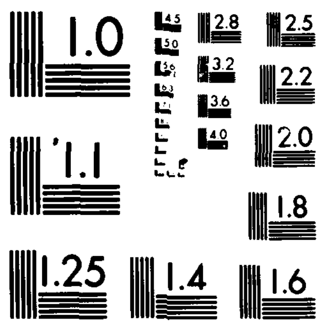
AD-A114 622 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
A BAYESIAN APPROACH TO MARKOVIAN MODELS FOR NORMAL AND POISSON --ETC(U)
FEB 82 T LEONARD DAAG29-80-C-0041
UNCLASSIFIED MRC-TSR-2337 NL

END
DATE
FILMED
6-82
DTIC



MICROCOPY RESOLUTION TEST CHART

(2)

MRC Technical Summary Report #2337

A BAYESIAN APPROACH TO MARKOVIAN MODELS FOR NORMAL AND POISSON DATA

Tom Leonard

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

February 1982

(Received January 11, 1982)

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

Approved for public release
Distribution unlimited

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

A BAYESIAN APPROACH TO MARKOVIAN MODELS FOR NORMAL AND POISSON DATA

Tom Leonard

Technical Summary Report #2337

February 1982

ABSTRACT

A Bayesian updating procedure is proposed for filtering the process parameters in the two-stage Markovian constant variance model for time varying normal data in the situation where the signal to noise ratio is unknown. A forecasting procedure is described which yields the entire predictive distribution of future observations; a numerical study involves an on-line analysis for chemical process concentration readings. A similar method is developed for Poisson data and applied to the analysis of an industrial control chart.

AMS (MOS) Subject Classifications: 62M05; 62M20

Key Words: Kalman filter, ARIMA models; Linear prediction; Variance components; Bayesian updating; Predictive distribution; Quality control; Poisson observations; Linear growth model; Box-Jenkins.

Work Unit Number 4 - Statistics and Probability

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

Simple two-stage Bayesian models are considered for time-varying normal and Poisson data, in the linear prediction situation. The normal model is equivalent to an ARIMA process and posterior estimates for the process levels and signal to noise ratio are compared with the Box-Jenkins likelihood procedure. The posterior distribution of the variance ratio may be updated on-line and the unconditional posterior densities and predictive distributions, of the process levels and future observations, may be calculated at each time stage, providing useful inference procedures for filtering and forecasting. The methodology is applied in some detail to the Box-Jenkins chemical process data. In the Poisson situations similar procedures are developed using a Gamma approximation to one of the updating distributions in the model. This method is illustrated by an analysis of an industrial control chart.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

A BAYESIAN APPROACH TO MARKOVIAN MODELS
FOR NORMAL AND POISSON DATA

Tom Leonard

1. Introduction

Following Kalman [5], Blight [1], and Harrison and Stevens [4], we consider the two-stage Markovian (constant variance) model

$$y_i = \theta_i + \delta_i \tag{1.1}$$

$$\theta_i = \theta_{i-1} + \varepsilon_i \tag{1.2}$$

$$(i = 1,2,\ldots,m)$$

where the θ_i represent the process parameters and the δ_i and ε_i are mutually independent and normally distributed error terms with zero means. The δ_i and ε_i are taken to possess respective common variances τ^2 and σ^2, with the exception of ε, which for convenience is taken to possess variance $\sigma_0^2 = \gamma\tau^2$ where γ is known. The initial mean θ_1 then possesses a normal prior distribution with mean θ_0 and variance σ_0^2.

For $i = 2,\ldots,m$ the model in (1.1) and (1.2) is mathematically equivalent to an ARIMA process, of the type discussed by Box and Jenkins ([2], p. 8) where

$$y_i - y_{i-1} = q_i - \xi q_{i-1} \tag{1.3}$$

with the q_i representing independent and normally distributed random variables with zero means and common variance $\xi^{-1}\tau^2$, and with ξ denoting the smaller root of the equation

$$2 + \alpha = \xi^{-1} + \xi \quad , \tag{1.4}$$

with

$$\alpha = \sigma^2/\tau^2 \quad .$$

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

The equivalence may be demonstrated by differencing out the θ_i from (1.1) and (1.2). In Chapter 7, Box and Jenkins recommend a procedure for the approximate maximum likelihood estimation of ξ. We will later compare this with our inferences based upon the posterior distribution for α.

From [4], we have that, when τ^2 and σ^2 are known, the posterior distribution after time m of the parameter θ_m is normal with mean a_m and variance v_m. The latter may be obtained from the updating relations

$$a_i = a_{i-1} + D_i(y_i - a_{i-1}) \qquad (i = 1,\ldots,m) \tag{1.5}$$

and

$$D_i = (D_{i-1} + \alpha)/(D_{i-1} + \alpha + 1) \qquad (i = 2,\ldots,m) \tag{1.6}$$

with $v_i = \tau^2 D_i$, $\alpha = \sigma^2/\tau^2$, $a_0 = \theta_0$, and $D_1 = \gamma/(1+\gamma)$.

Note that a_m provides a smoothed value after time m for the process parameter θ_m, it is also the j-step ahead forecast for θ_{m+j} for any $j = 1,2,\ldots$.

We propose to extend these results to the situation where τ^2 and σ^2 are unknown by supposing that any prior information about the variances may be adequately represented by taking $\nu_1\kappa_1/\tau^2$ and $\nu_2\kappa_2/\sigma^2$ to possess independent χ^2-distributions with ν_1 and ν_2 degrees of freedom respectively. The values κ_1^{-1} and κ_2^{-1} could be specified as the respective prior means of the precisions τ^{-2} and σ^{-2}; ν_1 and ν_2 are then prior 'sample sizes' measuring the strength of the prior information. For example, as $\nu_1 \to \infty$ the first-stage variance τ^2 becomes known and equal to κ_1. Ignorance priors will be discussed in Section 4; the presence of prior information is not requisite for the practical applicability of our method.

2. Bayesian Theory.

Under the prior assumptions described in the previous section the joint prior density of τ^2 and $\alpha = \sigma^2/\tau^2$ is given by

$$\pi(\tau^2,\alpha) \propto$$

$$(\tau^2)^{-\frac{1}{2}(\nu_1+\nu_2+2)}\,\alpha^{-\frac{1}{2}(\nu_2+2)}\exp\{-\tfrac{1}{2}\nu_1\kappa_1/\tau^2 - \tfrac{1}{2}\nu_2\kappa_2/\alpha\tau^2\} \tag{2.1}$$

$$(0 < \tau^2 < \infty;\ 0 < \alpha < \infty) \ .$$

We seek the posterior distribution of the variance ratio α after m observations $\underset{\sim}{y}^{(m)} = (y,\ldots,y_m)$. It is firstly necessary to find the distribution of $\underset{\sim}{y}^{(m)}$ given τ^2 and α.

For $i = 2,\ldots,m$, we have under the notation introduced in (1.5) and (1.6), that the conditional distribution of $\underset{\sim}{y}^{(i)}$ given $\underset{\sim}{y}^{(i-1)}$, τ^2, and α, is normal with mean a_{i-1} and variance $\tau^2(1 + \alpha + D_{i-1})$. Since y_1, given τ^2 and α is normal with mean θ_0 and variance $\tau^2(1+\gamma)$, the required distribution of $\underset{\sim}{y}^{(m)}$ is

$$\begin{aligned} p(\underset{\sim}{y}^{(m)}|\tau^2,\alpha) &= p(y_1|\tau^2,\alpha)\prod_{i=2}^{m} p(y_i|\underset{\sim}{y}^{(i-1)},\tau^2,\alpha) \\ &\propto (\tau^2)^{-\frac{1}{2}m}\, U_1(\alpha)\exp\{-\tfrac{1}{2}\tau^{-2}U_2(\alpha)\} \end{aligned} \tag{2.2}$$

with

$$U_1(\alpha) = (1+\gamma)^{-\frac{1}{2}}\prod_{i=2}^{m}(1 + \alpha + D_{i-1})^{-\frac{1}{2}} \tag{2.3}$$

and

$$U_2(\alpha) = (1+\gamma)^{-\frac{1}{2}}(x_1 - \theta_0)^2 + \sum_{i=2}^{m}(1 + \alpha + D_{i-1})^{-1}(y_i - a_{i-1})^2 \tag{2.4}$$

where the a_i and D_i may be obtained from the updating relations in (1.5) and (1.6).

By Bayes' theorem, the joint posterior distribution of τ^2 and α is given by

$$\pi(\tau^2,\alpha|\underset{\sim}{y}^{(m)}) \propto \pi(\tau^2,\alpha)p(\underset{\sim}{y}^{(m)}|\tau^2,\alpha) \qquad (0 < \tau^2 < \infty m\ 0 < \alpha < \infty) \tag{2.5}$$

where the first and second contributions to the right hand side are given in (2.1) and (2.2) respectively. Integrating out τ^2 we find that the marginal posterior distribution of the variance ratio α is

$$\pi(\alpha|\underset{\sim}{y}^{(m)}) \propto U_1^*(\alpha)\ \{U_2^*(\alpha)\}^{-\frac{1}{2}\nu_T} \qquad (0 < \alpha < \infty) \tag{2.6}$$

where

$$U_1^*(\alpha) = \alpha^{-\frac{1}{2}(\nu_2+2)}U_1(\alpha) \tag{2.7}$$

$$U_2^*(\alpha) = \nu_1\kappa_1 + \nu_2\kappa_2/\alpha + U_2(\alpha) \tag{2.8}$$

and

$$\nu_T = \nu_1 + \nu_2 + m \tag{2.9}$$

with $U_1(\alpha)$ and $U_2(\alpha)$ defined in (2.3) and (2.4) respectively.

We are now in a position to obtain the unconditional posterior mean a_m^*, of θ_m, after m stages. This is given by the expectation

$$a_m^* = E(\theta_m|\underset{\sim}{y}_m) = \int_0^\infty E(\theta_m|\underset{\sim}{y}^{(m)},\alpha)\pi(\alpha|\underset{\sim}{y}^{(m)})d\alpha \tag{2.10}$$

of the conditional posterior mean given α (i.e. a_m from (1.5)) with respect to the posterior distribution of α, given $\underset{\sim}{y}^{(m)}$, in (2.6). Note that a_m^* may be calculated by a straightforward one-dimensional integration.

The above procedure possesses the appealing property that it is easy to update in time; this aspect will be considered in the next Section. For $\nu_T > 2$ the posterior variance after m time stages is equal to the expectation of the quantity

$$(a_m - a_m^*)^2 + (\nu_T - 2)^{-1} U_2^*(\alpha) D_m \tag{2.11}$$

with respect to the same distribution of α in (2.6), where a_m, D_m, $U_2^*(\alpha)$, and ν_T are given in (1.5), (1.6), (2.8), and (2.10) respectively.

It is moreover possible to compute the whole posterior density of θ_m, given $\underset{\sim}{y}^{(m)}$, using

$$\begin{aligned} \pi(\theta_m|\underset{\sim}{y}^{(m)}) &= \int_0^\infty \pi(\theta_m,\alpha|\underset{\sim}{y}^{(m)})d\alpha \\ &\propto \int_0^\infty D_m^{-1/2} U_1^*(\alpha)\{U_2^*(\alpha) + D_m^{-1}(\theta_m - a_m)^2\}^{-1/2(\nu_T+1)} d\alpha \end{aligned} \tag{2.12}$$

with $U_1^*(\alpha)$ and $U_2^*(\alpha)$ defined in (2.7) and (2.8) respectively.

The posterior density of the first stage variance τ^2 may be obtained by integrating the quantity in (2.5) with respect to α; however when $\nu_T > 2$ the posterior mean of τ^2 may be obtained more simply, using

$$\begin{aligned} E(\tau^2|\underset{\sim}{y}^{(m)}) &= \int_0^\infty E(\tau^2|\underset{\sim}{y}^{(m)},\alpha)\pi(\alpha|\underset{\sim}{y}^{(m)})d\alpha \\ &= (\nu_T - 2)^{-1} \int_0^\infty U_2^*(\alpha)\pi(\alpha|\underset{\sim}{y}^{(m)})d\alpha \end{aligned} \tag{2.13}$$

where the first and second contribution to the integrand are given in (2.8) and (2.5) respectively.

Lastly, it is straightforward to compute the predictive distribution after m time stages for a future observation y_{m+j}. For $j = 1,2,\ldots$ this will possess mean a_m^* in (2.10), and density

$$p(y_{m+j}|\underset{\sim}{y}^{(m)}) = \int_0^\infty p(y_{m+j},\alpha|\underset{\sim}{y}^{(m)})d\alpha$$

$$\propto \int_0^\infty D_{m,j}^{-1/2} U_1^*(\alpha)\{U_2^*(\alpha) + D_{m,j}^{-1}(y_{m+j} - a_m)^2\}^{-1/2(\nu_T+1)} d\alpha \qquad (2.14)$$

$$(-\infty < y_{m+j} < \infty)$$

where $D_{m,j} = 1 + j\alpha + D_m$.

Our approach provides us with formally justified filtering and forecasting procedures, and enables us to make inferences at any time stage about the process parameter θ_i, the variances τ^2 and σ^2, and the future observations y_{m+j} by considering their posterior or predictive distributions.

3. Updating Procedures

In order to perform the numerical integration in (2.10), store a_m in (1.5) and the distribution in (2.6) for the values of α lying in a set

$$\Omega = \{h, 2h, 3h, \ldots, \ell h\} \quad . \tag{3.1}$$

The integer ℓ and width h choose be chosen after balancing considerations of computer speed and storage with the degree of accuracy required in the numerical integration. With Simpson's rule, $\ell = 100$ and $h = 0.01$ should suffice. We introduce the notation $W(\alpha)$ to denote the expression on the right hand side of (2.6).

After any time stage i it is only necessary to store the latest values of a_i, D_i, $U_1^*(\alpha)$, $U_2^*(\alpha)$ and $W(\alpha)$ for each $\alpha \in \Omega$. Previous values and observations may be discarded. The relevant quantities are defined in (1.5), (1.6), (2.7), (2.8), and (2.6) respectively. For example, after time $i = 1$ we have

$$a_1 = \theta_0 + D_1(x_1 - \theta_0) \quad ,$$

$$D_1 = \gamma/(1+\gamma)$$

$$U_1^*(\alpha) = \alpha^{-\frac{1}{2}(\nu_2+2)} \, , \; U_2^*(\alpha) = \nu_1\kappa_1 + \nu_2\kappa_2/\alpha$$

and

$$W(\alpha) = U_1^*(\alpha)\{U_2^*(\alpha)\}^{-\frac{1}{2}(\nu_1+\nu_2)} \quad .$$

Given the stored values after time $i - 1$ it is straightforward to update to new values after reading in a fresh observation y_i at time i. The validity of our routine may be checked from (1.5), (1.6), (2.3) and (2.4), step (ii) was devised to minimize problems of exponential overflow when calculating $W(\alpha)$ for large i. The following four steps should be completed in order for each $\alpha \in \Omega$.

(i) Calculate a new value for $U_1^*(\alpha)$ by dividing the old value by $(1 + \alpha + D_{i-1})^{1/2}$.

(ii) Calculate the quantity

$$z_i = (1 + \alpha + D_{i-1})^{-1}(y_i - a_{i-1})^2$$

and obtain a new value for $W(\alpha)$ by multiplying the old value by

$$(1 + \alpha + D_{i-1})^{-1/2} Q_i(z_i)$$

with

$$Q_i(q) = \{U_2^*(\alpha)\}^{-1/2}\{1 + q/U_2^*(\alpha)\}^{-1/2(\nu_1+\nu_2+i)} \quad . \tag{3.2}$$

(iii) Calculate a new value for $U_2^*(\alpha)$ by adding z_i to the old value.

(iv) Calculate new values a_i and D_i from (1.5) and (1.6) respectively.

After each time stage the function $W(\alpha)$ should be normalized by dividing through by its total over the set Ω.

We seem to have provided a simple procedure for updating, to any required accuracy, the posterior distribution in (2.6). Note, from (2.10) that the unconditional - an a_i^* of θ_i may now be computed by numerically integrating a_i over $\alpha \in \Omega$, and with respect to this distribution.

It is straightforward to employ the stored values after time i to calculate all the means, variances, and distributions described in the previous Section. It, for example, follows from (2.11) that the posterior distribution of θ_i after time i is proportional to the integral with respect to α of

$$W(\alpha)D_i^{-1/2} Q_i\{D_i^{-1}(\theta_i - a_i)^2\} \tag{3.3}$$

where $Q_i(q)$ is defined in (3.2). The expression in (3.3) may be averaged over $\alpha \in \Omega$ for each required value of θ_i. The obvious analogous procedure is available for the predictive distribution in (2.13).

4. An On-Line Analysis of Chemical Process Readings

The data in the second column of Table 1 were reported by Box and Jenkins (p. 525); we have subtracted 17.0 from each observation. On p. 239 Box and Jenkins fit in ARIMA model in the complete set of 197 observations. Their identified model takes the form given in (1.3), with $\zeta = 0.70$. This corresponds to a value for the variance ratio of $\alpha = 0.13$.

We firstly proceed under an assumption of prior ignorance about θ_1, σ^2, and τ^2 and set $\gamma = \infty$ and hence $D_1 = 1$. The term $(1+\gamma)^{-1/2}$ on the right hand side of (2.3) should then be removed together with the first term on the right hand side of (2.4).

We select the improper prior density

$$\pi(\tau^2,\alpha) \propto \tau^2 \quad (0 < \tau^2 < \infty;\ 0 < \alpha < \infty) \tag{4.1}$$

for τ^2 and α because this particular choice ensures that the posterior distribution will remain proper for an $i = 1,\ldots,m$. This is equivalent to setting $\kappa_1 = \kappa_2 = 0$ and replacing ν_1 and ν_2 by 2 and -2 respectively in the analysis of Section 3.

We obtained, under this ignorance prior, the smooth values a_i^*, for θ_i, listed under (A) in the third column of Table 1; the posterior means α_A of α are in the fourth column. The smoothed value a_i^* for, say, $i = 20$, was calculated from (2.10) and only depends upon the first twenty observations y^{20}. It is therefore the latest on-line value for the process level after 20 stages of the chemical experiment.

The estimate α_A is initially rather large, causing the a_i^* to, very reasonably, remain close to the observations. As the process proceeds, α_A tends to get smaller, and greater smoothing ensued. After time $m = 197$ our values are $x_m = 0.4$, $a_m^* = 0.49$ and $\alpha_A = 0.20$. The final posterior mean in (2.13) of the first stage variance is equal to 0.066. Our value for α_A

implies that this is five times as large as the second stage variance suggesting that a moderately large amount of noise in the process is accounted for by the first stage fluctuations, but that the process parameters still vary noticeably with time.

The final posterior density of α possessed a mode at $\alpha = 0.13$. This, quite interestingly, is absolutely identical to the value recommended by Box and Jenkins. However, both the posterior distribution and likelihood of α are positively skew with thick tails, suggesting that our recommended value 0.20 (i.e. the posterior mean) might also be plausible.

We calculated the whole posterior density curve of θ_{197} from (2.12) and found this to be almost exactly normal with mean 0.49 and variance 0.022. Similarly, the predictive density curves for $y_{198}, y_{199}, \ldots, y_{202}$ were all well-approximated by normal distributions with mean 0.49 and respective variances 0.101, 0.114, 0.127, 0.140, and 0.153. Note that these distributions are easily calculated after any time stage.

The analysis was repeated under an assumption of definite prior information about τ^2 and σ^2. We set $\kappa_1 = 0.05$, $\kappa_2 = 0.025$, and $\nu_1 = \nu_2 = 10$, leading to a prior mean of 0.63 for the variance ratio α. The smoothed values for analysis B are listed in the fifth column of Table 1, and the corresponding posterior means α_B for α in the sixth column.

A possible advantage of a proper prior distribution for α is that the smoothed process settles down more quickly; smaller values are observed for α_B in the initial stages. The estimate α_B however remains somewhat larger than α_A in our example, in view of the information introduced by the prior distribution.

Table 1: An On-Line Analysis of Chemical Process Concentration Readings

Time Stage (i)	Observation (x_i)	Smoothed Value A	α_A	Smoothed Value B	α_B
1	0.0	0.00	5.00	0.00	0.63
2	-0.4	-0.33	5.00	-0.24	0.48
3	-0.7	-0.63	5.25	-0.47	0.48
4	-0.0	-0.85	5.54	-0.68	0.49
5	0.1	-0.08	4.92	-0.31	0.39
6	-0.1	-0.11	4.90	-0.22	0.41
7	-0.2	-0.19	4.83	-0.21	0.42
8	0.4	0.28	4.87	0.06	0.43
9	0.1	0.12	4.72	0.08	0.44
10	0.0	0.02	4.67	0.04	0.44
11	-0.3	-0.23	4.70	-0.12	0.44
12	0.4	0.26	4.35	0.12	0.41
13	0.2	0.20	4.24	0.15	0.42
14	0.4	0.35	4.24	0.27	0.43
15	0.4	0.38	4.22	0.33	0.45
16	0.0	0.09	4.08	0.17	0.44
17	0.3	0.25	3.85	0.23	0.44
18	0.2	0.21	3.74	0.22	0.44
19	0.4	0.35	3.67	0.30	0.45
20	-0.2	-0.05	3.46	0.07	0.43
21	0.1	0.07	3.16	0.09	0.43
22	0.4	0.30	3.11	0.23	0.43
23	0.4	0.36	3.13	0.31	0.43
24	0.5	0.45	3.16	0.40	0.44
25	0.4	0.41	3.05	0.40	0.45
26	0.6	0.54	3.02	0.49	0.46
27	0.4	0.44	2.84	0.45	0.46
28	0.3	0.35	2.80	0.38	0.46
29	0.0	0.11	3.00	0.20	0.46
30	0.8	0.56	2.26	0.48	0.42
31	0.5	0.50	2.02	0.49	0.42
32	1.1	0.88	2.39	0.77	0.43
33	0.5	0.62	1.69	0.64	0.41
34	0.4	0.49	1.70	0.53	0.41
35	0.4	0.45	1.68	0.47	0.41
36	0.1	0.25	1.85	0.30	0.42
37	0.6	0.46	1.44	0.44	0.40
38	0.7	0.59	1.52	0.56	0.41
39	0.4	0.48	1.31	0.49	0.40
40	0.8	0.65	1.23	0.63	0.40
41	0.6	0.61	1.12	0.61	0.40
42	0.5	0.55	1.06	0.56	0.40
43	-0.5	-0.06	1.80	0.09	0.39
44	0.8	0.43	0.47	0.41	0.31
45	0.3	0.37	0.40	0.36	0.31
46	0.3	0.35	0.39	0.34	0.31
47	0.1	0.26	0.39	0.24	0.31
48	0.4	0.32	0.35	0.31	0.31
49	-0.1	0.16	0.35	0.14	0.31
50	0.3	0.22	0.31	0.21	0.31

Table 1 (continued)

Time Stage (i)	Observation (x_i)	Smoothed Value A	α_A	Smoothed Value B	α_B
51	0.6	0.36	0.30	0.37	0.30
52	-0.1	0.20	0.25	0.18	0.29
53	-0.3	0.02	0.31	-0.01	0.30
54	-0.2	-0.06	0.34	-0.09	0.31
55	-0.2	-0.11	0.35	-0.13	0.31
56	0.2	0.02	0.30	0.01	0.30
57	-0.2	-0.06	0.29	-0.07	0.30
58	0.6	0.18	0.27	0.19	0.29
59	0.2	0.18	0.25	0.19	0.29
60	-0.4	-0.02	0.23	-0.04	0.28
61	0.1	0.02	0.21	0.02	0.27
62	-0.1	-0.01	0.21	-0.03	0.27
63	-0.4	-0.14	0.22	-0.18	0.28
64	1.0	0.22	0.17	0.27	0.24
65	0.2	0.20	0.16	0.24	0.23
66	0.3	0.23	0.16	0.26	0.24
67	0.0	0.16	0.15	0.16	0.23
68	-0.1	0.08	0.15	0.07	0.23
69	0.3	0.15	0.14	0.15	0.23
70	-0.2	0.05	0.14	0.02	0.23
71	0.3	0.12	0.13	0.13	0.23
72	0.4	0.20	0.13	0.23	0.23
73	0.7	0.34	0.14	0.40	0.23
74	-0.2	0.18	0.12	0.18	0.22
75	-0.1	0.11	0.12	0.08	0.22
76	0.0	0.08	0.12	0.05	0.22
77	-0.1	0.03	0.12	0.00	0.22
78	0.0	0.03	0.12	0.00	0.22
79	-0.4	-0.09	0.13	-0.15	0.23
80	-0.3	-0.15	0.13	-0.20	0.23
81	-0.2	-0.16	0.13	-0.20	0.23
82	-0.3	-0.20	0.14	-0.24	0.23
83	-0.6	-0.32	0.15	-0.37	0.24
84	-0.5	-0.37	0.15	-0.42	0.24
85	-0.6	-0.44	0.16	-0.49	0.24
86	-0.4	-0.42	0.16	-0.45	0.24
87	-0.5	-0.44	0.16	-0.47	0.24
88	-0.3	-0.40	0.16	-0.41	0.24
89	-0.6	-0.46	0.16	-0.48	0.24
90	-0.6	-0.50	0.16	-0.52	0.24
91	-0.8	-0.59	0.16	-0.63	0.25
92	-0.6	-0.59	0.16	-0.62	0.25
93	-0.7	-0.62	0.16	-0.65	0.25
94	-0.6	-0.61	0.16	-0.63	0.25
95	0.0	-0.42	0.16	-0.39	0.24
96	-0.1	-0.32	0.17	-0.28	0.25
97	0.1	-0.18	0.19	-0.13	0.26
98	0.1	-0.09	0.20	-0.04	0.27
99	-0.3	-0.17	0.19	-0.15	0.26
100	-0.1	-0.15	0.19	-0.13	0.26

Table 1 (continued)

Time Stage (i)	Observation (x_i)	Smoothed Value A	α_A	Smoothed Value B	α_B
101	-0.5	-0.26	0.18	-0.27	0.26
102	0.2	-0.11	0.17	-0.09	0.25
103	-0.6	-0.27	0.16	-0.29	0.25
104	0.0	-0.18	0.16	-0.18	0.24
105	0.0	-0.13	0.16	-0.11	0.24
106	-0.3	-0.18	0.15	-0.18	0.24
107	-0.8	-0.37	0.16	-0.42	0.24
108	-0.4	-0.38	0.16	-0.41	0.24
109	-0.1	-0.29	0.15	-0.29	0.24
110	-0.5	-0.35	0.15	-0.37	0.24
111	-0.4	-0.37	0.15	-0.38	0.24
112	-0.4	-0.38	0.15	-0.39	0.24
113	0.0	-0.26	0.14	-0.24	0.24
114	0.1	-0.15	0.15	-0.11	0.24
115	0.1	-0.08	0.16	-0.03	0.25
116	-0.3	-0.15	0.15	-0.14	0.24
117	-0.2	-0.17	0.15	-0.16	0.24
118	-0.7	-0.33	0.15	-0.36	0.24
119	-0.4	-0.35	0.15	-0.38	0.24
120	-0.2	-0.30	0.14	-0.31	0.24
121	-0.1	-0.24	0.14	-0.23	0.24
122	0.1	-0.13	0.14	-0.11	0.24
123	-0.2	-0.16	0.14	-0.14	0.24
124	0.0	-0.11	0.14	-0.09	0.24
125	0.2	-0.02	0.15	0.02	0.24
126	0.3	0.08	0.16	0.13	0.25
127	0.2	0.12	0.16	0.15	0.25
128	0.3	0.17	0.17	0.21	0.25
129	0.2	0.18	0.17	0.20	0.25
130	0.2	0.18	0.17	0.20	0.25
131	0.5	0.29	0.17	0.32	0.26
132	-0.1	0.16	0.16	0.16	0.25
133	-0.1	0.08	0.16	0.06	0.25
134	-0.1	0.02	0.16	0.00	0.25
135	0.0	0.02	0.16	0.00	0.25
136	-0.5	-0.15	0.17	-0.19	0.26
137	-0.3	-0.20	0.18	-0.23	0.26
138	-0.2	-0.20	0.17	-0.22	0.26
139	-0.3	-0.23	0.18	-0.25	0.26
140	-0.3	-0.25	0.18	0.27	0.26
141	-0.4	-0.30	0.18	-0.32	0.26
142	-0.5	-0.36	0.18	-0.39	0.27
143	0.0	-0.24	0.17	-0.24	0.26
144	-0.3	-0.26	0.17	-0.26	0.26
145	-0.3	-0.27	0.17	-0.28	0.26
146	-0.1	-0.22	0.17	-0.21	0.26
147	0.4	-0.01	0.18	0.03	0.26
148	0.1	0.02	0.18	0.06	0.27
149	0.0	0.01	0.18	0.03	0.27
150	-0.2	-0.06	0.18	-0.06	0.26

Table 1 (continued)

Time Stage (i)	Observation (x_i)	Smoothed Value A	α_A	Smoothed Value B	α_B
151	0.2	0.03	0.18	0.04	0.26
152	0.2	0.08	0.18	0.10	0.26
153	0.4	0.19	0.19	0.22	0.27
154	0.2	0.19	0.18	0.21	0.27
155	-0.1	0.09	0.18	0.09	0.26
156	-0.2	0.01	0.18	-0.02	0.27
157	0.0	0.00	0.18	-0.01	0.27
158	0.4	0.13	0.17	0.15	0.26
159	0.2	0.15	0.17	0.17	0.26
160	0.2	0.17	0.18	0.18	0.26
161	0.1	0.14	0.17	0.15	0.26
162	0.1	0.13	0.17	0.13	0.26
163	0.1	0.12	0.17	0.12	0.26
164	0.4	0.21	0.17	0.23	0.26
165	0.2	0.21	0.17	0.22	0.26
166	-0.1	0.11	0.17	0.09	0.26
167	-0.1	0.04	0.17	0.02	0.26
168	0.0	0.03	0.17	0.01	0.26
169	-0.3	-0.08	0.18	-0.11	0.27
170	-0.1	-0.08	0.18	-0.11	0.27
171	0.3	0.04	0.17	0.05	0.26
172	0.8	0.29	0.18	0.35	0.27
173	0.8	0.48	0.21	0.53	0.29
174	0.6	0.52	0.22	0.56	0.29
175	0.5	0.51	0.22	0.53	0.29
176	0.0	0.33	0.20	0.32	0.28
177	-0.1	0.18	0.21	0.15	0.29
178	0.1	0.15	0.21	0.13	0.29
179	0.2	0.17	0.21	0.16	0.29
180	0.4	0.25	0.20	0.26	0.29
181	0.5	0.34	0.20	0.36	0.29
182	0.9	0.54	0.22	0.58	0.30
183	0.0	0.34	0.19	0.34	0.28
184	0.0	0.23	0.19	0.21	0.28
185	0.0	0.15	0.20	0.12	0.28
186	0.2	0.17	0.19	0.16	0.28
187	0.3	0.22	0.19	0.21	0.28
188	0.4	0.28	0.19	0.29	0.28
189	0.4	0.32	0.19	0.33	0.28
190	0.0	0.21	0.19	0.20	0.28
191	1.0	0.47	0.18	0.52	0.27
192	1.2	0.74	0.22	0.80	0.29
193	0.6	0.68	0.20	0.71	0.28
194	0.8	0.72	0.21	0.75	0.28
195	0.7	0.71	0.21	0.73	0.28
196	0.2	0.53	0.20	0.52	0.28
197	0.4	0.49	0.20	0.47	0.28

5. Poisson Observations

Suppose now that we replace the first stage of the model in (1.1) by the assumption that $y_1, y_2, \ldots, y_m$, are independent and Poisson distributed with respective means $\theta_1, \theta_2, \ldots, \theta_m$. The y_i might for example represent the numbers of items per successive batch by an industrial process. A similar analysis to the one described below may be developed if y_i instead possesses a binomial distribution with probability θ_i and sample size n_i (e.g. $n_i = 1$ for binary process).

In the Poisson situation we retain a similar second stage to the model by supposing that

$$\theta_i = \theta_{i-1} + \varepsilon_i \qquad (i = 1,\ldots,m) \tag{5.1}$$

where $\varepsilon_1,\ldots,\varepsilon_m$ are independent terms with zero means. For $i = 2,\ldots,m$ we suppose that ε_i possesses variance $\alpha\theta_{i-1}$, so that the variance changes with the non-negative mean in a sensible (and technically convenient) manner. The first error term ε_1 is taken to possess variance $\gamma\theta_0$, so that θ_1 has a prior distribution with mean θ_0 and variance $\gamma\theta_0$. No further distributional assumptions need to be made about $\varepsilon_1,\ldots,\varepsilon_m$ until we present ourselves with the problem of estimating α.

6. Linear Prediction

Let $\theta^*_{m,0}$ denote our smoothed value for θ_m, given $\underset{\sim}{y}^{(m)}$, and for $j = 1,2,\ldots$ let $\theta^*_{m,j}$ represent our j-step ahead forecast for θ_{m+j} after time m. Attention is for the moment restricted to linear predictors of the form

$$\theta^*_{m,j} = \beta_{0j} + \sum_{i=1}^{m} \beta_{ij} y_i \qquad (j = 0,1,2,\ldots) \tag{6.1}$$

where $\beta_{0j},\ldots,\beta_{mj}$ represent unknown constants not depending upon the data, and α is initially taken to be known. Optimal values for $\beta_{0j},\ldots,\beta_{mj}$ may be obtained by minimizing the Bayes risk of $\theta^*_{m,j}$ under the quadratic loss function

$$L(\theta^*_{m,j},\theta_{m+j}) = (\theta^*_{m,j} - \theta_{m+j})^2 \quad . \tag{6.2}$$

The first two moments of the joint distribution, given α, of $y_1,\ldots,y_m$ and $\theta_1,\ldots,\theta_m$ is described in Section 5, and the Bayes risk may be obtained by taking expectation of the loss function in (6.2) with respect to this joint distribution. Setting $\beta_{ij} = \beta_i$, for $i = 0,1,\ldots,m$ and $j = 0, 1,2,\ldots,$ for notational simplicity we find after some manipulation that the Bayes risk is given by

$$\begin{aligned} E\{L(\theta^*_{m,j},\theta_{m+j})|\alpha\} = {} & \{\beta_1\theta_0 +\cdots+ \beta_m\theta_0 + \beta_0 - \theta_0\}^2 \\ & + j\alpha\theta_0 \\ & + (\beta_1^2 +\cdots+ \beta_m^2)\theta_0 \\ & + \alpha\{(\beta_m - 1)^2 + (\beta_m + \beta_{m-1} - 1)^2 +\cdots+ (\beta_m +\cdots+ \beta_2 - 1)^2\}\theta_0 \\ & + \gamma\{\beta_m +\cdots+ \beta_1 - 1\}^2\theta_0 \quad . \end{aligned} \tag{6.3}$$

Differentiating the expression in (6.3) with respect to the β_i we find that the optimal values of β_i do not depend upon j. The optimal linear values $\theta^*_{m,j}$ in (6.1) may therefore be set equal to a common quantity a_m for all $j = 0, 1, 2, \ldots$.

We now introduce an analogy with the normal process of Section 1 in order to obtain updating formulae for a_m. It is straightforward to show that this normal process provides exactly the same Bayes risk as given in (6.3), but with θ_0, in all but the first term on the right hand side of (6.3), replaced by the first stage variance τ^2 and α replaced by σ^2/τ^2. The optimal linear predictors in the Poisson case are therefore, with appropriate substitutions identical to those in the normal case. However the optimal linear predictors $\theta^*_{m,j}$ in the normal case are identical to the posterior mean a_m, as *normality implies linearity*. We therefore have the pleasing result that a_m in the Poisson situation may be obtained from exactly the same updating formulae as employed in Section 1, namely

$$a_i = a_{i-1} + D_i(y_i - a_{i-1}) \qquad (i = 1,\ldots,m) \tag{6.4}$$

and

$$D_i = (D_{i-1} + \alpha)/(D_{i-1} + \alpha + 1) \qquad (i = 2,\ldots,m) \tag{6.5}$$

with $a_0 = \theta_0$ and $D_1 = \gamma/(1+\gamma)$.

The value a_m could be viewed as the 'best' linear approximation to the posterior mean of θ_m after time m. In the next Section we show how this result may be approximately generalized to the situation where α is unknown.

7. Estimation of α

In order to obtain an approximation to the posterior distribution of α we introduce an alternative method which also approximately justifies the updating formulae in (6.4) and (6.5). Suppose that after $i-1$ time stages the posterior distribution of θ_{i-1} given α, possesses mean a_{i-1} and variance v_{i-1}, i.e.

$$\theta_{i-1} | \underset{\sim}{y}^{(i-1)}, \alpha \sim (a_{i-1}, v_{i-1}) \quad . \tag{7.1}$$

Then

$$\theta_i | \underset{\sim}{y}^{(i-1)}, \alpha \sim (a_{i-1}, v_{i-1} + \alpha a_{i-1}) \quad . \tag{7.2}$$

Suppose that the distribution in (7.2) is approximately Gamma with the stipulated mean and variance. As θ_i in Poisson with mean θ_i we obtain, applying Bayes theorem and some manipulation

$$\theta_i | \underset{\sim}{y}^{(i)}, \alpha \sim (a_i, v_i) \tag{7.3}$$

where

$$a_i = a_{i-1} + v_i (y_i - a_{i-1}) \tag{7.4}$$

and

$$v_i = \left\{ \frac{v_{i-1} + \alpha a_{i-1}}{v_{i-1} + (\alpha+1) a_{i-1}} \right\} .$$

Setting $v_i = D_i a_i$ yields the relations in (6.4) and (6.5). The approximations also give

$$y_i | \underset{\sim}{y}^{(i-1)}, \alpha \sim (a_{i-1}, \{D_{i-1} + \alpha + 1\} a_{i-1}) \tag{7.5}$$

and that this distribution is approximately a Gamma mixture of a Poisson. The density is given by

$$p(y_i | \underset{\sim}{y}^{(i-1)}, \alpha) = \frac{\Gamma(a_{i-1} g_{i-1} + y_i) g_{i-1}^{a_{i-1} g_{i-1}}}{\Gamma(y_{i+1}) \Gamma(a_{i-1} g_{i-1}) (1 + g_{i-1})^{a_{i-1} g_{i-1} + y_i}} \tag{7.6}$$

where

$$g_{i-1} = (D_{i-1} + \alpha)^{-1} \quad . \tag{7.7}$$

The posterior density of α after m time stages is therefore approximated by

$$\pi(\alpha|\underset{\sim}{y}^{(m)}) \propto \pi(\alpha) \prod_{i=1}^{m} p(y_i|\underset{\sim}{y}^{(i-1)},\alpha)$$

$$\propto \pi(\alpha) \prod_{i=1}^{m} \{g_{i-1}/(1 + g_{i-1})\}^{a_{i-1}g_{i-1}} \prod_{k=0}^{y_{i-1}} \left\{\frac{a_{i-1}g_{i-1} + k}{g_{i-1} + 1}\right\} \tag{7.8}$$

$$(0 < \alpha < \infty)$$

where $\pi(\alpha)$ denotes the prior density; we recommend the flexible choice

$$\pi(\alpha) \propto \alpha^{\frac{1}{2}\nu_1 - 2} (\nu_1\kappa + \nu_2\alpha)^{-\frac{1}{2}(\nu_1+\nu_2)} \tag{7.9}$$

so that $\kappa^{-1}\alpha$ possesses an F-distribution with ν_1 and ν_2 degrees of freedom. For $\nu_2 > 2$ the prior mean of α is then equal to $\nu_2\kappa/(\nu_2-2)$.

The unconditional mean a_m^* and variance v_m^* of θ_m after m observations may be approximated by taking expectations with respect to the distribution in (7.8) of a_m in (6.4) and

$$(a_m - a_m^*)^2 + D_m a_m \quad .$$

The predictive mean of any future observation y_{m+j} is also approximated by a_m^* and the predicted variance of y_{m+j} may be approximated by taking expectations with respect to the distribution in (7.8) of the quantity

$$(a_m - a_m^*)^2 + (1 + j\alpha + D_m)a_m \quad .$$

It may be reasonable to take the unconditional predictive distributions of the y_{m+j} to be approximately Poisson-Gamma with the means and variances described above.

Updating in the Poisson situation is just as straightforward as for the method for normal observations described in Section 3. After any time stage i it is necessary to store the values a_i, D_i, and $W(\alpha)$ for all α lying in Ω in (3.1), where $W(\alpha)$ denotes the expression in the right hand side of (7.8).

Given the stores values after time $i-1$ we update to new values after time i by multiplying the old value of $W(\alpha)$ by the expression in the right hand side of (7.6), and using (6.4) and (6.5) to update a_i and D_i.

9. An Analysis of an Industrial Control Chart

The 52 Poisson observations in the second column of Table 2 were introduced by Hald ([3], p. 720), and represent numbers of defective items in consecutive shifts of an industrial process.

The smoothed values (A) in the third column correspond to a uniform prior distribution for α (set $\nu_1 = 2$, $\nu_2 = -2$, $\kappa = 0$, and $\gamma = \infty$ in the analysis of Section 7). The posterior distribution of α after 52 observations possessed a mode at $\alpha = 0$ and a very thick tail. The value $\alpha_A = 0.05$ suggests that whilst the random noise in the process is primarily caused by the first-stage fluctuations of the Poisson observations y_i about their mean θ_i, there is some definite evidence that the θ_i are changing in time. The final smoothed value of 2.93 may be compared with the average 3.23 of the 52 observations. The predictive variance of x_{53} is about 1.24 times its predictive mean of 2.93, suggesting that the predictive distribution of x_{53} is not approximately Poisson. A Poisson-Gamma distribution with this mean and variance might be more reasonable.

We also employed the F-distribution in (7.9) as prior for α. The choices $\nu_1 = \nu_2 = 10$, $\kappa = 0.20$, and $\gamma = \infty$ lead to a prior mean of 0.25 for α. The corresponding smoothed values (B) are listed in the fifth column of Table 2; the estimates α_B in the sixth column fluctuate rather less than the α_A in the fourth column. The final posterior distribution of α possessed a mode at $\alpha = 0.07$, its mean at $\alpha = 0.10$, and a thick positive tail.

Table 2: Analysis of an Industrial Control Chart

Time Stage (i)	Observation (x_i)	Smoothed Value A	α_A	Smoothed Value B	α_B
1	3	3.00	0.50	3.00	0.25
2	1	1.81	0.50	1.89	0.24
3	0	0.90	0.49	1.07	0.24
4	7	3.89	0.51	3.44	0.25
5	3	3.39	0.48	3.24	0.24
6	4	3.63	0.46	3.50	0.23
7	4	3.76	0.43	3.66	0.22
8	4	3.83	0.41	3.77	0.22
9	5	4.29	0.39	4.19	0.21
10	3	3.72	0.36	3.76	0.20
11	2	3.03	0.35	3.16	0.20
12	3	3.06	0.32	3.12	0.19
13	3	3.07	0.29	3.09	0.19
14	5	3.76	0.26	3.72	0.18
15	1	2.77	0.27	2.82	0.18
16	2	2.55	0.25	2.57	0.18
17	2	2.41	0.24	2.40	0.17
18	5	3.28	0.20	3.24	0.16
19	2	2.87	0.18	2.84	0.16
20	5	3.49	0.17	3.51	0.15
21	2	3.04	0.15	3.04	0.15
22	1	2.47	0.17	2.41	0.15
23	3	2.68	0.14	2.61	0.14
24	2	2.53	0.13	2.43	0.14
25	3	2.68	0.11	2.61	0.14
26	5	3.23	0.10	3.31	0.13
27	1	2.69	0.10	2.63	0.13
28	3	2.79	0.09	2.74	0.13
29	0	2.17	0.09	1.96	0.13
30	2	2.19	0.08	1.99	0.12
31	6	3.00	0.07	3.12	0.12
32	5	3.38	0.07	3.64	0.12
33	9	5.00	0.14	5.30	0.15
34	4	4.55	0.12	4.84	0.14
35	4	4.32	0.11	4.57	0.13
36	4	4.19	0.10	4.39	0.13
37	4	4.11	0.09	4.27	0.13
38	6	4.54	0.09	4.76	0.12
39	0	3.47	0.09	3.40	0.12
40	7	4.25	0.08	4.42	0.12
41	3	3.96	0.07	4.03	0.11
42	4	3.96	0.07	4.02	0.11
43	4	3 96	0.06	4.02	0.11
44	1	3.37	0.07	3.19	0.11
45	2	3.11	0.07	2.87	0.11
46	3	3.13	0.06	2.93	0.11
47	3	3.14	0.06	2.96	0.11
48	1	2.70	0.07	2.43	0.11
49	2	2.59	0.07	2.33	0.11
50	3	2.72	0.06	2.53	0.11
51	6	3.36	0.05	3.45	0.10
52	1	2.93	0.05	2.80	0.10

REFERENCES

[1] BLIGHT, B. (1974). Recursive solutions for the estimation of a stochastic parameter. J. Amer. Statist. Assoc. 69, 477-481.

[2] BOX, G. E. P. and JENKINS, G. M. Time Series Analysis, Holden Day.

[3] HALD, A. (1967). Statistical Theory with Engineering Applications, Wiley.

[4] HARRISON, P. J. and STEVENS, C. (1976). Bayesian forecasting (with discussion), J. Roy. Statist. Soc. B 39.

[5] KALMAN, R. E. (1960). A new approach to linear filtering and prediction problems, Journal of Basic Engineering, American Society of Mechanical Engineering Transactions, 83D, 299-306.

[6] LEONARD, T. (1975). A Bayesian approach to the linear model with unequal variances, Technometrics 17, ,5-102.

TL/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #2337	2. GOVT ACCESSION NO. AD-A114 612	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Bayesian Approach to Markovian Models for Normal and Poisson Data		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Tom Leonard		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE February 1982
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Kalman filter; ARIMA models; Linear prediction; Variance components; Bayesian updating; Predictive distribution; Quality control; Poisson observations; Linear growth model; Box-Jenkins.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A Bayesian updating procedure is proposed for filtering the process parameters in the two-stage Markovian constant variance model for time varying normal data in the situation where the signal to noise ratio is unknown. A forecasting procedure is described which yields the entire predictive distribution of future observations; a numerical study involves an on-line analysis for chemical process concentration readings. A similar method is developed for Poisson data and applied to the analysis of an industrial control chart.