CHAPTER 10

# Eruptions of the Old Faithful geyser

## 10.1 Introduction

There are many published analyses, from various points of view, of data relating to eruptions of the Old Faithful geyser in the Yellowstone National Park in the USA: for instance Cook and Weisberg (1982, pp. 40–42), Weisberg (1985, pp. 230–235), Silverman (1985; 1986, p. 7), Scott (1992, p. 278), and Aston and Martin (2007). Some of these accounts ignore the strong serial dependence in the behaviour of the geyser; see the comments of Diggle (1993).

In this chapter we present:

- an analysis of a series of long and short eruption durations of the geyser. This series is a dichotomized version of one of the two series provided by Azzalini and Bowman (1990).
- univariate models for the series of durations and waiting times, in their original, non-dichotomized, form; and
- a bivariate model for the durations and waiting times.

The models we describe are mostly HMMs, but in Section 10.2 we also fit Markov chains of first and second order, and compare them with the HMMs.

## 10.2 The binary time series of short and long eruptions

Azzalini and Bowman (1990) have presented a time series analysis of data on eruptions of Old Faithful. The data consist of 299 pairs of observations, collected continuously from 1 August to 15 August 1985. The pairs are $(w_t, d_t)$, with $w_t$ being the time between the starts of successive eruptions, and $d_t$ being the duration of the subsequent eruption.

It is true of both series that most of the observations can be described as either long or short, with very few observations intermediate in length, and with relatively low variation within the low and high groups. It is therefore natural to treat these series as binary time series; Azzalini and Bowman do so by dichotomizing the 'waiting times' $w_t$ at 68 minutes and the durations $d_t$ at 3 minutes, denoting short by 0 and long by 1. There is, in respect of the durations series, the complication that some of the eruptions were recorded only as short, medium or long, and the medium durations have to be treated as either short or long. We use the convention that the (two) mediums are treated as long*.

Table 10.1 *Short and long eruption durations of Old Faithful geyser (299 observations, to be read across rows).*

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 |
| 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | |

It emerges that $\{W_t\}$ and $\{D_t\}$, the dichotomized versions of the series $\{w_t\}$ and $\{d_t\}$, are very similar — almost identical, in fact — and Azzalini and Bowman therefore concentrate on the series $\{D_t\}$ as representing most of the information relevant to the state of the system. Table 10.1 presents the series $\{D_t\}$. On examination of this series one notices that 0 is always followed by 1, and 1 by either 0 or 1. A summary of the data is displayed in the 'observed no.' column of Table 10.3 on p. 144.

### *10.2.1 Markov chain models*

What Azzalini and Bowman first did was to fit a (first-order) Markov chain model. This model seemed quite plausible from a geophysical point of view, but did not match the sample ACF at all well. They then fitted a second-order Markov chain model, which matched the ACF much better, but did not attempt a geophysical interpretation for this second model. We describe what Azzalini and Bowman did, and then fit HMMs and compare them with their models.

Using the estimator of the ACF described by Box, Jenkins and Reinsel (1994, p. 31), Azzalini and Bowman estimated the ACF and PACF of $\{D_t\}$ as displayed in Table 10.2. Since the sample ACF is not even

* Azzalini and Bowman appear to have used one convention when estimating the ACF (medium=long), and the other when estimating the t.p.m. (medium=short). There are only very minor differences between their results and ours.

Table 10.2 *Old Faithful eruptions: sample autocorrelation function ($\hat{\rho}(k)$) and partial autocorrelation function ($\hat{\phi}_{kk}$) of the series $\{D_t\}$ of short and long eruptions.*

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $\hat{\rho}(k)$ | −0.538 | 0.478 | −0.346 | 0.318 | −0.256 | 0.208 | −0.161 | 0.136 |
| $\hat{\phi}_{kk}$ | −0.538 | 0.266 | −0.021 | 0.075 | −0.021 | −0.009 | 0.010 | 0.006 |

approximately of the form $\alpha^k$, a Markov chain is not a satisfactory model; see Section 1.3.4. Azzalini and Bowman therefore fitted a second-order Markov chain, which turned out not to be consistent with a first-order model. They mention also that they fitted a third-order model, which did produce estimates consistent with a second-order model.

An estimate of the transition probability matrix of the first-order Markov chain, based on maximizing the likelihood conditional on the first observation as described in Section 1.3.5, is

$$\begin{pmatrix} 0 & 1 \\ \frac{105}{194} & \frac{89}{194} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0.5412 & 0.4588 \end{pmatrix}. \tag{10.1}$$

Although it is not central to this discussion, it is worth noting that unconditional maximum likelihood estimation is very easy in this case. Because there are no transitions from 0 to 0, the explicit result of Bisgaard and Travis (1991) applies; see our Equation (1.6) on p. 22. The result is that the transition probability matrix is estimated as

$$\begin{pmatrix} 0 & 1 \\ 0.5404 & 0.4596 \end{pmatrix}. \tag{10.2}$$

This serves to confirm as reasonable the expectation that, for a series of length 299, estimation by conditional maximum likelihood differs very little from unconditional.

Since the sequence (0,0) does not occur, the three states needed to express the second-order Markov chain as a first-order Markov chain are, in order: (0,1), (1,0), (1,1). The corresponding t.p.m. is

$$\begin{pmatrix} 0 & \frac{69}{104} & \frac{35}{104} \\ 1 & 0 & 0 \\ 0 & \frac{35}{89} & \frac{54}{89} \end{pmatrix} = \begin{pmatrix} 0 & 0.6635 & 0.3365 \\ 1 & 0 & 0 \\ 0 & 0.3933 & 0.6067 \end{pmatrix}. \tag{10.3}$$

The model (10.3) has stationary distribution $\frac{1}{297}(104, 104, 89)$, and the

Table 10.3 *Old Faithful: observed numbers of short and long eruptions and various transitions, compared with those expected under the two-state HMM.*

| | observed no. | expected no. |
|---|---|---|
| short eruptions (0) | 105 | 105.0 |
| long eruptions (1) | 194 | 194.0 |
| *Transitions:* | | |
| from 0 to 0 | 0 | 0.0 |
| from 0 to 1 | 104 | 104.0 |
| from 1 to 0 | 105 | 104.9 |
| from 1 to 1 | 89 | 89.1 |
| from (0,1) to 0 | 69 | 66.7 |
| from (0,1) to 1 | 35 | 37.3 |
| from (1,0) to 1 | 104 | 104.0 |
| from (1,1) to 0 | 35 | 37.6 |
| from (1,1) to 1 | 54 | 51.4 |

ACF can be computed from

$$\begin{aligned}\rho(k) &= \frac{\mathrm{E}(D_t D_{t+k}) - \mathrm{E}(D_t)\mathrm{E}(D_{t+k})}{\mathrm{Var}(D_t)} \\ &= \frac{297^2\,\Pr(D_t = D_{t+k} = 1) - 193^2}{193 \times 104}.\end{aligned}$$

The resulting figures for $\{\rho(k)\}$ are given in Table 10.4 on p. 145, and match the sample ACF $\{\hat{\rho}(k)\}$ well.

### *10.2.2 Hidden Markov models*

We now discuss the use of HMMs for the series $\{D_t\}$. Bernoulli–HMMs with $m = 1$, 2, 3 and 4 were fitted to this series.

We describe the two-state model in some detail. This model has log-likelihood $-127.31$, $\boldsymbol{\Gamma} = \begin{pmatrix} 0.000 & 1.000 \\ 0.827 & 0.173 \end{pmatrix}$ and state-dependent probabilities of a long eruption given by the vector (0.225, 1.000). That is, there are two (unobserved) states, state 1 always being followed by state 2, and state 2 by state 1 with probability 0.827. In state 1 a long eruption has probability 0.225, in state 2 it has probability 1. A convenient interpretation of this model is that it is a rather special stationary two-state Markov chain, with some noise present in the first state; if the probability 0.225 were instead zero, the model would be exactly a Markov chain. Since (in the usual notation) $\mathbf{P}(1) = \mathrm{diag}(0.225, 1.000)$ and $\mathbf{P}(0) = \mathbf{I}_2 - \mathbf{P}(1)$, a long eruption has unconditional probability

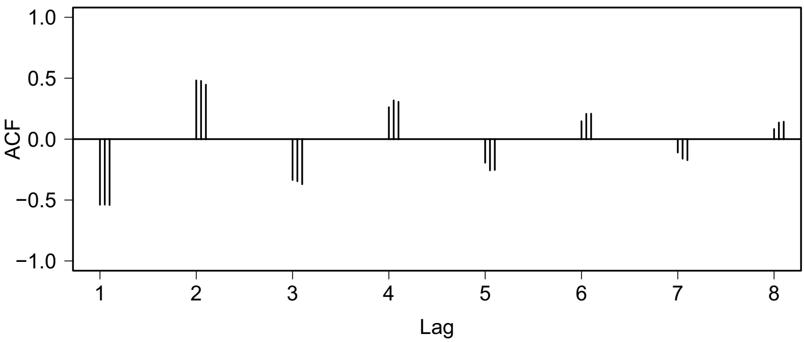


Figure 10.1 *Old Faithful, short and long eruptions: sample autocorrelation function, and ACF of two models. At each lag the centre bar represents the sample ACF, the left bar the ACF of the second-order Markov chain (i.e. model (10.3)), and the right bar that of the two-state HMM.*

Table 10.4 *Old Faithful, short and long eruptions: sample ACF compared with the ACF of the second-order Markov chain and the HMM.*

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $\rho(k)$ for model (10.3) | −0.539 | 0.482 | −0.335 | 0.262 | −0.194 | 0.147 | −0.110 | 0.083 |
| sample ACF, $\hat{\rho}(k)$ | −0.538 | 0.478 | −0.346 | 0.318 | −0.256 | 0.208 | −0.161 | 0.136 |
| $\rho(k)$ for HM model | −0.541 | 0.447 | −0.370 | 0.306 | −0.253 | 0.209 | −0.173 | 0.143 |

$\Pr(X_t = 1) = \boldsymbol{\delta}\mathbf{P}(1)\mathbf{1}' = 0.649$, and a long eruption is followed by a short one with probability $\boldsymbol{\delta}\mathbf{P}(1)\boldsymbol{\Gamma}\mathbf{P}(0)\mathbf{1}'/\boldsymbol{\delta}\mathbf{P}(1)\mathbf{1}' = 0.541$. A short is always followed by a long. A comparison of observed numbers of zeros, ones and transitions, with the numbers expected under this model, is presented in Table 10.3. (A similar comparison in respect of the second-order Markov chain model would not be informative because in that case parameters have been estimated by a method that forces equality of observed and expected numbers of first- and second-order transitions.)

The ACF is given for all $k \in \mathbb{N}$ by $\rho(k) = (1+\alpha)^{-1}w^k$, where $w = -0.827$ and $\alpha = 0.529$. Hence $\rho(k) = 0.654 \times (-0.827)^k$. In Figure 10.1 and Table 10.4 the resulting figures are compared with the sample ACF and with the theoretical ACF of the second-order Markov chain model (10.3). It seems reasonable to conclude that the HMM fits the sample

Table 10.5 *Old Faithful, short and long eruptions: percentiles of bootstrap sample of estimators of parameters of two-state HMM.*

| percentile: | 5th | 25th | median | 75th | 95th |
|---|---|---|---|---|---|
| $\hat{\gamma}_{21}$ | 0.709 | 0.793 | 0.828 | 0.856 | 0.886 |
| $\hat{p}_1$ | 0.139 | 0.191 | 0.218 | 0.244 | 0.273 |

ACF well; not quite as well as the second-order Markov chain model as regards the first three autocorrelations, but better for longer lags.

The parametric bootstrap, with a sample size of 100, was used to estimate the means and covariances of the maximum likelihood estimators of the four parameters $\gamma_{12}$, $\gamma_{21}$, $p_1$ and $p_2$. That is, 100 series of length 299 were generated from the two-state HMM described above, and a model of the same type fitted in the usual way to each of these series. The sample mean vector for the four parameters is (1.000, 0.819, 0.215, 1.000), and the sample covariance matrix is

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.003303 & 0.001540 & 0 \\ 0 & 0.001540 & 0.002065 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The estimated standard deviations of the estimators are therefore (0.000, 0.057, 0.045, 0.000). (The zero standard errors are of course not typical; they are a consequence of the rather special nature of the model from which we are generating the series. Because the model has $\hat{\gamma}_{12} = 1$ and $\hat{p}_2 = 1$, the generated series have the property that a short is always followed by a long, and all the models fitted to the generated series also have $\hat{\gamma}_{12} = 1$ and $\hat{p}_2 = 1$.)

As a further indication of the behaviour of the estimators we present in Table 10.5 selected percentiles of the bootstrap sample of values of $\hat{\gamma}_{21}$ and $\hat{p}_1$. From these bootstrap results it appears that, for this application, the maximum likelihood estimators have fairly small standard deviations and are not markedly asymmetric. It should, however, be borne in mind that the estimate of the distribution of the estimators which is provided by the parametric bootstrap is derived under the assumption that the model fitted is correct.

There is a further class of models that generalizes both the two-state second-order Markov chain and the two-state HMM as described above. This is the class of two-state second-order HMMs, described in Section 8.3. By using the recursion (8.5) for the probability $\nu_t(j,k;\mathbf{x}_t)$, with the appropriate scaling, it is almost as straightforward to compute the likelihood of a second-order model as a first-order one and to fit models by maximum likelihood. In the present example the resulting probabilities of a long eruption are 0.0721 (state 1) and 1.0000 (state 2). The parameter process is a two-state second-order Markov chain with associated first-order Markov chain having transition probability matrix

$$\begin{pmatrix} 1-a & a & 0 & 0 \\ 0 & 0 & 0.7167 & 0.2833 \\ 0 & 1.0000 & 0 & 0 \\ 0 & 0 & 0.4414 & 0.5586 \end{pmatrix}. \qquad (10.4)$$

Here $a$ may be taken to be any real number between 0 and 1, and the four states used for this purpose are, in order: (1,1), (1,2), (2,1), (2,2). The log-likelihood is $-126.9002$. (Clearly the state (1,1) can be disregarded above without loss of information, in which case the first row and first column are deleted from the matrix (10.4).)

It should be noted that the second-order Markov chain used here as the underlying process is the general four-parameter model, not the Pegram–Raftery submodel, which has three parameters. From the comparison which follows it will be seen that an HMM based on a Pegram–Raftery second-order chain is in this case not worth pursuing, because with a total of five parameters it cannot produce a log-likelihood value better than $-126.90$. (The two four-parameter models fitted produce values of $-127.31$ and $-127.12$, which by AIC and BIC would be preferable to a log-likelihood of $-126.90$ for a five-parameter model.)

### *10.2.3 Comparison of models*

We now compare all the models considered so far, on the basis of their unconditional log-likelihoods, denoted by $l$, and AIC and BIC. For instance, in the case of model (10.1), the first-order Markov chain fitted by conditional maximum likelihood, we have

$$l = \log(194/299) + 105\log(105/194) + 89\log(89/194) = -134.2426.$$

The comparable figure for model (10.2) is $-134.2423$; in view of the minute difference we shall here ignore the distinction between estimation by conditional and by unconditional maximum likelihood. For the second-order Markov chain model (10.3) we have

$$\begin{aligned} l = \log(104/297) &+ 35\log(35/104) + 69\log(69/104) \\ &+ 35\log(35/89) + 54\log(54/89) = -127.12. \end{aligned}$$

Table 10.6 presents a comparison of seven types of model, including for completeness the one-state HMM, i.e. the model which assumes independence of the consecutive observations. (Given the strong serial de-

Table 10.6 *Old Faithful, short and long eruptions: comparison of models on the basis of AIC and BIC.*

| model | $k$ | $-l$ | AIC | BIC |
|---|---|---|---|---|
| 1-state HM (i.e. independence) | 1 | 193.80 | 389.60 | 393.31 |
| Markov chain | 2 | 134.24 | 272.48 | 279.88 |
| second-order Markov chain | 4 | 127.12 | **262.24** | **277.04** |
| 2-state HM | 4 | 127.31 | 262.62 | 277.42 |
| 3-state HM | 9 | 126.85 | 271.70 | 305.00 |
| 4-state HM | 16 | 126.59 | 285.18 | 344.39 |
| 2-state second-order HM | 6 | 126.90 | 265.80 | 288.00 |

pendence apparent in the data, it is not surprising that the one-state model is so much inferior to the others considered.)

From the table it emerges that, on the basis of AIC and BIC, only the second-order Markov chain and the two-state (first-order) HMM are worth considering. In the comparison, both of these models are taken to have four parameters, because, although the observations suggest that the sequence (short, short) cannot occur, there is no *a priori* reason to make such a restriction.

While it is true that the second-order Markov chain seems a slightly better model on the basis of the model selection exercise described above, and possibly on the basis of the ACF, both are reasonable models capable of describing the principal features of the data without using an excessive number of parameters. The HMM perhaps has the advantage of relative simplicity, given its nature as a Markov chain with some noise in one of the states. Azzalini and Bowman note that their second-order Markov chain model would require a more sophisticated interpretation than does their first-order model. Either a longer series of observations or a convincing geophysical interpretation for one model rather than the other would be needed to take the discussion further.

### *10.2.4 Forecast distributions*

The ratio of likelihoods, as described in Section 5.2, can be used to provide the forecasts implied by the fitted two-state HMM. As it happens, the last observation in the series, $D_{299}$, is 0, so that under the model $\Pr(D_{300} = 1) = 1$. The conditional distribution of the next $h$ values given the history $\mathbf{D}^{(299)}$, i.e. the joint $h$-step ahead forecast, is easily computed. For $h = 3$ this is given in Table 10.7. The corresponding

Table 10.7 *Old Faithful, short and long eruptions: the probabilities* $\Pr(D_{300} = 1, D_{301} = i, D_{302} = j \mid \mathbf{D}^{(299)})$ *for the two-state HMM (left) and the second-order Markov chain model (right).*

| | $j=0$ | 1 |
|---|---|---|
| $i=0$ | 0.000 | 0.641 |
| 1 | 0.111 | 0.248 |

| | $j=0$ | 1 |
|---|---|---|
| $i=0$ | 0.000 | 0.663 |
| 1 | 0.132 | 0.204 |

probabilities for the second-order Markov chain model are also given in the table.

## 10.3 Univariate normal–HMMs for durations and waiting times

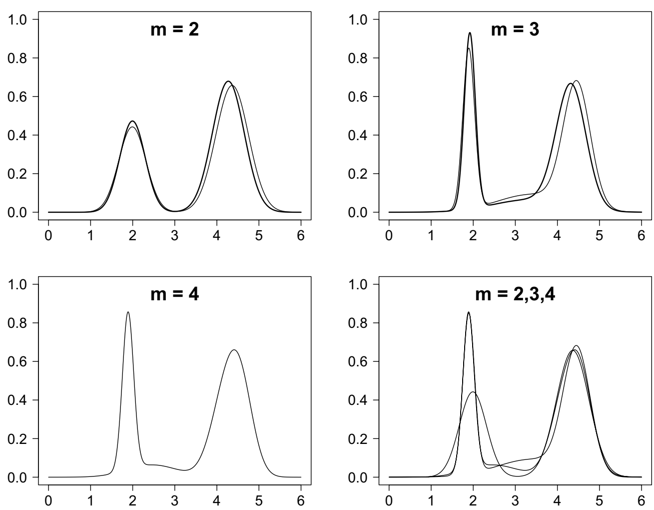


Figure 10.2 *Old Faithful durations, normal–HMMs. Thick lines (m = 2 and 3 only): models based on continuous likelihood. Thin lines (all panels): models based on discrete likelihood.*

Bearing in mind the dictum of van Belle (2002, p. 99) that one should not dichotomize unless absolutely necessary, we now describe normal–HMMs for the durations and waiting-times series in their original, non-dichotomized, form. Note that the quantities denoted $w_t$, both by Azzalini and Bowman (1990) and by us, are the times between the *starts*

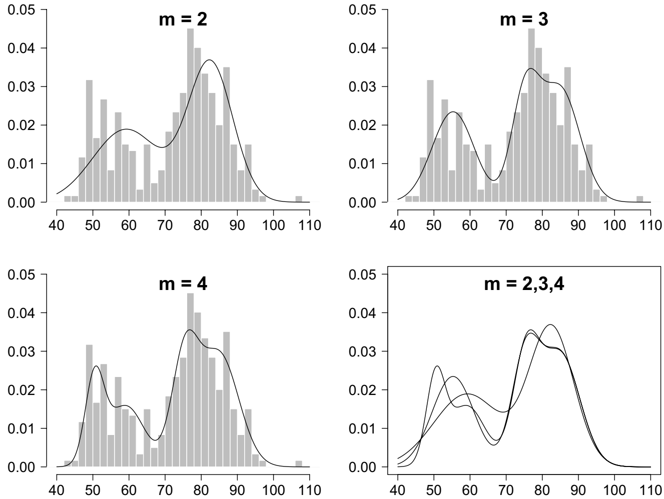


Figure 10.3 *Old Faithful waiting times, normal–HMMs. Models based on continuous likelihood and models based on discrete likelihood are essentially the same. Notice that the model for* $m = 3$ *is identical, or almost identical, to the three-state model of Robert and Titterington (1998): see their Figure 7.*

of successive eruptions, and each $w_t$ therefore consists of an eruption duration plus an interval that lies strictly between eruptions. (Azzalini and Bowman note that $w_t$ therefore exceeds the true waiting time, but ignore this because the eruption durations $d_t$ are small relative to the quantities $w_t$.)

We use here the discrete likelihood, which accepts observations in the form of upper and lower bounds. Section 1.2.4 contains a description of the bounds used for the durations: see p. 13. The waiting times are all given by Azzalini and Bowman to the nearest minute, so (for instance) $w_1 = 80$ means that $w_1 \in (79.5, 80.5)$.

First we present the likelihood and AIC and BIC values of the normal–HMMs. We fitted univariate normal–HMMs with two to four states to durations and waiting times, and compared these on the basis of AIC and BIC. For both durations and waiting times, the four-state model is chosen by AIC and the three-state by BIC. We concentrate now on the three-state models, which are given in Table 10.10. In all cases the model quoted is that based on maximizing the discrete likelihood, and the states have been ordered in increasing order of mean. The transition probability matrix and stationary distribution are, as usual, $\mathbf{\Gamma}$ and $\boldsymbol{\delta}$, and the state-dependent means and standard deviations are $\mu_i$ and $\sigma_i$,

Table 10.8 *Old Faithful durations: comparison of normal–HMMs and independent mixture models by AIC and BIC, all based on discrete likelihood.*

| model | $k$ | $-\log L$ | AIC | BIC |
|---|---|---|---|---|
| 2-state HM | 6 | 1168.955 | 2349.9 | 2372.1 |
| 3-state HM | 12 | 1127.185 | 2278.4 | **2322.8** |
| 4-state HM | 20 | 1109.147 | **2258.3** | 2332.3 |
| indep. mixture (2) | 5 | 1230.920 | 2471.8 | 2490.3 |
| indep. mixture (3) | 8 | 1203.872 | 2423.7 | 2453.3 |
| indep. mixture (4) | 11 | 1203.636 | 2429.3 | 2470.0 |

Table 10.9 *Old Faithful waiting times: comparison of normal–HMMs based on discrete likelihood.*

| model | $k$ | $-\log L$ | AIC | BIC |
|---|---|---|---|---|
| 2-state HM | 6 | 1092.794 | 2197.6 | 2219.8 |
| 3-state HM | 12 | 1051.138 | 2126.3 | **2170.7** |
| 4-state HM | 20 | 1038.600 | **2117.2** | 2191.2 |

Table 10.10 *Old Faithful: three-state (univariate) normal–HMMs, based on discrete likelihood.*

**Durations:**

| $\boldsymbol{\Gamma}$ | | |
|---|---|---|
| 0.000 | 0.000 | 1.000 |
| 0.053 | 0.113 | 0.834 |
| 0.546 | 0.337 | 0.117 |

| $i$ | 1 | 2 | 3 |
|---|---|---|---|
| $\delta_i$ | 0.291 | 0.195 | 0.514 |
| $\mu_i$ | 1.894 | 3.400 | 4.459 |
| $\sigma_i$ | 0.139 | 0.841 | 0.320 |

**Waiting times:**

| $\boldsymbol{\Gamma}$ | | |
|---|---|---|
| 0.000 | 0.000 | 1.000 |
| 0.298 | 0.575 | 0.127 |
| 0.662 | 0.276 | 0.062 |

| $i$ | 1 | 2 | 3 |
|---|---|---|---|
| $\delta_i$ | 0.342 | 0.259 | 0.399 |
| $\mu_i$ | 55.30 | 75.30 | 84.93 |
| $\sigma_i$ | 5.809 | 3.808 | 5.433 |

for $i = 1$ to 3. The marginal densities of the HMMs for the durations are presented in Figure 10.2, and for the waiting times in Figure 10.3.

One feature of these models that is noticeable is that, although the matrices $\mathbf{\Gamma}$ are by no means identical, in both cases the first row is (0,0,1) and the largest element of the third row is $\gamma_{31}$, the probability of transition from state 3 to state 1.

## 10.4 Bivariate normal–HMM for durations and waiting times

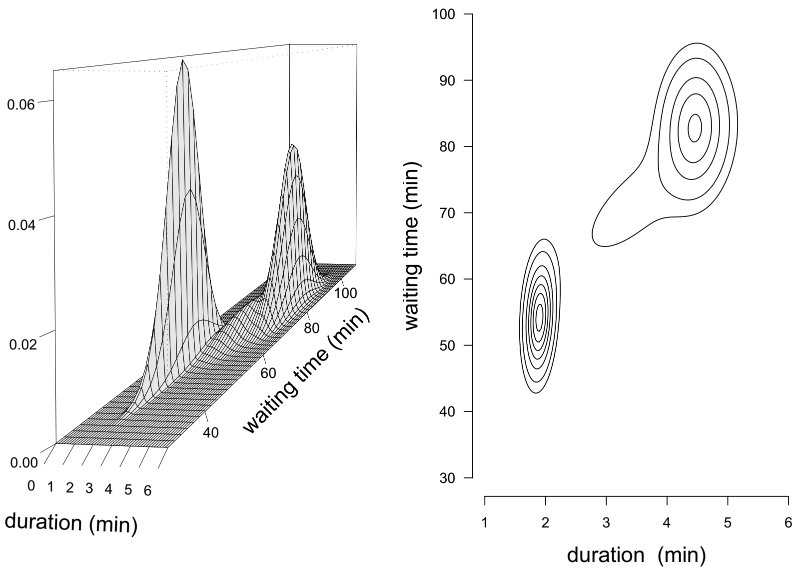


Figure 10.4 *Old Faithful durations and waiting times: perspective and contour plots of the p.d.f. of the bivariate normal–HMM.(Model fitted by discrete likelihood.)*

Finally we give here a stationary three-state bivariate model for durations $d_t$ and waiting times $w_{t+1}$, for $t$ running from 1 to 298. The pairing $(d_t, w_t)$ would be another possibility; in Exercise 2 the reader is invited to fit a similar model to that bivariate series and compare the two models. The three state-dependent distributions are general bivariate normal distributions. Hence there are in all $21 = 15 + 6$ parameters: 5 for each bivariate normal distribution, and 6 for the transition probabilities. The model was fitted by maximizing the discrete likelihood of the bivariate observations $(d_t, w_{t+1})$; that is, by maximizing

$$L_T = \boldsymbol{\delta}\mathbf{P}(d_1, w_2)\boldsymbol{\Gamma}\mathbf{P}(d_2, w_3)\cdots\boldsymbol{\Gamma}\mathbf{P}(d_{T-1}, w_T)\mathbf{1}',$$

where $\mathbf{P}(d_t, w_{t+1})$ is the diagonal matrix with each diagonal element not a bivariate normal density but a probability: the probability that the $t$th pair (duration, waiting time) falls in the rectangle $(d_t^-, d_t^+) \times (w_{t+1}^-, w_{t+1}^+)$. Here $d_t^-$ and $d_t^+$ represent the lower and upper bounds available for the $t$th duration, and similarly for waiting times. The code used can be found in A.3. Figure 10.4 displays perspective and contour plots of the marginal p.d.f. of this model, the parameters of which are given in Table 10.11. Note that here $\boldsymbol{\Gamma}$, the t.p.m., is of the same form as the matrices displayed in Table 10.10.

Table 10.11 *Old Faithful durations and waiting times: three-state bivariate normal–HMM, based on discrete likelihood.*

| $\boldsymbol{\Gamma}$ | | |
|---|---|---|
| 0.000 | 0.000 | 1.000 |
| 0.037 | 0.241 | 0.722 |
| 0.564 | 0.356 | 0.080 |

| $i$ (state) | 1 | 2 | 3 |
|---|---|---|---|
| $\delta_i$ | 0.283 | 0.229 | 0.488 |
| mean duration | 1.898 | 3.507 | 4.460 |
| mean waiting time | 54.10 | 71.59 | 83.18 |
| s.d. duration | 0.142 | 0.916 | 0.322 |
| s.d. waiting time | 4.999 | 8.289 | 6.092 |
| correlation | 0.178 | 0.721 | 0.044 |

## Exercises

1.(a) Use the code in A.3 to fit a bivariate normal–HMM with *two* states to the observations $(d_t, w_{t+1})$.

(b) Compare the resulting marginal distributions for durations and waiting times to those implied by the three-state model reported in Table 10.11.

2. Fit a bivariate normal–HMM with three states to the observations $(d_t, w_t)$, where $t$ runs from 1 to 299. How much does this model differ from that reported in Table 10.11?

3.(a) Write an **R** function to generate observations from a bivariate normal–HMM. (Hint: see the code in A.2.1, and use the package `mvtnorm`.)

(b) Write a function that will find MLEs of the parameters of a bivariate normal–HMM when the observations are assumed to be known exactly; use the ‘continuous likelihood’, i.e. use densities, not probabilities, in the likelihood.

(c) Use the function in 3(a) to generate a series of 1000 observations, and use the function in 3(b) to estimate the parameters.

(d) Apply varying degrees of interval censoring to your generated series and use the code in A.3 to estimate parameters. To what extent are the parameter estimates affected by interval censoring?