Chapter 2

Linear Time Series Models

Auto-regressive time series models are central to modern stationary time series data analysis and, as components of larger models or in suitably modified and generalised forms, underlie non-stationary time-varying models. The concepts and structure of linear auto-regressive models also provide important background material for appreciation of non-linear models. This chapter discusses model forms and inference for AR models, and related topics. This is followed by discussion of the class of stationary auto-regressive, moving average models, one which a large area of traditional linear time series analysis is predicated.

2.1 Structure of Auto-regressions

Consider the time series of equally-spaced quantities x_t, $(t = 1, 2, \ldots)$, arising from the model

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + \epsilon_t \tag{2.1}$$

where ϵ_t is a sequence of uncorrelated error terms and the ϕ_j are constant parameters. This is a sequentially defined model; x_t is generated as a function of past values, parameters and errors. The ϵ_t are termed innovations, and are assumed to be conditionally independent of the past values of the series. They are also often assumed normally distributed, and so then are independent. This is a standard auto-regressive model framework, AR(p) for short; p is the order of the auto-regression. The usual normal model assumes the innovations to be independent $N(\epsilon_t|0, v)$.

AR models may be viewed from a purely empirical standpoint; the data are assumed related over time and the AR form is about the simplest class of empirical models for exploring dependencies. A more formal motivation is, of course, based on the genesis in stationary stochastic process theory. Here we proceed to inference in the model class.

The sequential definition of the model and its Markovian nature imply a sequential structuring of the data density, viz

$$p(x_T, x_{T-1}, \ldots, x_1) = p(x_p, \ldots, x_1) \prod_{t=p+1}^{T} p(x_t|x_{t-1}, \ldots, x_{t-p}) \tag{2.2}$$

for any $T > p$. The leading term is the joint density of the p initial values of the series, as yet undefined. Here the densities are conditional on (ϕ, v);

though this is not made explicit in notation here, it will be later when considering inference about these parameters. If the first p values of the series are known and viewed as fixed constants, and $T = n + p$ for some $n > 1$, then the conditional density of $x = (x_T, x_{T-1}, \ldots, x_{p+1})$ given the first p values is

$$\begin{aligned} p(x_T, x_{T-1}, \ldots, x_{p+1} | x_p, \ldots, x_1) &= \prod_{t=p+1}^{T} p(x_t | x_{t-1}, \ldots, x_{t-p}) \\ &= \prod_{t=p+1}^{T} N(x_t | f_t' \phi, v) = N(x | F\phi, vI_n), \end{aligned} \tag{2.3}$$

where $\phi' = (\phi_1, \ldots, \phi_p)$, $f_t' = (x_{t-1}, \ldots, x_{t-p})$ and F is $n \times p$ matrix with rows f_t'. This has the linear model form and the standard methods apply, as will be reviewed below.

Practically useful extensions of the model (2.1) include models with additional regression terms for the effects of independent regressor variables on the series, differing variances for the ϵ_t over time, and non-normal error distributions. This standard normal linear model is a very special case of auto-regressions which, generally, define models via sequences of conditional distributions for $(x_t | f_t)$ over time.

Stationary AR processes

The series x_t, assumed (at least in principle) to extend over all time $t = 0, \pm 1, \pm 2, \ldots$, follows a stationary auto-regressive model of order p, if the stationarity conditions are satisfied. With the innovations independent $N(\epsilon_t | 0, v)$, the stationary distribution of each x_t, and of any set of $k > 1$ of the x_t, is zero-mean normal. Extending the model to include a non-zero mean μ for each x_t gives $x_t = \mu + (f_t - \mu l)'\phi + \epsilon_t$ where $l = (1, \ldots, 1)'$, or $x_t = \beta + f_t'\phi + \epsilon_t$ where $\beta = (1 - l'\phi)\mu$. The special case of $p = 2$ is discussed in detail in the following section.

In the case $p = 1$, with $x_t = \phi_1 x_{t-1} + \epsilon_t$, the process is stationary for $-1 < \phi_1 < 1$ when the stationary distribution of each of the x_t is $N(x_t | 0, v/(1 - \phi_1^2))$. At the boundary $\phi_1 = 1$ the model becomes a non-stationary random walk. The bivariate stationary distribution of $(x_t, x_{t-1})'$ is normal with correlation $\rho_1 = \phi_1$; that of $(x_t, x_{t-k})'$ for any k is $\rho_k = \phi_1^k$. A positive auto-regressive parameter ϕ_1 leads to processes that wanders away from the stationary mean of the series, with such excursions being more extensive when ϕ_1 is closer to unity; $\phi_1 < 0$ leads to more oscillatory behaviour about the mean.

With $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t$ the process is stationary for parameter values lying in the region $-2 < \phi_1 < 2$, $\phi_1 < 1 - \phi_2$ and $\phi_1 > \phi_2 - 1$. Further discussion appears in the following section.

In the case of general order p, the stationarity condition imposes a set of restrictions on the coefficients ϕ best represented in terms of the roots of the auto-regressive polynomial $\phi(u) = 1 - \sum_{j=1}^{p} \phi_j u^j$, for $|u| \leq 1$. This arises through the representation of (2.1) as

$$\phi(B)x_t = \epsilon_t$$

using the back-shift operator B. The process is stationary if, and only if, the inversion of this equation, namely

$$x_t = \phi(B)^{-1}\epsilon_t = \sum_{j=0}^{\infty} \pi_j \epsilon_{t-j}$$

exists and converges, and this is true only if the roots of the polynomial $\phi(u)$ have moduli greater than unity. Write $\phi(u) = \prod_{j=1}^{p}(1 - \alpha_j u)$ so that the roots are the reciprocals of the α_j. Generally, the α_j may be real-valued or appear as pairs of complex conjugates. Either way, the process is stationary if $|\alpha_j| < 1$ for all j, and non-stationary otherwise.

State-Space Representation

The state-space representation of an AR(p) model has utility in both exploring mathematical structure and, as we shall see later, in inference and data analysis. One version of this representation of (2.1) is simply

$$\begin{aligned} x_t &= F' z_t \\ z_t &= G z_{t-1} + \omega_t \end{aligned} \tag{2.4}$$

where $z_t = (x_t, x_{t-1}, \ldots, x_{t-p+1})'$, the state vector at time t, the innovation at time t appears in the error vector $\omega_t = (\epsilon_t, 0, \ldots, 0)'$, and with

$$F = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \text{and} \quad G = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & & & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & 0 \end{pmatrix}. \tag{2.5}$$

The expected behaviour of the future of the process may be exhibited through the forecast function $f_t(k) = E(x_{t+k}|x_t, \ldots, x_1)$ as a function of integers $k > 0$ for any fixed "origin" $t \geq p$, conditional on the most recent p values of the series in the current state vector $z_t = (x_t, x_{t-1}, \ldots, x_{t-p+1})'$. We have $f_t(k) = F'G^k z_t$. The form is most easily appreciated in cases when the matrix G has distinct eigenvalues, real and/or complex. It easily follows that these eigenvalues are precisely the reciprocals of the roots of the auto-regressive polynomial equation $\phi(u) = 0$, namely the α_j above. Then

$$f_t(k) = \sum_{j=1}^{p} c_{tj} \alpha_j^k \tag{2.6}$$

where the c_{tj} are (possibly complex valued) constants depending on ϕ and the current state z_t, and the α_j are the p distinct eigenvalues/reciprocal roots. The c_{tj} coefficients are given by $c_{tj} = d_j e_{tj}$ where d_j and e_{tj} are the elements of the $p-$vectors $d = E'F$ and $e_t = E^{-1}z_t$ where E is the eigenmatrix of G; i.e. E is the $p \times p$ matrix whose columns are the eigenvectors in order corresponding to the eigenvalues λ_j.

The form of forecast function now depends on the combination of real and complex valued eigenvalues of G. Suppose α_p, for example, is real and positive; the contribution to the forecast function is then $c_{tp}\alpha_p^k$. If the process stationary $|\alpha_j| < 1$ for all j so that this function of k decays exponentially to zero, monotonically if $\alpha_p > 0$, otherwise oscillating between consecutive positive and negative values. If $|\alpha_p| \geq 1$ the process is non-stationary and the forecast function is explosive. The relative contribution to the overall forecast function is measured by the decay rate and the initial amplitude c_{tp}, the latter depending explicitly on the current state, and therefore having different impact at different times t as the state varies in response to the innovations sequence.

In the case of complex eigenvalues, the fact that G is real-valued implies that any complex eigenvalues appear in pairs of complex conjugates. Suppose, for example, that α_1 and α_2 are complex conjugates $\alpha_1 = r\exp(i\omega)$ and $\alpha_2 = r\exp(-i\omega)$ with modulus r and argument ω. In this case, the corresponding complex factors c_{t1} and c_{t2} are conjugate, $a_t\exp(\pm i b_t)$, and the resulting contribution to $f_t(k)$, which must be real-valued, is

$$c_{t1}\alpha_1^k + c_{t2}\alpha_2^k = 2a_t r^k \cos(\omega k + b_t).$$

Hence ω determines the constant frequency of a sinusoidal oscillation in the forecast function, the corresponding wavelength or period being $\lambda = 2\pi/\omega$. In a stationary model $|r| < 1$ so the sinusoidal oscillations over times $k > 0$ are subject to exponentially decay through the damping factor r^k, with additional oscillatory effects if $r < 0$. In non-stationary cases the sinusoidal variation explodes in amplitude as $|r|^k$ increases. The factors a_t and b_t determine the relative amplitude and phase of the component. The amplitude factor $2a_t$ measures the initial magnitude of the contribution of this term to the forecast function, quite separately from the decay factor r. At a future time epoch $s > t$, the new state vector z_s will define an updated forecast function $f_s(k)$ with the same form as (2.6) but with updated coefficients c_{sj} depending on z_s, and so affecting the factors a_s and b_s. Hence, as time evolves, the relative amplitudes and phases of the individual components vary according to the changes in state induced by the sequence of innovations.

Generally, the forecast function (2.6) is a linear combination of exponentially decaying or exploding terms, and decaying or exploding factors multiplying sinusoids of differing periods. Returning to the model (2.1), this basic expected behaviour translates into a process that has the same form but in which, at each time point, the innovation ϵ_t provides a random shock to the current state of the process. This describes a process that exhibits such exponentially damped or exploding behaviour, possibly with periodic components, but in which the amplitudes and phases of the components are randomly varying over time in response to the innovations.

AR(2) *Models*

The special case of $p = 2$ is illuminating and of practical importance in its own right. The process is stationary if $-2 < \phi_1 < 2$, $\phi_1 < 1 - \phi_2$ and $\phi_1 > \phi_2 - 1$. In such cases, the quadratic $\phi(u) = 0$ has reciprocal roots α_i lying within the unit circle, and these define:

- Two real roots when $\phi_1^2 + 4\phi_2 \geq 0$, in which case the the auto-correlations of the process decay exponentially; the forecast function shares this behaviour.
- A pair of complex conjugate roots $r\exp(\pm i\omega)$ when $\phi_1^2 + 4\phi_2 < 0$. The roots have modulus $r = \sqrt{-\phi_2}$ and argument given by $\cos(\omega) = |\phi_1|r/2$. The auto-correlations and forecast function behave as an exponentially damped cosine.

We already know that $-2 < \phi_2 < 2$ for stationarity; for complex roots, we have the additional restriction to $-1 < \phi_2 < -\phi_1^2/4 < 0$. So, in these cases, the model $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t$ represents a quasi-cyclical process, behaving as a damped sine wave of fixed period $2\pi/\omega$ but with amplitude and phase characteristics randomly varying over time in response to the innovations ϵ_t. A large innovations variance v induces greater degrees of variation in this dynamic, quasi-cyclical process. If the innovation variance is very small, or were to become zero at some point, the process would decay to zero in amplitude due to the damping factor. On the boundary of this region at $\phi_2 = -1$, the modulus is $r = 1$ and the forecast function is sinusoidal with no damping; in this case, $\phi_2 = 2\cos(\omega)$. So, for $|\phi_1| < 2$, the model $x_t = \phi_1 x_{t-1} - x_{t-2} + \epsilon_t$ is the one of a sinusoid with randomly varying amplitude and phase; with a small or zero innovation variance v the sinusoidal form sustains, representing essentially a fixed sine wave of constant amplitude and phase. It is easily seen that the difference equation $x_t = 2\cos(\omega)x_{t-1} - x_{t-2}$ defines, for given initial values, a sine wave of period $2\pi/\omega$.

2.2 Basic Inference for AR Models

Return to the basic model (2.1) and the conditional sampling density (2.3), and suppose that the data $y \stackrel{\text{def}}{=} (x_{p+1}, \ldots, x_T)$ are observed. Now make the parameters (ϕ, v) explicit in the notation, so that (2.3) is formally $p(y|\phi, v, z_p)$. Observing y this defines the resulting likelihood function of (ϕ, v). This is a conditional likelihood function, conditional on the assumed initial values $z_p = (x_p, \dots, x_1)'$, so that resulting inferences, reference posterior inferences or otherwise, are also explicitly conditional on these initial values. More on dealing with this later. For now, we have a linear model $p(y|\phi, v, z_p) = N(y|F\phi, vI_n)$ and apply standard theory. In particular, the reference posterior analysis of chapter 1 applies to provide baseline inferences for (ϕ, v).

An Example: EEG data analysis

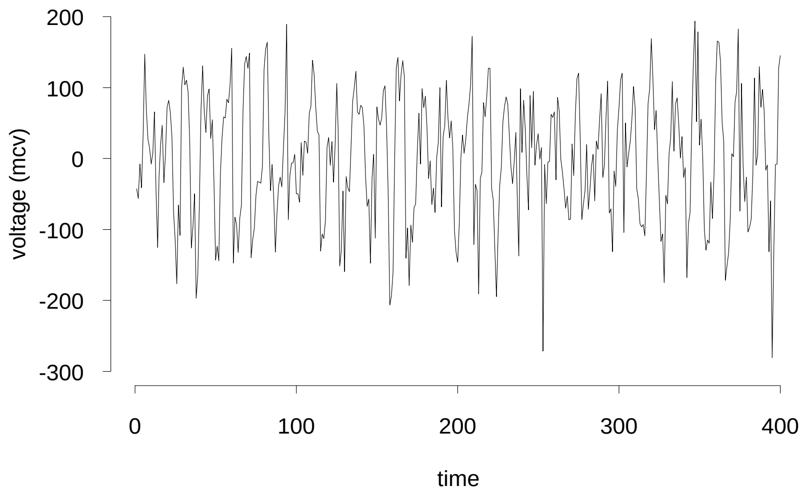

Figure 2.1. A section of an EEG trace

Figure 2.1 displays recordings of an electro-encephalogram (EEG) trace recorded by Dr Andrew Krystal, of Duke University, on a patient in electro-convulsive therapy. This is a very small, sub-sampled segment of a long EEG trace arising in a study of waveform characteristics in multi-channel EEG signal analyses. These studies are germane to assessments of differing ECT protocols. Comparison of two or more such time series underlies part of the study, and appropriate modelling of individual time series represents a starting position for comparative analyses. The data displayed represent variations in scalp potentials in micro-volts during a seizure, the time intervals being just less than one fortieth of a second (the original data is sampled at 256 observations per second, and the 400 points in the figure were obtained by selecting every sixth observation from a mid-seizure section.)

The sample auto-correlations have an apparent damped sinusoial form, indicative of the periodic behaviour evident from the data plot, with a period around 12-14 time units. The samping towards zero evident in the sample auto-correlations is consistent with stationary auto-regressive components with complex roots. The sample partial auto-correlations are evidently strongly negative at lags bewteen 2 and 7 or 8, but appear to drop off there-after, suggesting an auto-regression of order at 7 or 8.

This an AR(8) model is explored as an initial model for these data; $p = 8$ and y represents the final $n = 392$ observations, the first 8 being conditioned upon for initial values. The posterior T distribution has 384 degrees of freedom so is practically indistinguishable from a normal; it has mean

$$b = (0.27, 0.07, -0.13, -0.15, -0.11, -0.15, -0.23, -0.14)'$$

and approximately common standard deviations at 0.05. This illustrates quite typical variation and, to some degree, decay of coefficients with increasing lag. The innovations variance has posterior estimate $s = 61.52^2$.

Fix $\phi = b$ to explore the model based on this point estimate of the parameter vector. The corresponding auto-regressive equation $\phi(u) = 0$ has four pairs of complex conjugate roots; the corresponding moduli and wavelength pairs (r_j, λ_j) are (in order of decreasing modulus)

$$0.97, 12.73; \quad 0.81, 5.10; \quad 0.72, 2.99; \quad 0.66, 2.23.$$

The first term here represents the apparent cyclical pattern of wavelength around $12 - 13$ time units, and has a damping factor close to unity, indicating a rather persistent waveform; the half-life is about $k = 23$, i.e. 0.97^k decays to about 0.5 at $k = 23$, so that, with zero future innovations, the amplitude of this waveform is expected to decay to half a starting level in about two full cycles. By comparison, the three other, higher frequency components have much faster decay rates. The pattern here is quite typical of quasi-cyclical series. The high frequency terms, close to the Nyquist frequency limit, represent terms capturing very short run oscillations in the data of very low magnitude, essentially tailoring the model to low level noise features in the data rather than representing meaningful cyclical components in the series.

At time $t = n = 392$, the current state vector z_t together with the estimated parameter $\phi = b$ implies a forecast function (2.6) in which the four component, damped sinusoids have relative amplitudes $2a_{tj}$ of approximately 157.0, 6.9, 18.0 and 7.0. So the first component of wavelength around 12.73 is quite dominant at this time epoch (as it is over the full span of the data), both in terms of the initial amplitude and in terms of a much lower decay rate. Thus the description of the series as close to a time-varying sine wave is reinforced.

Figure 2.2 displays the data and the forecast function from the end of the series over the next $k = 200$ time epochs based on the estimated value $\phi = b$. Figure 2.3 represents more useful extrapolation, displaying is a single 'sampled future' based on estimated parameter values. This is generated simply by successively simulating future values $y_{T+k} = \sum_{j=1}^{p} b_j y_{T+k-j} + \epsilon_{T+k}$ over $k = 1, 2, \ldots,$ etc., where the ϵ_{T+k} are drawn from $N(\cdot|0, s)$, and substituting sampled values as regressors for the future. This give some flavour of likely development and the form is apparently similar to that of the historical data, suggesting a reasonable model description. These forecasts do not account for uncertainties about the estimated parameters (b, s), so do not represent formal predictive distributions though are quite close approximations. This point is explored further below. Further insight into the nature of the likely development, and also of aspects of model fit, are often gleaned by repeating this exercise, generating and comparing small sets of possible futures.

Simulation of Posterior Distributions

Inferences for other functions of model parameters and formal forecast distributions may be explored via simulation. Suppose interest lies in more formal inference about, for example, the period λ_1 of the dominate cyclical component in the above analysis of the EEG series, and other features of the structure of the roots of the AR polynomial. Though the posterior for (ϕ, v) is analytically manageable, that for the α_i is not; posterior simulation may be used to explore these analytically intractable distributions. Similarly, sampled futures incorporating posterior uncertainties about (ϕ, v) may be easily computed.

To exemplify, 5,000 draws are made from the full normal/inverse-gamma posterior distribution for (ϕ, v). For each such draw, generate a sampled future $y_{T+1}, \ldots, y_{T+k}$, for any horizon k, as before but now based on the simulated values (ϕ, b) at each sample, rather than the estimates (b, s). This delivers a sample of size 5,000 from the full joint posterior predictive distribution for $(y_{T+1}, \ldots, y_{T+k})$. Averaging value across samples provides a Monte Carlo approximation to the forecast function. Exploring sampled futures provides figures like Figure 2.3 though now with additional parameter uncertainties incorporated. In this analysis, the additional uncertainties are small and have slight effects; other applications may be different.

Turn now to inference on the AR polynomial roots α. Each posterior draw (ϕ, v) delivers a corresponding root vector α which represents a random sample from the full posterior $p(\alpha|y, z_p)$. Various features of this posterior sample for α may be summarised. Note first the inherent identification issue, that the roots are unidentifiable as the AR models is unchanged

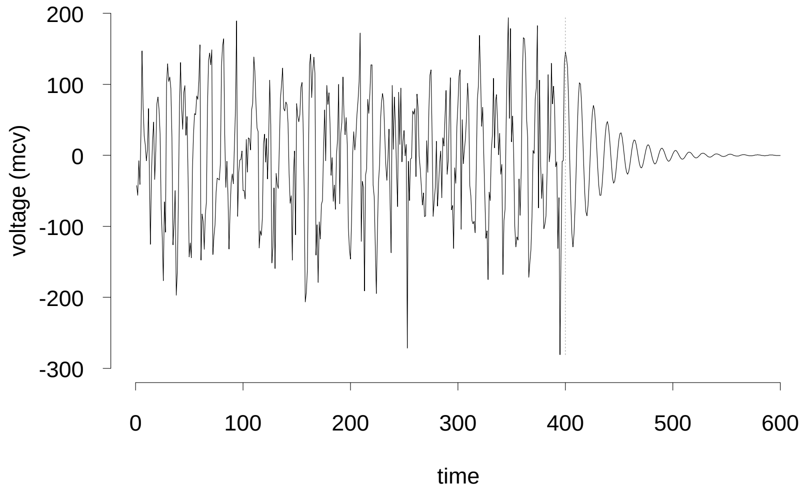

Figure 2.2. EEG trace and forecast function from end of series

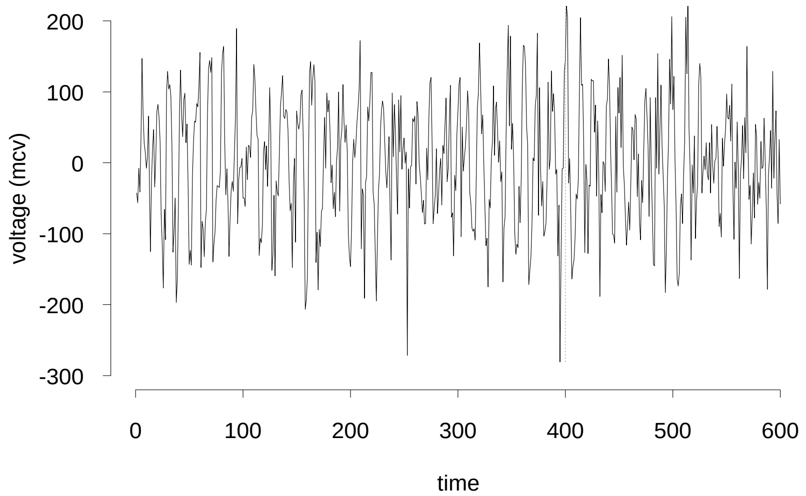

Figure 2.3. EEG trace and sampled future conditional on parameter estimates (b, s)

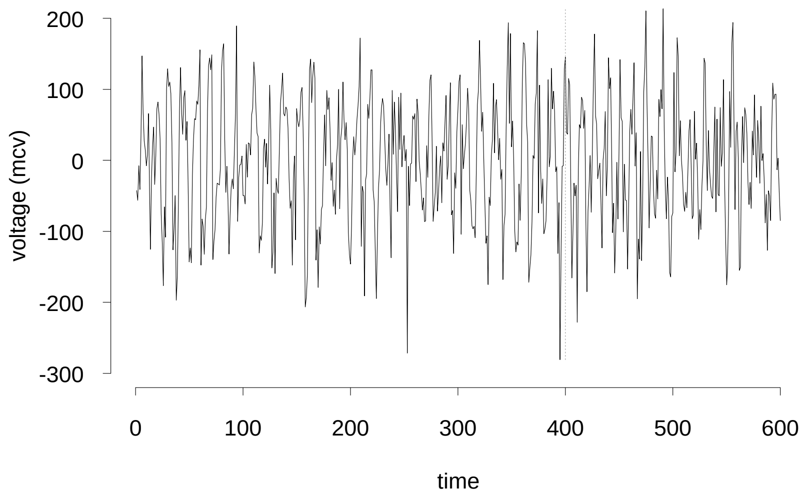

Figure 2.4. EEG trace and sampled future from full posterior predictive distribution

under permutations of the subscripts on the α_i without some modification. One way around this difficulty is to consider inference on roots ordered by moduli or frequency (note the case of real roots formally corresponds to zero frequency). For example, the dominant component of the EEG model has been identified as that corresponding to the complex conjugate roots with the largest period, around $12 - 13$ time units. Ordering the complex values of each sampled set of roots leads to those with the largest period representing a sample from the posterior distribution for the period of the dominant component, and similarly for the samples of the corresponding modulus. Figure 2.5(a) and (b) display the corresponding histograms in this analysis.

Note that no mention of stationarity has been made in this analysis. The reference posterior for ϕ, a multivariate T distribution, is unconstrained and does not theoretically respect a constraint such as stationarity. In some applications, it may be physically meaningful and desirable to impose such an assumption and the analysis should then be modified; theoretically, the prior for (ϕ, v) should be defined as zero outside the stationarity region, whatever the form inside. The simplest approach is to proceed as in the unconstrained analysis, but to simply reject sampled (ϕ, v) values if the ϕ vector lies outside the stationarity region, a condition that is trivially

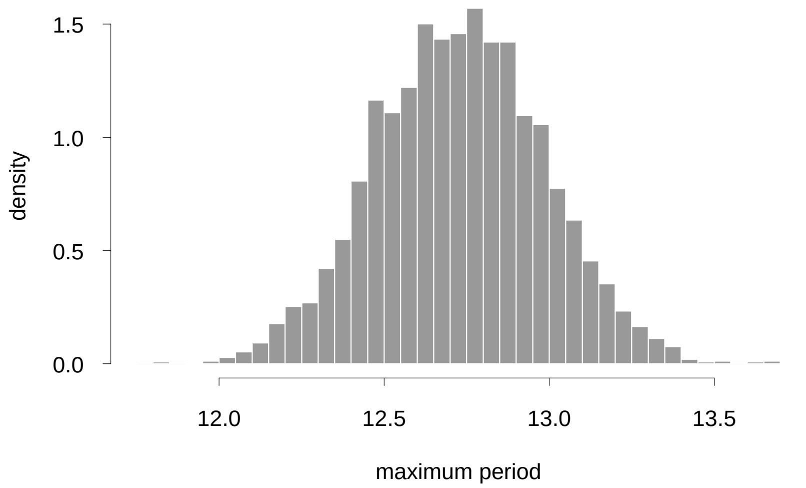

Figure 2.5(a). Posterior for maximum period of sinusoidal components of the EEG series

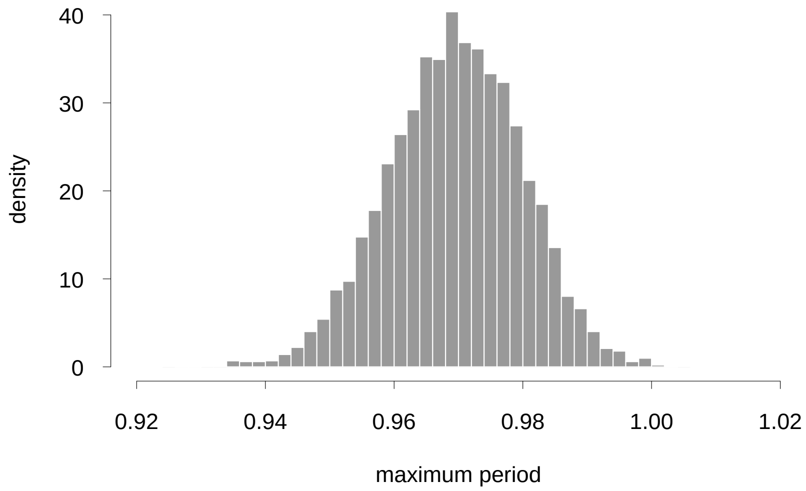

Figure 2.5(b). Posterior for modulus of the damped sinusoidal component of maximum period in the EEG analysis

checked by evaluating the roots of the implied AR polynomial. In cases where the data/model match really supports a stationary series, the rejection rate will be low and this provides a reasonable and efficient approximation to the analysis imposing the stationarity constraint through the prior. In other cases, evidence of non-stationary features may lead to higher rejection rates and an inefficient analysis; other methods are then needed. Some references below indicate work along these lines. Of course, an over-riding consideration is the suitability of a strict stationarity assumption to begin with; if the series, conditional on the appropriateness of the assumed model, is really consistent with stationarity, this should be evidenced automatically in the posterior for the AR parameters, whose mass should be concentrated on values consistent with stationarity. This is, in fact, true in the unconstrained EEG data analysis. Here the estimated AR polynomial root structure (at the reference posterior mean b) has all reciprocal roots with moduli less than unity, so suggesting stationarity. In addition, the 5,000 samples from the posterior can be checked similarly; in fact, the actual sample drawn has no values with roots violating stationarity, indicating high posterior probability (probability one on the Monte Carlo posterior sample) on stationarity. In other applications, sampling the posterior may give some values outside the stationary region; whatever the values, this provides a Monte Carlo approach to evaluating the posterior probability of a stationary series (conditional on the assumed AR model form).

Order Assessment

Analysis may be repeated for different values of model order p, it being useful and traditional to explore variations in inferences and predictions across a range of increasing values. Larger values of p are limited by the sample size, of course, and fitting high order models to only moderate data sets produces meaningless reference posterior inferences; large number of parameters, relative to sample size, can be entertained only with informed and proper prior distributions for those parameters, such as smoothness priors and others mentioned below. Otherwise, increasing p runs into the usual regression problems of over-fitting and collinearity.

Simply proceeding to sequentially increase p and exploring fitted residuals, changes in posterior parameter estimates, and so forth is a very valuable exercise. Various numerical summaries may be easily computed as adjunct to this, the most widely known and used being the so-called Akaike information criterion, or AIC, now described together with a more formal, reference Bayesian measure of model fit. As we are comparing models with differing numbers of parameters, we do so based on a common sample size; thus, fix a maximum order p^* and, when comparing models of various orders $p \leq p^*$, we do so in conditional reference analyses using the latter $n = T - p^*$ of the full T observations in the series.

For a chosen model order p, make explicit the dependence on p by writing b_p for the MLE of the AR parameters, and s_p for the corresponding posterior estimate of innovations variance, i.e. the residual sum of squares divided by $n-p$. For our purposes, the AIC measure of model fit is taken as $2p+n\log(s_p)$. Values of p leading to small AIC values are taken as indicative of relatively good model fits, within the class of AR models so explored (they may, of course, be poor models compared with other classes). Larger values of p will tend to give smaller variance estimates, so decreasing the second term here, but this increase is penalised for parameter dimension by the first term. For the EEG series, negated AIC values, normalised to zero at the maximum, appear in Figure 2.6, based on $p^* = 25$. Also displayed there is a plot of what the corresponding log-likelihood function for model order, computed as follows.

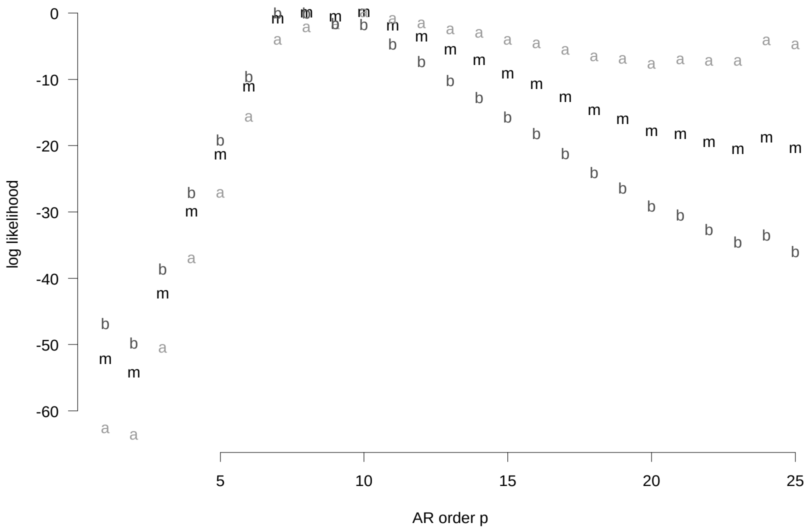

Figure 2.6. Log of log-likelihood function for AR model order, computed from marginal data densities (labelled M), together with negated AIC criterion (labelled A)

In a formal Bayesian analysis, the order p is viewed as an uncertain parameter and so any prior over p is updated via a likelihood function proportional to the marginal data density $p(y|z_p) = \int p(y|\phi, v, z_p)dp(\phi, v)$ where $p(\phi, v)$ is the prior under the AR(p) model and it should be remembered that the dimension of ϕ depends on p. Given proper priors $p(\phi, v)$ across the interesting range of order values $p \leq p^*$, a direct numerical measure of relative fit is available through this collection of marginal densities which defines a valid likelihood function for the model order. To do this, however, requires proper prior $p(\phi, v)$ that naturally depend on the parameter dimension p and this dependency is important in determining the resulting likelihood function. The use of the traditional reference prior invalidates these calculations due to impropriety. Alternative approaches to constructing proper but, in some senses, uninformative priors may be pursued (see later references) but the critical need for priors to be consistent as model dimension varies remains and needs consideration. Nevertheless, under the assumedly common reference prior $p(\phi, v) \propto$ constant, the marginal data densities are defined up to proportionality (and follow directly from the reference Bayesian analysis of the linear regresison model in chapter 1) and are closely related to the AIC values. The reference log-likelihood function so computed for the EEG series, with $p^* = 25$, appears in Figure 2.6. Apparently, both this reference log-likelihood function and the usual AIC criterion suggest orders between 8 and 10 as preferable, hence the earlier analysis based on $p = 8$.

Various alternatives based on different priors give similar results, at least in terms of identifying $p = 8$ or 9 as most appropriate. We note also that formal computation of, for example, predictive inferences involving averaging over p with respect to computed posterior probabilities on model order is possible, in contexts where proper priors for (ϕ, v) are defined across models.

2.3 Initial Values and Missing Data

The above analysis partitions the full data series $x = (x_1, \ldots, x_T)$ into the p initial values z_p and the final $n = T - p$ values y, and is then conditional on z_p. Turn now to the unconditional analysis, in which the full likelihood function for (ϕ, v) is

$$p(x|\phi, v) = p(y|\phi, v, z_p)p(z_p|\phi, v); \tag{2.7}$$

the conditional analysis simply ignores the second component here. Apparently, whether or not this is justifiable or sensible depends on context, as follows.

Analytic Considerations

In some applications, it is appropriate to assume some form of distribution for the initial values z_p that does not, in fact, depend on (ϕ, v) at all. For example, it is perfectly reasonable to specify a model in which, say, the distribution $N(z_p|0, A)$ is assumed, for some specified variance matrix A. In such cases, (2.7) reduces to the first component alone, and the conditional analysis is exact.

Otherwise, there will be a contribution to the likelihood from the initial values, and the conditional analysis is only approximate. Note, however, that, as the series length T increases, the first term of the likelihood, based on $n = T - p$ observations, becomes more and more dominant; the effect of the initial values in the second likelihood factor is fixed based on these values, and does not change with n. On a log-likelihood scale, the first factor behaves in expectation as $o(n)$, and so the conditional and unconditional analyses are asymptotically the same. In real problems with finite n but in which p is usually low compared to n, experience indicates that the agreement is typically close even with rather moderate sample sizes. It is therefore common practice, and completely justifiable in applications with reasonable data sample sizes, to adopt the conditional analysis.

The situation has been much studied under a stationarity assumption, and a variation of the reference Bayesian analysis is explored here to exemplify. Under stationarity, any subset of the x_t have a marginal multivariate normal distribution, with zero mean and a variance matrix whose elements are determined by the model parameters. In particular, the initial values follow $N(z_p|0, vA(\phi))$ where the $p \times p$ matrix $A(\phi)$ depends (only) on ϕ through the defining equations for auto-correlations in AR models. So (2.7), as a function of (ϕ, v), is

$$p(x|\phi, v) \propto v^{-n/2}|A(\phi)|^{-1/2}\exp(Q(x,\phi)/2v)$$

where $Q(x,\phi) = \sum_{t=p+1}^{T}(x_t - f_t'\phi)^2 + z_p'A(\phi)^{-1}z_p$. As developed in Box and Jenkins (1976, chapter 7, especially pp.275-277), this reduces to a quadratic form $Q(x,\phi) = a + 2b'\phi + \phi'C\phi$, where the quantities a, b, C are easily calculable, as follows. Define the the symmetric $(p+1) \times (p+1)$ matrix $D = \{D_{ij}\}$ by elements $D_{ij} = \sum_{r=0}^{n+1-j-i} x_{i+r}x_{j+r}$; then D is partitioned as

$$D = \begin{pmatrix} a & b' \\ b & C \end{pmatrix}.$$

One immediate consequence of this is that, if we ignore the determinant factor $|A(\phi)|$, the likelihood function is of standard linear model form. The traditional reference prior $p(\phi, v) \propto v^{-1}$ induces a normal/inverse gamma posterior, for example; other normal/inverse gamma priors might be used similarly. In the reference case, full details of the posterior analysis can be worked through by the reader; obviously, the posterior mode for ϕ is now $b^* = C^{-1}b$. For the EEG series, the calculations lead to

$$b^* = (0.273, 0.064, -0.128, -0.149, -0.109, -0.149, -0.229, -0.138)'$$

to three decimal places. The approximate value based on the conditional analysis is

$$b = (0.272, 0.068, -0.130, -0.148, -0.108, -0.148, -0.226, -0.136)',$$

earlier quoted to only two decimal places in light of the corresponding posterior standard deviations around 0.05 in each case. The differences, in the third decimal place in each case, are negligible, entirely so in the context of spread of the posterior. Here we are in the (common) context where n is large enough compared to p that the effect of the initial values in (2.7) is really negligible. Repeating the analysis with just the first $T = 100$ EEG observations, the elements of b and b^* differ by only about 0.01, whereas the associated posterior standard errors are around 0.1; the effects become more marked with smaller sample sizes though are still well within the limits of posterior standard deviations with much smaller values of n. In other applications the effects may be more substantial.

Ignoring the determinant factor can be justified by the same, asymptotic reasoning. Another justification is based on the use of an alternative reference prior, that based on Jeffrey's rule. In this case, as shown in Box and Jenkins (1977, §7.43), the Jeffrey's prior is approximately $p(\phi, v) \propto |A(\phi)|^{1/2}v^{-1}$; this results in cancellation of the determinant factor so the above analysis is exact.

Otherwise, under different prior distributions, the exact posterior involves the factor $|A(\phi)|$, a complicated polynomial function of ϕ. However, $|A(\phi)|$ can be evaluated at any specified ϕ value, and numerical methods can be used to analysis the complete posterior. Numerical evaluation of the exact MLE is now a standard feature in some software packages, for example. Bayesian analysis using Monte Carlo methods are easy to implement. For example, the posterior from the conditional analysis, or that above that just ignores the determinant factor, provides a candidate importance sampling distribution, and importance sampling weights are then just based on the determinant factor itself. Alternative simulation methods are now mentioned.

Initial Values Revisited Via Simulation

Introduce the truly uncertain initial values $z_0 = (x_0, x_{-1}, \ldots, x_{-(p-1)})'$. Adjust the earlier conditional analysis to be based on all T observations x, and now to be conditional on these (imaginary) initial values z_0. Then, whatever the prior, we have posterior $p(\phi, v|x, z_0)$. In the reference analysis, we have a normal/inverse gamma posterior now based on all T observations rather than just the last $n = T - p$, with obvious modifications. Note that this posterior can be simulated, to deliver draws for (ϕ, v) conditional on any specific initial vector z_0. This can be embedded in an iterative simulation of the full joint posterior $p(\phi, v, z_0|x)$ if, in addition, we can sample z_0 vectors from the conditional posterior $p(z_0|\phi, v, x)$ for any specified (ϕ, v) parameters.

In the case of a stationary series, stationarity and the linear model form imply reversibility with respect to time; that is, the basic AR model holds backwards, as well as forwards, in time. Hence, conditional on (ϕ, v)

and future series values $x_{t+1}, x_{t+2}, \ldots,$ the current value x_t follows the distribution $N(x_t|g_t'\phi, v)$ where $g_t = rev(z_{t+p}) = (x_{t+1}, \ldots, x_{t+p})'$; here the operator $rev(\cdot)$ simply reverses the elements of its vector argument. Applying this to the initial values at $t = 0, -1, \ldots,$ leads to

$$p(z_0|\phi, v, x) = \prod_{t=0}^{-(p-1)} N(x_t|rev(z_{t+p})'\phi, v).$$

Hence, given (ϕ, v), a vector z_0 is simulated by sequentially sampling the individual component normal distributions in this product: first draw x_0 given the known data z_p and the parameters; the substitute the sampled value x_0 as the first element of the otherwise know data vector z_{p-1}, and draw x_1; continue this way down to $x_{-(p-1)}$. This is technically similar to the process of simulating a future of the series illustrated earlier; now we are simulating the past.

In the modern, computational world of applied statistics, this approach is both trivially implemented and practically satisfying as it provides, modulo the Monet Carlo, exact analysis. Further, extensions of basic AR models to incorporate various practically relevant additional features, naturally lead to Markov chain simulations as natural, and typically necessary, approaches to analysis, so that dealing with the starting value issue in this framework makes good sense.

It should also be clear that the same principle applies to problems of missing data. For any set of indices t such that the values x_t are missing (at random, that is, the reasons for missing data do not have a bearing on the values of the model parameters), then iterative simulation analysis can be extended and modified to incorporate the missing values as additional uncertain quantities to be estimated. Further details can be worked out in the framework here, as with the missing initial values above, and details are left to the reader. We revisit missing values later in the context of general state space models, and state space representations of auto-regessions in paricular.

2.4 Alternative Priors and Further Inference

Additional analyses explore inference based on longer order AR models with various proper priors for the AR coefficients. One interest is in exploring the sensitivity of the earlier, reference inferences under ranges of proper and perhaps more plausible prior assumptions. In each case the model is based on (a maximum lag) $p = 25$, assuming that higher order models would have negligible additional coefficients and that, in any case, the higher order coefficients in the model are likely to decay. The two priors for ϕ are centred around zero and so induce shrinkage of the posteriors towards the prior means of zero for all parameters. In each case, the first p values of z are fixed to provide conditional analyses comparable to that earlier discussed at length.

Analysis Based on Normal Prior

A first analysis assumes a traditional prior with the coefficients i.i.d. normal; the joint prior is $N(\phi|0, wI)$, for some scalar variance w, and so induces shrinkage of the posterior towards the prior mean of zero for all parameters. The hyperparameter w will be estimated together with the primary parameters (ϕ, v), analysis using Gibbs sampling to simulate the full posterior for (ϕ, v, w). We assume prior independence of v and w and adopt uniform priors, so $p(v)$ and $p(w)$ are constant over a wide range; in each analysis we assume this range is large enough so that the corresponding conditional posteriors are effectively proportional to the appropriate conditional likelihood functions, i.e. the truncation implied under the prior has little effect. Posterior simulations draw sequentially from the following three conditional posteriors, easily deduced from the model form and general normal linear model theory reviewed in chapter 1.

(a) Given (v, w), posterior for ϕ is $N(\phi|b, B)$ where $B^{-1} = w^{-1}I + v^{-1}F'F$ and $b = B^{-1}v^{-1}F'y$.

(b) Given (ϕ, w), posterior for v^{-1} is $Ga(v^{-1}|n/2, e'e/2)$ based on residual vector $e = y - F\phi$.

(c) Given (ϕ, v), posterior for w^{-1} is $Ga(w^{-1}|p/2, \phi'\phi/2)$.

For the EEG series, Figure 2.7 graphs the approximate posterior means of the ϕ_j based on a Monte Carlo sample of size 5,000 from the simulation analysis so specified. This sample is saved following burn-in of 500 iterations. Also plotted are the reference posterior means with two posterior standard deviation intervals, for comparison. Some shrinkage of the coefficients is evident, though apparently not dramatic in extent, and the posterior means are not incomparable with the reference values, indicating some robustness to prior specification. Inferences and forecasts based on the normal prior will not differ substantially from those based on the reference prior. In this analysis, the posterior for the shrinkage parameter $\sqrt{w}$ is apparently unimodal, centred around 0.12 with mass predominantly concentrated in the range 0.08-0.16.

Discrete Normal Mixture Prior and Subset Models

A further analysis illustrates priors inducing differential shrinkage effects for across the ϕ_j parameters; some of the ϕ_j may indeed be close to zero, others quite clearly distinct from zero, and a prior view that this may be the case can be embodied in standard modifications of the above analysis. One such approach uses independent priors conditional on individual scale factors, namely $N(\phi_j|0, w/\lambda_j)$, where the weights λ_j are random quantities to be estimated. For example, a model in which only one or two of the

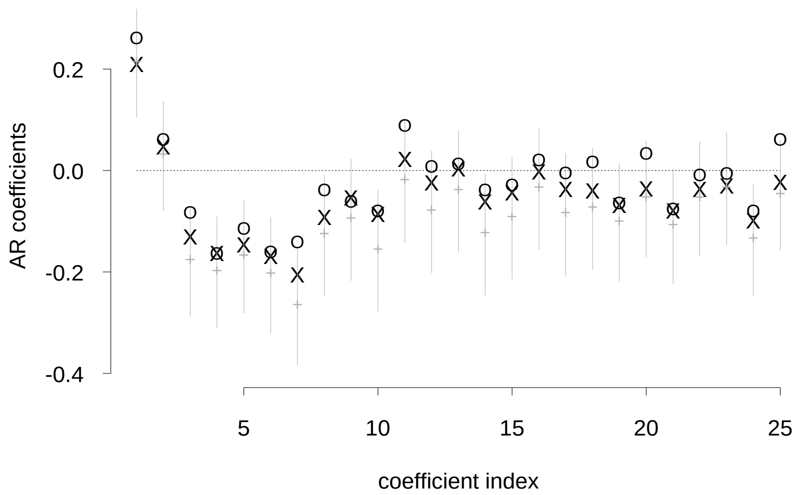

Figure 2.7. Estimates of ϕ in EEG analyses. The vertical bars indicate approximate 95% posterior intervals for the ϕ_j from the reference analysis, centred about reference posterior means. The symbols X indicate approximate posterior means from the analysis based on independent normal priors. Symbols O indicate approximate posterior means from the analysis based on the two-component, normal mixture priors.

ϕ_j are really significant is induced by weights λ_j close to unity for those parameters, the other weights being relatively large resulting in priors and posteriors concentrated around zero for the negligible weights. This links to the concept of subset auto-regressions, in which only a few parameters at specific lags are really relevant, the others, at possibly intervening lags, being zero or close to zero. A class of priors for ϕ that embody this kind of qualitative view provides for automatic inference on relevant subsets of non-negligible parameters and, effectively, addresses the variable selection question.

Probably the simplest approach extends the case of independent normal priors above, in which each $\lambda_j = 1$, to the case of independent priors that are two-component normals, namely

$$\pi N(\phi_j|0, w) + (1 - \pi)N(\phi_j|0, w/L)$$

where π is a probability and L a specified precision factor. If $L >> 1$, the second normal component is very concentrated around zero, so this mixture prior effectively states that each ϕ_j is close to zero, with probability $1-\pi$, and is otherwise drawn from the earlier normal with variance w.

Assume L is specified. Introduce indicators u_j such that $u_j = 1$ or 0 according to whether ϕ_j is drawn from the first or the second of the normal mixture components. These u_j are latent variables that may be introduced to enable the simulation analysis. Write $u = (u_1, \dots, u_p)$ and, for any set of values u, write $\lambda_j = u_j + (1-u_j)L$, so that $\lambda_j = 1$ or L; also, define the matrix $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$. Further, write $k = \sum_{j=1}^p u_j$ for the number of coefficients drawn from the first normal component; k can be viewed as the number of non-neglible coefficients, the others being close to zero. Note that, given π, k has a prior binomial distribution with success probability π.

For completeness and robustness, π is usually viewed as uncertain too; in the analysis below, π is assigned a beta prior, $Be(\pi|a,b)$, independently of the other random quantities in the model. This implies, among other things, a beta-binomial marginal prior for the number k of significant coefficients, namely $p(k) = \binom{n}{k}\beta(a,b)/\beta(a+k, b+p-k)$, over $k = 0, \dots, p$, where $\beta(\cdot,\cdot)$ is the beta function.

Under this model and prior specification, the various conditional posterior distributions to be used in Gibbs sampling of the full posterior for (ϕ, v, w, u, π) are as follows.

(a) Given (v, w, u, π), posterior for ϕ is $N(\phi|b, B)$ where $B^{-1} = w^{-1}\Lambda + v^{-1}F'F$ and $b = B^{-1}v^{-1}F'y$.

(b) Given (ϕ, w, u, π), posterior for v^{-1} is $Ga(v^{-1}|n/2, e'e/2)$ based on residual vector $e = y - F\phi$.

(c) Given (ϕ, v, u, π), posterior for w^{-1} is $Ga(w^{-1}|p/2, q/2)$ with scale factor defined by $q = \sum_{j=1}^p \phi_j^2 \lambda_j$.

(d) Given (ϕ, v, w, π), the u_j are independent with conditional posterior probabilities $\pi_j = Pr(u_i = 0|x, \phi, v, w, \pi)$ given, in odds form, by

$$\frac{\pi_j}{1-\pi_j} = \frac{\pi}{1-\pi}\exp(-(L-1)\phi_j^2/2w)/\sqrt{L}.$$

(e) Given (ϕ, v, w, u), posterior for π is beta, namely $Be(\pi|a+k, b+p-k)$ where $k = \sum_{j=1}^p u_j$.

Iterative sampling of these conditional distributions provides samples of ϕ, v, w, u, π for inference. The additional symbols in Figure 2.7 indicate the posterior means for the ϕ_j from such an analysis, again based on a simulation sample size of 5,000 from the full posterior; the analysis adopts $a = 1, b = 4$ and $L = 25$. We note little difference in posterior means relative to the earlier analyses, again indicating robustness to prior specifications as there is a good deal of data here.

The implied beta-binomial prior for k appears in Figure 2.8, indicating mild support for smaller values consistent with the view that, though there is much prior uncertainty, several or many of the AR coefficients are likely to be negligible. The posterior simulation analysis provides posterior samples of k, and the relative frequencies estimate the posterior distribution, as plotted in Figure 2.8. This indicates a shift to favouring values in the 5–15 ranges based on the data analysis under this specific prior structure; there is much uncertainty about k represented under this posterior, though the indication of a evidence for more than just a few coefficients is strong. Additional information is available in the full posterior sample; it carries, for instance, Monte Carlo estimates of the posterior probabilities that individual coefficients ϕ_j are drawn from the first or second mixture component, simply the approximate posterior means of the corresponding indicators u_j. This information can be used to assess subsets of significant coefficients, as adjunct to exploring posterior estimates and uncertainties about the coefficients, as in Figure 2.7.

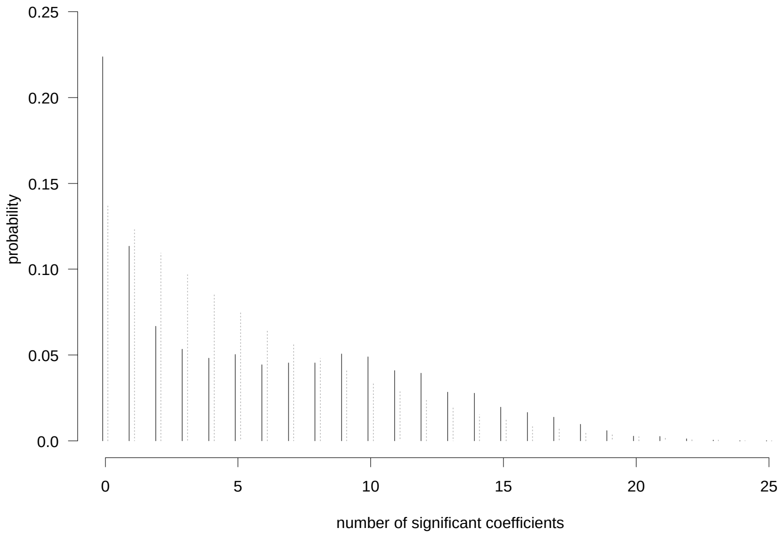

Figure 2.8. Prior and approximate posterior distribution for the number of non-negligible AR coefficients, out of the total $p = 25$, in the EEG analysis under the two-component mixture prior.

Other priors

Analyses based on alternative priors may be similarly explored; some examples are mentioned here, and may be explored by the reader. For instance, the second analysis is an example of a prior constructed via scale-mixtures of a basic normal prior for the individual coefficients. The mixing distribution in that case is discrete, placing mass of π at $\lambda_j = 1$ and $\lambda_j = 25$. Other mixing distributions are common in applied Bayesian work, a key example being the class of gamma distributions. For instance, take the weights λ_j to be independently drawn from a gamma distribution with shape and scale equal to $k/2$ for some $k > 0$; this implies that the resulting marginal prior for each ϕ_j is a Student T distribution with k degrees of freedom, mode at zero and scale factor $\sqrt{w}$. This is , in some senses, a natural heavy-tailed alternative to the normal prior, assigning greater prior probabilities to ϕ_j values further from the prior location at zero. This can result in differential shrinkage, as in the case of the discrete normal mixture in the example.

A further class of priors incorporate the view that AR coefficients are unlikely to be large at higher lags, and ultimately decay towards zero. This kind of qualitative information may be important in contexts where p is large relative to expected sample sizes. This can be incorporated in the earlier normal prior framework, for example, by generalising to independent priors $N(\phi_j|0, w/\lambda_j)$ where the weights are now fixed constants that concentrate the priors around zero for larger lags j; an example would be $\lambda_j = j^2$. Note that this may be combined with additional, random weights to develop decaying effects within a normal mixture prior, and is trivially implemented.

Traditional smoothness priors operate on differences of parameters at successive lags, so that priors for $|\phi_{j+1} - \phi_j|$ are also centred around zero to induce a smooth form of behaviour of ϕ_j as a function of lag j, a traditional 'distributed lag' concept; a smooth form of decay of the effects of lagged values of the series is often naturally anticipated. This is again a useful concept in contexts where long order models are being used. One example of a smoothness prior is given by generalising the normal prior structure as follows. Take the normal margin $N(\phi_1|0, w/\lambda_1)$ and, for $j > 1$, assume conditional priors $N(\phi_j|\phi_{j-1}, w/\lambda_j)$; here the λ_j weights are assumed to increase with lag j to help induce smoothness at higher lags. This specification induces a multivariate normal prior (conditional on the λ_j and w), $p(\phi) = p(\phi_1)\prod_{j=2}^{p} p(\phi_j|\phi_{j-1}) = N(\phi|0, A^{-1}w)$, where the precision matrix $A = B'\Lambda B$ is defined by $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ and

$$B = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}.$$

Again, the λ_j weights may be either specified or random, or a mix of the two. Posterior inferences follow easily using iterative simulation, via straightforward modifications of the analyses above.

2.5 Other Issues

Some very brief comments, and references to later chapters as well as papers etc, on

Sampling Intervals

Continuous Auto-regression

Non-Normal Innovations

Non-Normal Linear Auto-Regressions

2.6 Auto-Regressive, Moving Average Models

Structure of ARMA models

some basic theoretical stuff

Inversion of AR Components

In contexts where data series are of reasonable length, we can fit longer order AR models rather than ARMA or other, more complex forms. One key reason is that the statistical analysis, at least conditional analyses based on assumedly fixed initial values, is much easier. The reference analysis above, for example, is essentially trivial compared with the numerical analysis required to produce samples from posterior distributions in ARMA models (see below). Another driving motivation is that longer order AR models will closely approximate ARMA forms. The proliferation of parameters is an issue, though with longer series and possible use of smoothness priors or other constraints, such as in using subset AR models, this is not an over-riding consideration.

If this view is adopted in a given problem, it may be useful and informative to use the results of an AR analysis to explore possible MA component structure using the device of inversion, or partial inversion, of the AR model. This is described here. Assume an AR(p) model with parameter ϕ, so we can write

$$\phi(B)x_t = \prod_{i=1}^{p}(1-\alpha_i B)x_t = \epsilon_t$$

where the α_i are the auto-regressive characteristic roots. Often there will be subsets of pairs of complex conjugate roots corresponding to quasi-cyclical components, perhaps with several real roots. Stationary components are implied by roots with moduli less than unity.

For some positive integer $r < p$, suppose that the final $p - r$ roots are identified as having moduli less than unity; some or all of the first r roots may also represent stationary components, though that is not necessary for the following development. Then we can re-write the model as

$$\prod_{i=1}^{r}(1 - \alpha_i B)x_t = \prod_{i=r+1}^{p}(1 - \alpha_i B)^{-1}\epsilon_t = \theta(B)\epsilon_t$$

where the (implicitly) infinite order MA component has the coefficients of the infinite order polynomial $\theta(u) = 1 + \sum_{j=1}^{\infty}\theta_j u^j$, defined by

$$1 = \theta(u)\prod_{j=r+1}^{p}(1 - \alpha_i u).$$

So we have the representation

$$x_t = \sum_{j=1}^{r}\phi_j^* x_{t-j} + \epsilon_t - \sum_{j=1}^{\infty}\theta_j\epsilon_{t-j},$$

where the r new AR coefficients are defined by the AR polynomial $\phi^*(u) = \prod_{i=1}^{r}(1 - \alpha_i u)$. The MA terms θ_j can be easily calculated recursively, up to some appropriate upper bound on their number, say q. Explicitly, they are recursively computed as follows:

* initialise by setting $\theta_i = 0$ for $i = 1, 2, \ldots, q$; then,
* for $i = r + 1, \ldots, p$,
  - first update $\theta_1 = \theta_1 - \alpha_i$, and then,
    for $j = 2, \ldots, q$, update $\theta_j = \theta_j + \alpha_i\theta_{j-1}$.

Suppose ϕ is set at some estimate, such as a posterior mean, in the AR(p) model analysis. The above calculations can be performed for any specified value of r to compute the corresponding MA coefficients in an inversion to the approximating ARMA(r, q) model. If the posterior for ϕ is sampled in the AR analysis, the above computations can be performed repeatedly for all sampled ϕ vectors, so producing corresponding samples of the ARMA parameters ϕ^* and θ. Thus inference in various relevant ARMA models can be directly, and quite easily, deduced by inversion of longer order AR models. Typically, various values of r will be exlored. Guidance is derived from the estimated amplitudes and, in the case of complex roots, periods of the roots of the AR model. Analyses indicating some components that are persistent, i.e. that have moduli close to unity and, in the cases of complex roots, longer periods, suggest that these components be retained in the AR description. The remaining roots, corresponding to high frequency characteristics in the data with lower moduli and, if complex, high frequency oscillations, are then the candidates for inversion to what will often be a low order MA component. The calculations can be repeated, sequentially increasing q and exploring inferences about the MA parameters, to assess a relevant approximating order. These kinds of analyses will be illustrated later, in chapter 6.

Likelihood and Bayesian Inference

ref some of the recent MCMC stuff, also likelihood, initial values, etc etc

Some Data Analyses

including some stuff on long-order AR models as a way to identify and fit ARMA models, use EEG data

2.7 References

The classic reference for auto-regressive models is, of course, Box and Jenkins (1976). Additional material on Bayesian inference in auto-regressive, and related time series models appears in ...

Many others – add some refs

Chib and Greenberg, Gelfand and Marriott, McCulloch and Tsay all on Gibbs in AR models –talk about some specific foci here in refs

Tong's book, Müller, West and MacEachern

West etc etc on mixture priors

smoothness priors, distributed lags, others

something on convergence and diagnostics ?

Akaike and model choice stuff

Rubin et al new book on Bayes

Gibbs sampling and others

Implicit Bayes factors for AR models – Berger and student