2 \quad \text{for } n = 1, 2, \cdots, \frac{N_{\text{zp}}}{2} + 1,$
>
> where $[H_n]_{\text{Math}} =$ **Fourier[data]**, **data** is a list of d_j values,
> and the zero frequency corresponds to the $n = 1$ term.

P. 366, Equation 13.13

Replace:

$$\theta = \frac{1}{2}\tan^{-1}\left[\frac{\sum_{i=1}^{N_R}\sin(4\pi f t_i - \theta)Z(t_i)^2 - \sum_{j=1}^{N_I}\sin(4\pi f t'_j - \theta)Z(t'_j)^2}{\sum_{i=1}^{N_R}\cos(4\pi f t_i - \theta)Z(t_i)^2 - \sum_{j=1}^{N_I}\cos(4\pi f t'_j - \theta)Z(t'_j)^2}\right]. \tag{13.13}$$

By:

$$\theta = \frac{1}{2}\tan^{-1}\left[\frac{\sum_{i=1}^{N_R}\sin(4\pi f t_i)Z(t_i)^2 - \sum_{j=1}^{N_I}\sin(4\pi f t'_j)Z(t'_j)^2}{\sum_{i=1}^{N_R}\cos(4\pi f t_i)Z(t_i)^2 - \sum_{j=1}^{N_I}\cos(4\pi f t'_j)Z(t'_j)^2}\right]. \tag{13.13}$$

P. 367, Section 13.5.1, add an 'of' to the last line.

Replace:

range problems and involves a generalized version of the Lomb-Scargle statistic.

By:

range of problems and involves a generalized version of the Lomb-Scargle statistic.

P. 375, item (a), delete second sentence.

Remove following sentence:

Assume the variance of the data set = 1.

P. 373, change to column headings of Table 13.1.

t (s)	mK	t (s)	mK	t (s)	mK	t (s)	mK
1	0.474	17	-0.865	33	-0.225	49	0.369
2	0.281	18	0.206	34	-1.017	50	0.695
3	1.227	19	-0.926	35	0.817	51	1.291
4	-1.523	20	2.294	36	-2.064	52	0.978
5	-0.831	21	0.786	37	-0.103	53	-0.592
6	-0.978	22	0.522	38	1.878	54	-0.986
7	0.169	23	-1.04	39	0.625	55	-1.005
8	0.04	24	-0.181	40	1.418	56	-1.268
9	0.76	25	-1.47	41	0.464	57	-0.571
10	0.847	26	-1.837	42	-1.182	58	1.128
11	0.106	27	0.523	43	-1.319	59	0.64
12	-1.814	28	0.605	44	1.354	60	0.144
13	-1.16	29	-1.595	45	-1.784	61	-1.468
14	0.249	30	-0.413	46	-0.989	62	-0.71
15	-1.054	31	1.275	47	-1.52	63	-1.486
16	-0.359	32	-1.644	48	1.239	64	-0.129

Table 13.1: The table contains 64 samples of a simulated times series consisting of a single sinusoidal signal with additive IID Gaussian noise.

Chapter 14 errata and revisions

P. 387, two lines above Equ. 14.31, change 'become' to 'becomes'.

Replace:

Δt become infinitesimal

By:

Δt becomes infinitesimal

P. 387, line below Equ. 14.30, replace '$r(t)\Delta t$' by '$r(t_j)\Delta t$'.

P. 388, Section 14.6, add sentence at end of problems 4 and 5.

Add following sentence:

Assume a uniform prior for r_0 in the range 0 to 5 counts s^{-1}, and a Jeffreys prior for τ in the range 2 to 300 s.

Appendix B errata and revisions

P. 409, change equation (B.47) to make it clearer.

Change equation (B.47) from:

$$f_n \equiv \frac{n}{NT}, \qquad n = -\frac{N}{2}, \cdots, \frac{N}{2}, \tag{B.47}$$

To:

$$f_n \equiv \frac{n}{NT}, \qquad n = -\frac{N}{2}, \cdots, -1, 0, 1, \cdots, \frac{N}{2}, \tag{B.47}$$

P. 411, changes to equations B.54 and B.55.

Change:

$$H_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} h_k \; e^{i2\pi nk/N} \tag{B.54}$$

$$h_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} H_n \; e^{-i2\pi nk/N}. \tag{B.55}$$

To:

$$H_n = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} h_k \; e^{i2\pi(n-1)(k-1)/N} \tag{B.54}$$

$$h_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} H_n \; e^{-i2\pi(n-1)(k-1)/N}, \tag{B.55}$$

and the zero frequency corresponds to the $n = 1$ term.

P. 411, Box B.1, remove outer parentheses on the third line.

Change:
(**Fourier**[$\{\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_n}\}$]).
To:
Fourier[$\{\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_n}\}$].

P. 413, second sentence needs to end with a period.

Corrected sentence:
These results are shown in panels (c) and (d) of Figure B.8.

P. 420, 6 lines below equation (B.78)

Change the sentences:
If we only had $|C(n)|^2$, we could estimate $|S(n)|^2$ by extrapolating the spectrum at high values of n to zero. Similarly, we can estimate the $|N(n)|^2$ by extrapolating back into the signal region.
To:
If we only had $|C(n)|^2$, we could estimate $|N(n)|^2$ by extrapolating back into the signal region from high values of n. $|S(n)|^2$ is what stands above the estimate of $|N(n)|^2$ at low values of n. Finally, it is necessary is to extrapolate the portion of $|S(n)|^2$ that sits above the estimated $|N(n)|^2$ to zero.

Appendix D errata and revisions

P. 445, 2nd line of Equ. (D.2), Change 'N' to 'N_{on}'.

P. 446, change first sentence.
Change:
Substituting equations (D.5), (D.4), (D.3) and (D.2) into equation (D.1), we obtain
To:
Substituting equations (D.5), (D.4), (D.3), (D.2) and (14.15) into equation (D.1), we obtain

P. 448, sentence below Equ. (D.16).
Change:
Substitution of equation (D.14) and (D.16) into equation (D.17) yields
To:
Substitution of equations (D.14) and (D.16) into equation (D.6) yields

Added new Appendix F, begins on next page.

Appendix F

Highly Correlated Parameters

In Section 10.5.2, we introduced the correlation coefficient as a useful to summary of the correlation between estimates of any two model parameters. The correlation coefficient is defined by

$$\rho_{\alpha\beta} = \frac{\sigma_{\alpha\beta}}{\sqrt{\sigma_{\alpha\alpha}\sigma_{\beta\beta}}} = \frac{[\mathbf{\Psi}^{-1}]_{\alpha\beta}}{\sqrt{[\mathbf{\Psi}^{-1}]_{\alpha\alpha}[\mathbf{\Psi}^{-1}]_{\beta\beta}}}, \tag{F.1}$$

where $\mathbf{\Psi}^{-1}$ is the variance-covariance matrix of the parameter errors. It ranges from -1 to $+1$, where -1 indicates complete negative correlation, $+1$ indicates complete positive correlation, and 0 indicates no correlation.

In the extreme case of $\rho_{\alpha\beta} = \pm 1$, the elliptical contours will be infinitely wide in one direction (with only information in the prior preventing this catastrophe) and oriented at an angle $\pm \tan^{-1}\left[\sqrt{([\,\boldsymbol{\psi}^{-1}]_{\beta\beta}/[\,\boldsymbol{\psi}^{-1}]_{\alpha\alpha})}\right]$. In this case, the parameter error bars σ_α and σ_β will be infinite as well, saying that our individual estimates of A_α and A_β are completely unreliable, but we can still infer a linear combination of the parameters quite well. For ρ large and positive, the probability contours will be very elongated and bunch up close to the line $A_\beta = \mathrm{b} + mA_\alpha$, where $m = \sqrt{([\,\boldsymbol{\psi}^{-1}]_{\beta\beta}/[\,\boldsymbol{\psi}^{-1}]_{\alpha\alpha})}$. Figure F.1 shows a plot of the inner regions of these elongated contours (68% and 95%) and the equation of the line for a case where $\rho = 0.997$. This line intersects the A_β axis (i.e., $A_\alpha = 0$) at b. We can rewrite this as $A_\beta - mA_\alpha = \mathrm{b}$. Varying the intercept b corresponds to a parallel translation of this line. The value of b that causes the line to fall on the upper 68% contour corresponds to the 68% constraint we can set on the linear combination $A_\beta - mA_\alpha$. Since the contours are closely spaced, this indicates that the data contain a lot of information about the difference $A_\beta - mA_\alpha$. If ρ is large and negative, then we can infer the sum $A_\beta + mA_\alpha$.

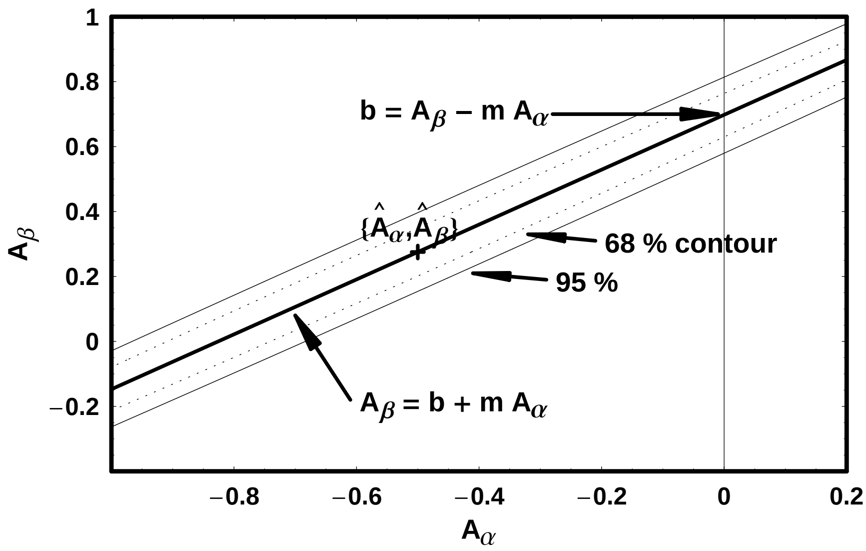


Figure F.1: A contour plot of the joint posterior $p(A_\alpha, A_\beta \mid D, I)$ where the correlation coefficient $\rho = 0.997$. Only the inner region close to the most probable value $\{\hat{A}_\alpha, \hat{A}_\beta\}$ is shown. The probability contours bunch up close to the line $A_\beta = \mathrm{b} + mA_\alpha$. For such a large positive correlation the individual estimates of A_α and A_β are completely unreliable, but we can still infer a linear combination of the parameters $A_\beta - mA_\alpha$ quite well.