Corrections to

Yuri Suhov and Mark Kelbert

Probability and Statistics by Example. Vol. 2: Markov Chains: A Primer in Random Processes and Their Applications

Cambridge University Press, 2008

Many thanks to all readers who pointed at mistakes and discrepancies.

Counting lines down always begins with the top line containing the running (short) title of the chapter and the section. Counting lines up begins with the bottom line of the page. A displayed equation is counted as one line (possibly containing several lines which – when needed – are identified within this equation). Tables and diagrams are also counted as single lines (with a similar principle of identification of lines inside them).

1. Page 3, Line 13 down:
   is represented in Figure 1.1., bottom.
   $\rightarrow$
   is represented in Figure 1.1., bottom. The states are labelled 1, 2, 3, 4.

2. Page 10, Line 3 Line 2 up:
   Again we use three initial conditions, with P^0, P and P^2. For instance,
   $\rightarrow$
   Again we use three initial conditions, with P^0, P and P^2. (As always, 0^0 is taken as 1.) For instance,

3. Page 11, Line 2 down:
   whence $A = 1/3$, $B = 0$, $C = 2/3$ and $p_{11}^{(n)} \equiv 1/3$. Similarly,
   $\rightarrow$
   whence $A = 1/3$, $B = 0$, $C = 2/3$ and $p_{11}^{(n)} \equiv 1/3$, $n \geq 1$. Similarly,

4. Page 15, Line 9 up, the displayed equation:

$$p_{AB}p_{BC}p_{CA} + p_{AC}p_{CB}p_{BA} = \frac{2s_As_Bs_C}{(s_A + s_B)(s_B + s_C)(s_C + s_A)}.$$

$\rightarrow$

$$p_{AA}^{(3)} = p_{AB}p_{BC}p_{CA} + p_{AC}p_{CB}p_{BA} = \frac{2s_As_Bs_C}{(s_A+s_B)(s_B+s_C)(s_C+s_A)}.$$

5. Page 18, Line 13 up, Definition 1.2.2:
   a state cannot escape $\rightarrow$ the chain cannot escape

6. Page 22, Lines 3-4 down:
   renumeration $\rightarrow$ renumbering

7. Page 26, Line 5 up:
   From now on we denote $\rightarrow$ Recall, we denote

8. Page 27, Line 7 down:
   so that if $\mathbb{P}_i(H^A = \infty) > 0$, then $k_i^A = \infty$. In other
   $\rightarrow$
   so that if $\mathbb{P}_i(H^A = \infty) > 0$, then $k_i^A = \infty$. (By convention, $\infty \cdot 0 = 0$). In other

9. Page 22, Line 8:
   Next, we have square blocks of varying size positioned on the main diagonal. These blocks correspond with closed communicating classes C_1, ..., C_m (and we will refer to them as C_1, ..., C_m); they form stochastic submatrices.
   $\rightarrow$
   Next, we have square blocks of varying size positioned on the main diagonal. These blocks correspond with closed communicating classes C_1, ..., C_m (and we will refer to them as C_1, ..., C_m); they form stochastic submatrices.

10. Page 23, Line 20 and below, and the whole of Page 24 should read as follows:
   For simplicity, let us now assume that a finite matrix P is irreducible. It is not hard to guess that, after an appropriate re-enumeration of states, we will have, inside block C, a 'periodic' picture, with a number v of rectangual 'cells' that are cyclically permuted by P: cell 1 is taken to cell 2, and so forth, cell v to cell 1. The horizontal length of the previous cell will be equal to the vertical height of the subsequent one (in the cyclic order). See Fig. 2.6. The space outside cells is again filled with zeros.

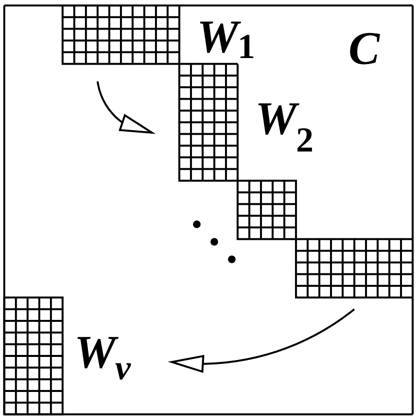
This picture corresponds with a partition of space I into *periodic subclasses* W_1, ..., W_v such that a one-step transition is only possible from a state $j \in W_i$ to a state $k \in W_{i+1}$, $i = 1, \dots, v$. (The sum $i \pm 1$ is understood as a sum modul;o v, so that $W_{v+1} = W_1$ and $M_0 = M_v$.) Recall, according to our assumption, I is reduced to C, a single (and closed) communicating class. The number v is called the *period* of class C. Subclass W_i is identified with the collection of rows in cell i or with the collection of columns in cell $i - 1$. Consequently, if W_i contains n_i states, with $n_1 + \ldots + n_v$ being the total number of states, then cell i is $n_i \times n_{i+1}$.

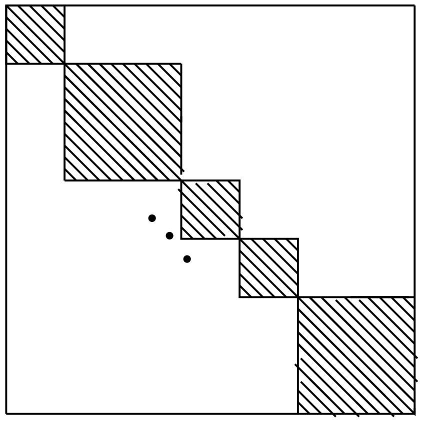
In the majority of our exmples, the period of the closed communicating class equals one, i.e., the class consists of a single subclass, and the matrix contains a single cell. Such classes (as well as their transition matrices) are called *aperiodic*.

In general, if a transition matrix P corresponding to a closed communicating class C of period v is raised to the power v then matrix P^v is divided into diagonal square blocks each of whch forms an $n_i \times n_i$ stochastic matrix of period one. Pictorially speaking, the periodic subclasses W_1, ..., W_v play a rôle of closed communicating classes for matrix P^v.

More precisely, each of the aforementioned $n_i \times n_i$ stochastic submatrices of P^v is irreducible and aperiodic and coincides with the product of rectangular blocks of P taken in the corresponding order. (Consequently, each such submatrix of P^v has the property that if we raise it to a power that is high enough then every entry of the resulting stochastic matrix will be positive.)

A formal definition of a periodic subclass is as follows. We say that state $i \in I$ has period $v = v(i)$ if $p_{ii}^{(n)} > 0$ only when $n = vm$, and v is the largest positive integer with this property. That is, v is the greatest common divisor of the values of n for which $p_{ii}^{(n)} > 0$. In case $p_{ii}^{(n)} \equiv 0$, we set: $v(i) = 0$. An important property is that, within a given closed communicating class C, the period $v(i)$ is the same for all states: it is the period of the class. We put this fact as Worked Example 1.2.7.





P $\qquad\qquad$ $P^{\,v}$

Figure 2.6

Worked Example 1.2.7. Given a state j, define the *period* $v(j)$ of this state as the greatest common divisor of numbers n such that $p_{jj}^{(n)} > 0$. Prove that if states i and j are from the same communicating class then $v(i) = v(j)$. (This justifies the term the 'period of a communicating class'.)

Solution Let i and j be two distinct communicating states. Then $p_{ij}^{(k)} > 0$ for some $k \geq 1$ and $p_{ji}^{(l)} > 0$ for some $l \geq 1$. Assume that $p_{jj}^{(n)} > 0$, then $p_{ii}^{(n+k+l)} \geq p_{ij}^{(k)} p_{jj}^{(n)} p_{ji}^{(l)} > 0$. Therefore, $v(i)$ divides $n + k + l$. Next, $v(i)$ divides $2n + k + l$ as $p_{ii}^{(2n+k+l)} \geq p_{ij}^{(k)} \big(p_{jj}^{(n)}\big)^2 p_{ji}^{(l)} > 0$. Thus, $v(i)$ divides the difference $(2n+k+l)-(n+k+l) = n$. This is true $\forall\, n$ with $p_{jj}^{(n)} > 0$. Then $v(i)$ must divide $v(j)$, as $v(j)$ is the greatest common divisor. A similar argument leads to the conclusion that $v(j)$ divides $v(i)$. Therefore, $v(i) = v(j)$.

Now let $v = v(C)$ be the period of class C. Select a state $i_0 \in C$ and set:

$$W_1 = \{j \in C : \ p^{(n)}_{i_0 j} > 0 \text{ only if } n = vm,$$
$$W_2 = \{j \in C : \ p^{(n)}_{i_0 j} > 0 \text{ only if } n = vm + 1,$$
$$\cdots \quad \cdots \quad \cdots$$
$$W_v = \{j \in C : \ p^{(n)}_{i+0 j} > 0 \text{ only if } n = vm + v - 1,$$

Then, obviously, $C = W_1 \cup \ldots W_v$, and sets W_i are pairwise disjoint. We claim that W_1, ..., W_v are the periodic subclasses. To this end, suppose that state $k \in W_i$ and that $p^{(n)}_{i_0 k} = \sum_{j \in C} p^{(n-1)}_{i_0 j} p_{jk} > 0$. Then, for some j, both probabilities $p^{(n-1)}_{i_0 j}$ and p_{jk} are positive. Consequently, $n - 1 = vm + i'$ for some $i' = i'(j) \in \{0, \ldots, v-1\}$. Such a representation, for a given k, has to be unique. That is, i' is uniquely determined for all j's with the above property. Since we assumed that $p^{(n)}_{i_0 k} > 0$, we obtain that $n = vm + i - 1$. Accordingly, $n - 1 = vm + i - 2$, i.e., $i' = i - 2$ and $j \in W_{i-1}$, for any j as above. Therefore, the inverse is also true: if $j \in W_i$ and $p_{jk} > 0$ then $k \in W_{i+1}$. This establishes the claim under consideration.

11. Page 45, Line 9 down:

take the sequence $i_0, \ldots, i_m$ as above, then $\widehat{p}_{i_l i_{l+1}} > 0$. Now check

$\longrightarrow$

take a sequence of non-repeated states $i = i_0$, i_1, ..., $i_m = j$ with $p_{i_l i_{l+1}} > 0$ (cf. Page 18). Then $\widehat{p}_{i_l i_{l+1}} > 0$. Now check

12. Page 53, Line 11 up:

Let $a, b \geq N$, $a, b, N \in \mathbb{Z}_+$. Consider a birth-death Markov chain on $n = 0, 1, \ldots, N$ with

$$\lambda_n = (N - n)(a - n), \mu_n = n(b - (N - n)).$$

$\longrightarrow$

Let $a, b \geq N$, $a, b, N \in \mathbb{Z}_+$. Consider a birth-death Markov chain on states $n = 0, 1, \ldots, N$ with probabilities p_n and $q_n = 1 - p_n$ of moving to $n + 1$ and to $n - 1$ from n given by $p_n = \lambda_n/(\lambda_n + \mu_n)$, $q_n = \mu_n/(\lambda_n + \mu_n)$, where

$$\lambda_n = (N - n)(a - n), \mu_n = n(b - (N - n)).$$

13. Page 59, Lines 2,3 down:

(I) Irreducible DTMCs with more than one state have transition probabilities $0 < p_{ij} < 1$ $\forall$ states $i, j \in I$ (no absorption).

$\rightarrow$

(I) Irreducible DTMCs with more than one state have transition probabilities $0 < p_{ij}^{(m)} < 1$ $\forall$ states $i, j \in I$ where $m \geq 1$ may depend on i, j (no absorption).

14. Page 60, Line 2 down, Figure 1.13:
$H_l \rightarrow H^{(l)}$ $\quad H_1 \rightarrow H^{(1)}$ $\quad H_2 \rightarrow H^{(2)}$ $\quad H_3 \rightarrow H^{(3)}$

Page 60, Lines 4-6 down:
$H_l(= H_l^i) \rightarrow H(l)(= H_i^{(l)})$ $\quad H_0 \rightarrow H^{(0)}$ $\quad H_1 \rightarrow H^{(1)}$ $\quad H_l \rightarrow H^{(l)}$
$H_{l-1} \rightarrow H^{(l-1)}$

15. Page 66, Line 9 up:
To be in event $\{N = n\}$, a sample
$\rightarrow$
To be in event $\{N = n\}$ with $n \geq 1$, a sample

16. Page 74, Line 3 down:
is geometric. This means that for some $m \geq 1$

$$p_{ij}^{(m)} \geq \rho \;\; \forall \text{ states } i, j. \tag{1.83}$$

$\rightarrow$

is geometric: see Theorem 1.9.3. In fact, if a chain is finite, irreducible and aperiodic, then $\exists\, m \geq 1$ and $\rho \in (0, 1)$ such that

$$p_{ij}^{(m)} \geq \rho \;\; \forall \text{ states } i, j. \tag{1.83}$$

17. Page 77, Line 2 up, Figure 1.15:
$H^{(i)}_{V_i(n)-1} \rightarrow H_i^{(V_i(n)-1)}$ $\quad T_1^{(i)} \rightarrow T_i^{(1)}$ $\quad T_2^{(i)} \rightarrow T_i^{(2)}$ $\quad T_3^{(i)} \rightarrow T_i^{(3)}$
$T^{(i)}_{V_i(n)-1} \rightarrow T_i^{(V_i(n)-1)}$ $\quad T^{(i)}_{V_i(n)} \rightarrow T_i^{(V_i(n))}$

This implies the change of notations in (1.66) as well.

18. Page 86, Lines 10-15 down:
and

$$u_1 + \cdots + u_i = h_0 - h_i, h_i = 1 - A(\gamma_0 + \cdots + \gamma_{i-1}).$$

The condition $\sum\limits_i \gamma_i = \infty$ forces $A = 0$ and hence $h_i = 1$ for all i. Here,

$$\gamma_{2^m-1} = 2^{-m},$$

so $\sum\limits_{\substack{i \\ \longrightarrow}} \gamma_i = \infty$ and the walk is recurrent.

and so

$$u + \cdots + u_i = h_0 - h_i, \quad \text{hence} \quad h_i = 1 - u_1(\gamma_1 + \ldots + \gamma_{i-1}),$$

Here $\gamma_{2^m-1} = 2^{-m}$, and γ_i is constant for blocks of length 2^m, so $\sum_i \gamma_i = +\infty$. This forces $u_1 = 0$, hence $h_i = 1$ for all i, and the walk is recurrent.

19. Page 89, after Line 7 down, the following text should be inserted:

We finish this section by discussing a general picture of asymptotic behaviour of iterations P^n of a finite transition matrix P with several communicating classes. See Figs 2.4 – 2.6. As was said in Section 1.2, the block O_0 tends to 0. The results of the preceding sections show that if a block C_i corresponds to an aperiodic closed communicating class, then in the course of iterations it will converge, as $n \to \infty$, to a block Π_i formed by the repeated row $\pi^{(i)} = (\pi_j^{(i)},\ j \in C_i)$ representing the unique equilibrium distribution supported by class the C_i. Next, if a block O_i is positioned over an aperiodic closed class C_i then it has a limit as $n \to \infty$. And then, the limiting block is a *sub-stochastic* matrix where the sum of the entries along any row is between 0 and 1 but not necessarily 1. (It may be that this sum equals 0, implying that the whole row vanishes). Furthermore, the limiting block is non-zero if the original block was non-zero. (More precisely, $\forall$ state $j \in I$, the sum $\sum\limits_{k \in C_i} p_{jk}^{(n)}$ representing the probability of a transition from j to the closed class C_i in n steps is non-decreasing with n.) Moreover, if the block O_i was zero initially, it will remain zero in matrix P^n as well.

If, on the other hand, C_i is periodic, of some finite period $p_i > 1$, and with periodic sub-classes W_{i1}, ..., W_{ip_i}, then it is convenient to think in terms of the power P^{p_i}. For convenience, assume now that P is irreducible, so there is a unique communicating class C formed by the whole state space I; this class is closed and split into periodic subclasses W_1, ..., W_p. See Fig. 2.6 **??? kotoruyu ispravit'??**. We noted that ordered lists W_i, W_{i+1}, $\ldots$, W_{i-1}, $i = 1, \ldots, v$ (with $W_0 = W_v$) will represent a communicating class for the matrix P^p. Each such collection gives rise to an irreducible and aperiodic stochastic submatrix in P^p which can be considered as a transition matrix in its own right.

20. Page 89, Line 18 down, the displayed equation:
    if the ith object is the first one to be better than anything before,
    $\rightarrow$
    if the ith object is the first one to be better than anything before, $i \geq 2$,
    21. Page 89, Line 2 up:
    In general, $\rightarrow$ In general, for $r \geq 2$,
    Page 89, Line 1 up, the displayed equation:
    X_{r-1}st $\rightarrow$ X_{r-1}th

22. Page 90, Line 1 up:

$$\text{Then } X_1,\ X_2,\ ...,\ X_m = \begin{cases} 2 \\ 3 \\ \vdots \\ m \\ m+1 \end{cases} \qquad (\text{and } X_n \equiv m+1 \text{ for } n > m).$$

$\rightarrow$

Then X_1, X_2, ..., X_m take values $2, 3, \ldots, m+1$ (and $X_n \equiv m+1$ for $n > m$).

23. Page 91, Lines 2-3 down:
    Further,

$$\begin{aligned}\mathbb{P}_1\big(X_2 = j|X_1 = i\big) &= \frac{\mathbb{P}_1\big(X_2 = j, X_1 = i\big)}{\mathbb{P}_1\big(X_1 = i\big)} \\ &= \frac{1}{\mathbb{P}\big(i \text{ the best, } 1 \text{ the 2nd best among } \{1, \ldots, i\}\big)} \\ &\quad \times \mathbb{P}\Big(j \text{ the best, } i \text{ the 2nd best among}\{1, \ldots, j\}; \\ &\qquad\qquad 1 \text{ the 2nd best among}\{1, \ldots, i\}\Big) \\ &= \frac{1/j(j-1) \cdot 1/(i-1)}{1/i(i-1)} = \frac{i}{j(j-1)}, \quad 1 \leq i < j \leq m,\end{aligned}$$

and

$$\mathbb{P}_1\big(X_2 = m+1 | X_1 = i\big) = \frac{\mathbb{P}_1\big(X_2 = m+1, X_1 = i\big)}{\mathbb{P}_1\big(X_1 = i\big)}$$
$$= \frac{\mathbb{P}\big(1 \text{ the 2nd best among } \{1,\ldots,i\};\ i \text{ the absolute best}\big)}{\mathbb{P}\big(i \text{ the best, } 1 \ \text{ 2nd best among } \{1,\ldots,i\}\big)}$$
$$= \frac{1/m \cdot 1/(i-1)}{1/i(i-1)} = \frac{i}{m}, \quad 1 < i \leq m.$$

$\rightarrow$

Further, for $1 \leq i < j \leq m$,

$$\mathbb{P}_1\big(X_2 = j | X_1 = i\big) = \frac{\mathbb{P}_1\big(X_2 = j, X_1 = i\big)}{\mathbb{P}_1\big(X_1 = i\big)}$$
$$= \frac{1}{\mathbb{P}\big(i \text{ the best, } 1 \ \text{ the 2nd best among } \{1,\ldots,i\}\big)}$$
$$\times \mathbb{P}\Big(j \text{ the best, } i \ \text{ the 2nd best among} \{1,\ldots,j\};$$
$$1 \ \text{ the 2nd best among} \{1,\ldots,i\}\Big)$$
$$= \frac{1/j(j-1) \cdot 1/(i-1)}{1/i(i-1)} = \frac{i}{j(j-1)},$$

and for $1 < i \leq m$,

$$\mathbb{P}_1\big(X_2 = m+1 | X_1 = i\big) = \frac{\mathbb{P}_1\big(X_2 = m+1, X_1 = i\big)}{\mathbb{P}_1\big(X_1 = i\big)}$$
$$= \frac{\mathbb{P}\big(1 \text{ the 2nd best among } \{1,\ldots,i\};\ i \text{ the absolute best}\big)}{\mathbb{P}\big(i \text{ the best, } 1 \ \text{ 2nd best among } \{1,\ldots,i\}\big)}$$
$$= \frac{1/m \cdot 1/(i-1)}{1/i(i-1)} = \frac{i}{m}.$$

24. Page 103, Line 13 down:
    For $p = 0$, this is trivial, as $\rightarrow$ For $p = 0$, equality $\overline{\psi_p} = \psi_{\ell - p}$ is trivial, as

25. Page 107, Lines 19-20 down:
    black and white. $\rightarrow$ black or white.

26. Page 108, Line 17 down:

suspensed $\to$ suspended

27. Page 112, Line 6 up:
has a periodic sub-class, $\mathcal{S}_1$, ..., $\mathcal{S}_{k-1}$, of period k. Then, under the matrix P^k, $\forall\ j = 1, \ldots, k-1$, states
$\to$
has periodic sub-classes, $\mathcal{S}_1$, ..., $\mathcal{S}_k$, of period k. Then, under the matrix P^k, $\forall\ j = 1, \ldots, k$, states

28. Page 119, Line 2 down:
See Diaconis, P., Stroock, D. $\to$ We give a formal proof of this inequality in the next section, following the paper Daiconis, P. Stroock, D.

29. Page 120, Line 6 down:
From (1.156) we obtain $\to$ From (1.157) we obtain
Page 120, Line 7 down:
but an incorrect constant. $\to$ although the constant is not as good as the value 8 obtained earlier.

30. Page 121, Line 11 down:
In Example 1.12.2, $\to$ The proof of Cheeger's inequality is also given in Section 1.14. In Example 1.12.2,

31. Page 126, Line 3 up:
atleast $\to$ at least

32. Page 128, Line 8 down:
agrees $\to$ agree

33. Page 168, Solution, Line 5 up (the first line of the solution):
subsequent $\to$ successive

34. Page 169, Lines 1-2 up:
(ii) the long-run proportion of heads which occur within runs of k or more consecutive heads
$\to$
(ii) define a block of order k as a part of a sequence of heads and tails between subsequent appearences of series of at least k consecutive heads, with the agreement that a block finishes with such a series followed by a tail,

after which a new block of order k starts. The series of at least k consecutive heads within a block is called a burst. Find the ratio

(the mean length of the burst)/(the means length of the block).

(For $k = 1$, it gives p.)

35. Page 170, Lines 9-10 down:
Now, (ii) consider subsequent independent 'blocks' of trials formed by k or more consecutive heads followed by a tail. Within a single block, the expected
$\rightarrow$
Now, (ii) within a burst of trials formed by k or more consecutive heads, the expected

36. Page 193, Line 8 up, the displayed equation:

$$\begin{pmatrix} -2 & 1 & 1 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 1 & 0 & 1 & -2 \end{pmatrix}$$

$\rightarrow$

$$Q = \begin{pmatrix} -2 & 1 & 1 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 1 & 0 & 1 & -2 \end{pmatrix}$$

37. Page 193, Lines 2,3 up:
(2.5) with $k = 0, 1, 2, 3$. $\rightarrow$ (2.5) with $n = 0, 1, 2, 3$.

38. Page 193, Line 1 up, the displayed equation, matrix entry $(4, 3)$:
$-6 + 6e^{-3t} \rightarrow 6 - 6e^{-3t}$

39. Page 203, Line 10 up
At the same time
$\rightarrow$
Note that the $Q-$matrix has the form $Q = A - \mathrm{I}$ where A has mainly zeros, but 1s in places (1,2), (2,3) and (3,1). Hence, $e^{tQ} = e^{-t\mathrm{I}}e^{tQ}$. Next, observe that $Q^3 = \mathrm{I}$ and Q^k have zero entries along the main diagonal unless k is divisible by 3. This implies that

40. Page 225, Line 5 up:
concepts → concept

38. Page 230, Line 3 up:
(ii) Owing to
→
(ii) (Cf. the solution to Worked Example 2.2.6.) Owing to

39. Page 236, Line 12 up:
feartures → features

40. Page 237, Line 2 down:
of record values forms an inhomogeneous
→
of record values generates the points of an inhomogeneous

41. Page 278, Line 8 down:
being the number of immature → being the mean number of immature

42. Page 279, Line 1 up:
neigbour → neighbour.

43. Page 280, Line 9 down:
horisontal → horizontal

44. Page 288, Line 16 down, displayed equation (2.142), the first row:
$i + k \to i_k$

45. Page 289, Line 1 up:
Karlin 1968. → Karlin, 1966.

46. Page 320, Lines 15-16 down:
find the probability that both end up busy at a later time t.
→
find a) the probability that both end up free at a later time t, b) the probability that both end up busy at time t.

47. Page 321, between Lines 17 and 18 up, insert:
and similarly,

48. Page 329, Line 1 up, displayed equation (2.178).
$\sum_{i=0}^{i-1} \mu_i z_i - (M_{i-1} + \theta + \lambda_i) z_i + \lambda_i z_{i+1} \rightarrow \sum_{j=0}^{i-1} \mu_j z_j - (M_{j-1} + \theta + \lambda_j) z_j + \lambda_j z_{j+1}$

49. Page 334, Line 6 up:
How they do it. $\rightarrow$ *'How they do it.'*

50. Page 338, Line 1 up:
$\sum_{i=0}^{2} \rightarrow \sum_{i=0,1}$

51. Page 339, Line 1 down:
distributed with parameter 1 $\rightarrow$ distributed with parameter λ

52. Page 339, Line 3 down:
$\alpha = \dfrac{2}{1+\sqrt{5}} \rightarrow \alpha = \dfrac{2\rho}{1+\sqrt{1+4\rho}}$, with $\rho = \dfrac{\lambda}{\mu} < 2$.

53. Page 339, Line 4 up, the displayed equation:
$\sum_{i=0}^{2} \rightarrow \sum_{i=0,1}$

54. Page 378, Line 2 down:
(3.16). $\rightarrow$ (3.14).

55. Page 401, Line 8 up:
a *Liouville* distributions $\rightarrow$ of *Liouville* distributions

56. Page 404, Line 7 down:
Now, integrating in the variable $\rightarrow$ Now, integrating out the variable

57. Page 408, Line 8 down:
Dirichlet distribution Dir $(a, \ldots, a)$, prove that
$\rightarrow$
Dirichlet distribution Dir $(\mathbf{a})$, when $\mathbf{a} = \begin{pmatrix} a \\ \vdots \\ a \end{pmatrix}$, prove that

58. Page 418, 14 up:

between statistics and the information theory. → between statistics and information theory.

59. Page 419, Line 5 down, displayed equation (3.162):
$\sum\limits_{x\in\mathbb{S}} f(x;\theta) \to \sum\limits_{x\in\mathbb{S}} f(x;\theta)^{-1}$ $\qquad \int_{\mathbb{S}} f(x;\theta) \to \int_{\mathbb{S}} f(x;\theta)^{-1}$

60. Page 419, Line 10 up, displayed equation (3.163):
$\sum\limits_{\mathbf{x}\in\mathbb{S}} f(\mathbf{x};\theta) \to \sum\limits_{\mathbf{x}\in\mathbb{S}} f(\mathbf{x};\theta)^{-1}$ $\qquad \int_{\mathbb{S}} f(\mathbf{x};\theta) \to \int_{\mathbb{S}} f(\mathbf{x};\theta)^{-1}$

61. Page 420, Line 5 down:
briefly the divergence. → briefly the divergence. In the case of two-point distributions, on $\mathbb{S} = \{0,1\}$, given by pairs of complementary probabilities $p_0, 1-p_0$ and $p_1, 1-p_1$, the divergence $D(p_1, 1-p_1||p_0, 1-p_0) = p_1 \ln(p_1/p_0) + (1-p_1)\ln[(1-p_1)/(1-p_0)]$.

62. Page 421, Line 3 down:
using $\ln(1+\epsilon) = \epsilon + o(\epsilon)$: → using $\ln(1+\epsilon) = \epsilon - (\epsilon^2/2) + o(\epsilon^2)$:

63. Page 438, Line 2 up:
minimiser → maximiser

64. Page 446, Line 8 up:
An remarkable → A remarkable

65. Page 447, Line 4 up, displayed equation (3.243):
$\sum\limits_{l=1}^{q_i} p_{ij} = 1, \to \sum\limits_{l=1}^{q_i} p_{il} = 1,$

66. Page 456, Line 1 down:
constitue → constitute

67. Page 461, Line 10 down
many British universities had a post such as ‘computor’ attached to mathematics professors
→
some British universities (for instance, Swansea) had a post such as ‘computor’ attached to mathematics professors

68. Page 476, Line 15 down:

as usually $\rightarrow$ as usual

The authors thank John Haigh for pointing at mistakes and suggestions for improvements.