STOCHASTIC MODELING SERIES

Series Editors

Laurence Baxter *State University of New York at Stony Brook*, USA
Marco Scarsini *Universita D'Annuzio*, Italy
Moshe Shaked *University of Arizona*, USA
Shaler Stidham, Jr. *University of North Carolina*, USA

G. Samorodnitsky and M.S. Taqqu
Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance

K. Sigman
Stationary Marked Point Processes: An Intuitive Approach

P. Guttorp
Stochastic Modeling of Scientific Data

M. Neuts
Algorithmic Probability

A. Shwartz and A. Weiss
Large Deviations for Perfomance Analysis: Queues, Communications and Computing

Stochastic Modeling of Scientific Data

PETER GUTTORP
Professor of Statistics
University of Washington
Seattle, USA

CHAPMAN & HALL
London · Glasgow · Weinheim · New York · Tokyo · Melbourne · Madras

Published by Chapman & Hall, 2–6 Boundary Row, London SE1 8HN, UK

Chapman & Hall, 2–6 Boundary Row, London SE1 8HN, UK

Blackie Academic & Professional, Wester Cleddens Road, Bishopbriggs, Glasgow G64 2NZ, UK

Chapman & Hall GmbH, Pappelallee 3, 69469 Weinheim, Germany

Chapman & Hall USA, 115 Fifth Avenue, New York, NY 10003, USA

Chapman & Hall Japan, ITP-Japan, Kyowa Building, 3F, 2-2-1 Hirakawacho, Chiyoda-ku, Tokyo 102, Japan

Chapman & Hall Australia, 102 Dodds Street, South Melbourne, Victoria 3205, Australia

Chapman & Hall India, R. Seshadri, 32 Second Main Road, CIT East, Madras 600 035, India

First edition 1995

Printed in Great Britain by
St Edmundsbury Press Limited, Bury St Edmunds, Suffolk

ISBN 0 412 99281 7

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

∞ Printed on permanent acid-free text paper, manufactured in accordance with ANSI/NISO Z39.48-1992 and ANSI/NISO Z39.48-1984 (Permanence of Paper).

To Kenji

CHAPTER 2

Discrete time Markov chains

Starting from a very simple model of daily precipitation, we build some theory for discrete time Markov chains. We develop some estimation and testing theory. and look at the goodness of fit of this simple model in different ways. A parsimonious model for higher order of dependence is applied to some meteorological data. Harmonic functions are introduced as a tool towards computing how long it takes a random walk on a graph to hit a subset of the boundary states. We analyze some problems in epidemiology and genetics using branching processes. A hidden Markov model categorizing atmospheric variables yields an improved fit to the precipitation data. A similar model is used to describe whether a chemical transmission channel in a nerve cell is open or closed.

2.1. Precipitation at Snoqualmie Falls

The US Weather Service maintains a large number of precipitation monitors throughout the United States. One station is located at the Snoqualmie Falls in the foothills of the Cascade Mountains in western Washington. A day is defined as wet if at least 0.01 inches of precipitation falls during a precipitation day: 8 a.m. through 8 a.m. the following calendar day. To start with, we shall ignore the amounts of rainfall, and just look at the pattern of wet and dry days. Using data from 1948 through 1983, and looking at January rainfall only, there were 325 dry and 791 wet days. Let $X_{ij}=1$(day i of year j wet), where $1(A)$ is 1 if the event A occurs, and 0 otherwise. A very simple model, which we can call the Bernoulli model, is that $X_{ij}\sim\text{Bin}(1,p)$, with the X_{ij} independent, i.e., an iid model, and with p being the probability of rain at Snoqualmie Falls on a January day. The **likelihood** (probability of the observed data as a function of p) is

$$L(p) \propto p^{791}(1-p)^{325}. \tag{2.1}$$

Appendix A contains a brief review of likelihood theory for multinomial data to illustrate some of the central ideas. Edwards (1985) is a good reference for more general likelihood theory. The maximum point of $L(p)$ is the maximum

likelihood estimate (mle) $\hat{p}$ of p. Letting n be the number of days observed, it is easy to see that $\hat{p}=\sum x_{ij}/n=0.709$. A standard error for this estimate is $(p(1-p)/n)^{1/2}$ which we estimate (using $\hat{p}$ in place of p) to be 0.014.

In order to assess the fit of the binomial model for rainfall, we first try to see if the independence assumption seems reasonable. We may suspect a certain amount of persistence, i.e., stretches of like weather, in the data. This would be induced by the relatively slow movement of large weather systems through an area. In the winter, a typical front may take up to three days to pass through from the Pacific Ocean. In order to study this hypothesis, let us look at consecutive pairs of days. Figure 2.1 shows the pattern of rainfall.

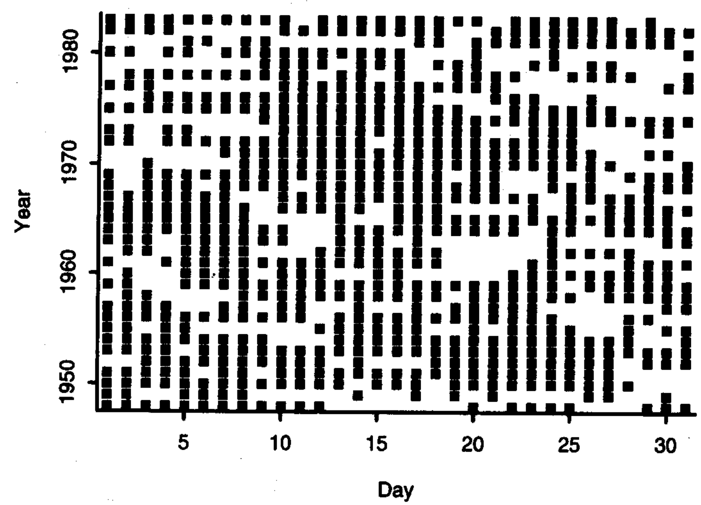

Figure 2.1. The pattern of January precipitation at Snoqualmie Falls. Each square is a day with measurable precipitation. Rows correspond to years, columns to days.

If the independence model is correct, we would expect to see $36\times30\times\hat{p}(1-\hat{p})=223$ dry days following wet days, since we have 36 years of data, and 30 consecutive pairs of days for each January. Table 2.1 contains the total counts, with expected counts under the independence assumption shown in parenthesis. There seems to be a lot more dry days followed by dry days, and wet days followed by wet days, than what the simple iid model predicts. To build a better model of this phenomenon, let us introduce two parameters:

$$p_w = \mathbf{P}(\text{wet today} \mid \text{wet yesterday}) \tag{2.2}$$

$$p_d = \mathbf{P}(\text{wet today} \mid \text{dry yesterday}). \tag{2.3}$$

Table 2.1 Observed precipitation

	Today dry		Today wet		Total
Yesterday dry	186	(91)	123	(223)	309
Yesterday wet	128	(223)	643	(543)	771
Total	314		766		1080

If the X_i are not independent, we must specify the conditional probabilities

$$\mathbf{P}(X_{i+1}=l \mid X_0=k_0, \ldots, X_i=k_i) \tag{2.4}$$

for all i, l, and $k_1, \ldots, k_i$. Note that we will assume unless otherwise specified that the process is observed from time 0. A simple (and perhaps natural) way to specify the probabilities in (2.4) is to assume that the conditional probability only depends on what happened at the previous time point. This assumption was first studied systematically by the Russian probabilist Markov[1] in a sequence of papers, starting in 1907, on generalizing various limit laws to dependent data. Formally we write the **Markov assumption** for a random process (X_n) with discrete state space

$$\begin{aligned}\mathbf{P}(X_{n+1}=l \mid X_0=k_0, \ldots, X_n=k_n)\\ = \mathbf{P}(X_{n+1}=l \mid X_n=k_n) = p_{k_n l}(n).\end{aligned} \tag{2.5}$$

If (X_n) satisfies (2.5) it is called a **Markov chain**. Two seemingly more general forms of (2.5) are outlined in Exercise 1: in part (a) we show that the conditional distribution of the process at any set of future times, given any set of times up to and possibly including the present, only depends on the last of the times in the condition, and in part (b) we show that an equivalent, and rather colorful, way of stating the Markov property is that the future is independent of the past, given the present.

The functions $p_{ij}(n)$ are called **transition probabilities**. We can write the transition probabilities in matrix form. The matrices $\mathbb{P}(n)=(p_{ij}(n))$ are called **transition matrices**.

In order to prove the existence of a Markov chain with a given set of transition matrices and distribution of X_0 one has to verify the Kolmogorov consistency condition (1.22). This is made precise, e.g., in Freedman (1983, pp. 7–8). Here is a simple fact about transition matrices:

[1] Markov, Andrei Andreevich (1856–1922). Russian probabilist in the St Petersburg School. He was a student of Chebyshev, and proved the law of large numbers rigorously in a variety of cases, including dependent sequences.

Proposition 2.1 If $\mathbb{P}$ is a sequence of transition matrices for a Markov chain with state space $S=\{0,\ldots,K\}$, where K may be finite or infinite, then $\sum_{j=0}^{K} p_{ij}(n)=1$ for any n.

Proof We have that $p_{ij}(n)=\mathbf{P}(X_{n+1}=j \mid X_n=i)$, so

$$\begin{aligned}\sum_{j=0}^{K} p_{ij}(n) &= \sum_{j=0}^{K} \mathbf{P}(X_{n+1}=j \mid X_n=i)\\ &= \mathbf{P}\Big(\bigcup_{j=0}^{K}\{X_{n+1}=j\} \mid X_n=i\Big) = \mathbf{P}(X_{n+1}\in S \mid X_n=i) = 1\end{aligned} \tag{2.6}$$

since the process must go somewhere. □

It is often a reasonable simplifying assumption that the transition probabilities are independent of time; such Markov chains are said to have **stationary transition probabilities**. In that case we just need a single transition matrix $\mathbb{P}=\mathbb{P}(1)$. For our rainfall model, we are only considering January. This makes the assumption of stationary transition probabilities reasonable, if we believe (at least approximately) that this month is meteorologically homogeneous. The state space is {dry,wet}, which we can map into $\{0,1\}$. Then, using (2.2) and (2.3), $p_{00}=1-p_d$, $p_{01}=p_d$, $p_{10}=1-p_w$, and $p_{11}=p_w$. In matrix notation,

$$\mathbb{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 1-p_d & p_d \\ 1-p_w & p_w \end{pmatrix}. \tag{2.7}$$

A matrix of non-negative elements with all row sums equal to one is often called a **stochastic** matrix. From now on we will, unless specifically stating otherwise, assume that all transition probabilities are stationary. Here are some elementary properties of stochastic matrices.

Proposition 2.2 (i) A stochastic matrix has at least one eigenvalue equal to one.

(ii) If $\mathbb{P}$ is stochastic, then $\mathbb{P}^k$ is also stochastic for all $k=1,2,3,\ldots$.

Proof (i) is a consequence of the definition of a stochastic matrix, which can be written $\mathbb{P}\mathbf{1}^T=\mathbf{1}^T$, where $\mathbf{1}$ is a row vector of ones (recall that all vectors are assumed to be row vectors). Hence $(\mathbb{I}-\mathbb{P})\mathbf{1}^T=\mathbf{0}$, where $\mathbb{I}$ is the identity matrix, so $\mathbf{1}$ is a right eigenvector of $\mathbb{P}$ corresponding to the eigenvalue 1. Now (ii) follows easily, writing

$$\mathbb{P}^k\mathbf{1}^T = \mathbb{P}^{k-1}\mathbb{P}\mathbf{1}^T = \mathbb{P}^{k-1}\mathbf{1}^T = \cdots = \mathbf{1}^T. \tag{2.8}$$

□

The likelihood for a Markov chain can be written, using (2.5) and (1.13), as

$$L(\mathbb{P}) = \mathbf{P}(X_0=x_0)\prod_{i=0}^{n-1}\mathbf{P}(X_{i+1}=x_{i+1} \mid X_i=x_i)$$
$$= \mathbf{P}(X_0=x_0)\prod_{i=0}^{n-1} p_{x_i x_{i+1}} = \mathbf{P}(X_0=x_0)\prod_{k,l=0}^{K} p_{kl}^{n_{kl}}, \tag{2.9}$$

where n_{kl} is the number of transitions from k to l observed in the chain. In our example we have the additional complication that we are considering 36 years. A simple model is to assume that years are independent. While quasi-periodic large-scale meteorological oscillations such as El Niño may make this hypothesis somewhat suspect (cf. Woolhiser, 1992), it nevertheless allows us to proceed. Furthermore, we shall be able to test it later (Exercise **D1**). Under the assumption of year-to-year independence the likelihood is a product of 36 factors, each of the form (2.9). Clearly, the product collapses, and we can use the data in Table 2.1 to compute

$$L(\mathbb{P}) = L(p_{01},p_{11}) = \left[\prod_{i=1}^{36}\mathbf{P}(X_0^i=x_0^i)\right] p_{00}^{186}p_{01}^{123}p_{10}^{128}p_{11}^{643} \tag{2.10}$$

Assuming that the starting values X_0^i for each year i are fixed (this assumption will be discussed in more detail in section 2.7), so that the beginning term in the right-hand side of (2.10) is 1, we find that L is maximized by

$$\hat{p}_{01} = \frac{123}{309} = 0.398 \quad \hat{p}_{11} = \frac{643}{771} = 0.834 \tag{2.11}$$

so

$$\hat{\mathbb{P}} = \begin{bmatrix} 0.398 & 0.602 \\ 0.166 & 0.834 \end{bmatrix}. \tag{2.12}$$

These estimates are substantially different from the estimate $\hat{p}=0.709$ from the iid model. However, we may question whether such a difference could occur by chance. At a first glance this seems very unlikely, since $\hat{p}_{01}$ is 22 standard errors (of $\hat{p}$) away from $\hat{p}$. For a formal test of significance we use the likelihood ratio test. Recall (or see Appendix A) that under suitable regularity conditions, the **log likelihood ratio** $2(\log L(\hat{\mathbb{P}})-\log L(\hat{p}))$ has a χ^2 distribution with degrees of freedom equal to the difference in the dimension of the parameter spaces; in this case 2-1=1. Although this result was developed for iid processes, it is also true in the Markov chain case. We will return to it in section 2.7. In order to be able to compare the likelihoods we need to exclude the January 1 measurements when computing the iid mle, since those observations cannot be used to compute the Markov chain mle's. This yields $\hat{p}=771/1080=0.714$, slightly higher than the 0.709 we obtained from the full data set. Computing the log likelihood ratio we get

$$2(\log L(\hat{\mathbb{P}})-\log L(\hat{p})) = 2(643\log 0.834+128\log 0.166$$
$$+123\log 0.398+186\log 0.602 \tag{2.13}$$
$$-771\log 0.714-309\log 0.286) = 184.5.$$

which under the null hypothesis of the Bernoulli model is distributed $\chi^2(1)$, corresponding to a P-value of 0. We therefore reject the iid model at all reasonable levels.

2.2. The marginal distribution

Although the Markov assumption tells us how to compute conditional probabilities, one often wants marginal probabilities. It is relatively straightforward to compute these. For example, in a 0-1 chain we have that

$$\mathbf{P}(X_{n+1}=1) = \mathbf{P}(X_{n+1}=1, X_n=0) + \mathbf{P}(X_{n+1}=1, X_n=1)$$
$$= \mathbf{P}(X_n=0)p_{01} + \mathbf{P}(X_n=1)p_{11} \tag{2.14}$$
$$= \mathbf{P}(X_n=1)(p_{11}-p_{01}) + p_{01}.$$

Define the **initial distribution** $\mathbf{p}_0 = (p_0(0),\ldots,p_0(K))$ where $p_0(i) = \mathbf{P}(X_0=i)$. In the 0-1 case we write $p_0(1)\equiv p_1$. Then (2.14) can be written

$$\mathbf{P}(X_1=1) = p_1(p_{11}-p_{01}) + p_{01},$$
$$\mathbf{P}(X_2=1) = (p_{11}-p_{01})\mathbf{P}(X_1=1) + p_{01}$$
$$= p_1(p_{11}-p_{01})^2 + p_{01}(1+(p_{11}-p_{01})) \tag{2.15}$$
$$\cdots$$
$$\mathbf{P}(X_n=1) = (p_{11}-p_{01})^n p_1 + p_{01}\sum_{j=0}^{n-1}(p_{11}-p_{01})^j.$$

If $p_{00}=p_{11}=1$ we have $\mathbf{P}(X_n=1)=p_1$. If $p_{01}\neq p_{11}$ we can write

$$\mathbf{P}(X_n=1) = \frac{p_{01}}{1-(p_{11}-p_{01})} + \left[p_1 - \frac{p_{01}}{1-(p_{11}-p_{01})}\right](p_{11}-p_{01})^n. \tag{2.16}$$

Notice that the effect of the initial distribution p_1 is dampened exponentially, and disappears completely when $p_1=p_{01}/(1-(p_{11}-p_{01}))$. In that situation $\mathbf{P}(X_n=1)$ is the same for each n. This choice of p_1 is called the **stationary initial distribution**. We will return to this in section 2.4.

More generally, let the state space S be an arbitrary countable set, which we identify with the integers $\mathbf{Z}$, and define $p_{jk}^{(n)} = \mathbf{P}(X_n=k \mid X_0=j)$. Here is an important computation, called the **Chapman[1]–Kolmogorov equation,**

[1] Chapman, Sydney (1888–1970). Leading British astro- and geophysicist. Major contributions to the understanding of the aurora; space physics; and convection in the atmosphere.