Change-Point problems: approaches and applications

A.F.M. SMITH

University of Nottingham

SUMMARY

Problems of making inferences about abrupt changes in the mechanism underlying a sequence of observations are considered in both retrospective and on-line contexts. Among the topics considered are the Lindisfarne scribes problem; switching straight lines; manoeuvering targets, and shifts of level or slope in linear time series models. Summary analyses of data obtained in studies of schizophrenic and kidney transplant patients are presented.

Keywords: CHANGE-POINT; SWITCHING STRAIGHT LINES; BAYES FACTOR; KALMAN FILTER.

1. INTRODUCTION

In the simplest possible case, a sequence of random quantities $\tilde{y}_1,\ldots,\tilde{y}_n$ is said to have a change-point at r $(1 \leq r < n)$ if $\tilde{y}_1,\ldots,\tilde{y}_r$ and $\tilde{y}_{r+1},\ldots,\tilde{y}_n$ are exchangeable subsequences, but the combined sequence is not exchangeable. Assuming the usual mixture representation of exchangeable sequences, the most frequently used model of a change-point at r can be written in terms of densities as

$$p\,(y_1,\ldots,y_n|M_r) = \int\int \Pi_{i=1}^{r} p_1(y_i|\theta_1)\,\Pi_{i=r+1}^{n} p_2(y_i|\theta_2)p\,(\theta_1,\theta_2)d\theta_1 d\theta_2, \qquad (1)$$

where M_r denotes the model which assumes a change-point at r, and $p_1(y|\theta_1) \neq p_2(y|\theta_2)$, $p\,(\theta_1,\theta_2)$ have obvious interpretations. It is convenient to denote by M_0 the model which assumes the entire sequence exchangeable and defines

$$p\,(y_1,\ldots,y_n|M_0) = \int\int \Pi_{i=1}^{n} p_1(y_i|\theta_1)p\,(\theta_1,\theta_2)d\theta_1 d\theta_2. \qquad (2)$$

With such a formulation, inference about change-points, given $\tilde{y}_1 = y_1, \ldots, \tilde{y}_n = y_n$, reduces to consideration of the set of alternative models M_0, $M_1, \ldots, M_{n-1}$. These may be conveniently compared pairwise using Bayes factors - ratios of posterior to prior odds - so that, as is easily seen from Bayes theorem,

$$B_{ij} = \frac{p(y_1, \ldots, y_n | M_i)}{p(y_1, \ldots, y_n | M_j)} , \tag{3}$$

the required densities being obtained from (1) and (2). A detailed study of this approach for univariate sequences and a variety of standard parametric distributions is given in Smith (1975). In Section 2 of this paper, we shall outline the extension to more than one possible change-point and illustrate the approach by applying it to the Lindisfarne Scribes problem (Ross, 1950).

In the more general setting of changes in structure of a regression or time series model, the simple characterization in terms of exchangeable subsequences no longer applies, but, provided we specify the model, M_r, corresponding to a change at r, we can use (3) directly to compare alternative models. This approach will be presented for regression models in Section 3 and a possible extension to linear time series models will be outlined in Section 4. Also in Section 3, we shall comment briefly on special problems of interest that arise in the case of switching straight lines.

The analysis in Sections 2-4 concentrates on *retrospective* analysis. In Section 5, we shall consider an alternative linear model formulation, in terms of Kalman filters (Harrison and Stevens, 1976), that seems more suited to *on-line* detection of changes.

2. BINOMIAL DATA: THE LINDISFARNE SCRIBES PROBLEM

The Lindisfarne Scribes problem (Ross, 1950; Silvey, 1958) is of the type described at the beginning of Section 1, but admitting more than one possible change-point. A text divides into n sections, and it is assumed that only one scribe was involved in the writing of any one section, and that sections written by any one scribe are consecutive. We wish to infer how often, and where, changes of scribe occurred. The analysis is to be based on the frequency of occurrence of a certain word which has just two alternative forms. A version of some of the data, taken from Ross (1950), is set out in Table 1.

TABLE 1

Number of occurrences of present indicative 3rd. singular endings s and ð for different sections of Lindisfarne

Section

	1	2	3	4	5	6	7	8	9	10	11	12	13
s...	12	26	31	24	28	34	39	46	41	19	17	17	16
ð...	9	10	13	6	24	11	9	11	7	3	3	4	4
Total...	21	36	44	30	52	45	48	57	48	22	20	21	20

The assumption is made that a scribe is characterized by the propensity with which, when using the present indicative third person singular, he adopts one or other of the two variants. We thus arrive at an example of a change-point problem, with many possible changes, where it might be reasonable to assume underlying binomial distributions.

If $M(r_1,\ldots,r_K)$ is the model which assumes K changes of scribe, with change-points $r_1,\ldots,r_K$, then if $\theta_1,\ldots,\theta_{K+1}$ denote the propensities of the assumed $K+1$ scribes, and m_i, y_i, $i = 1,\ldots,n$, the numbers of word uses and ð-variant uses, respectively, in each of the n sections, we have

$$p\,[y_1,\ldots,y_n \,|\, M(r_1,\ldots,r_K)] = \Pi_{i=1}^{n} \binom{m_i}{y_i} \times$$
$$\int \cdots \int \Pi_{j=1}^{k+1} \theta_j^{s_j} (1-\theta_j)^{f_j}\, p\,(\theta_1,\ldots,\theta_{k+1}) d\theta_1,\ldots,d\theta_{k+1}, \tag{4}$$

where

$$s_j = s(r_{j-1}+1, r_j) = y_{r_{j-1}+1} + \ldots + y_{r_j}$$
$$f_j = f(r_{j-1}+1, r_j) = m_{r_{j-1}+1} + \ldots + m_{r_j} - s(r_{j-1}+1, r_j). \tag{5}$$

There are, of course, no general prescriptions for the choice of $p\,(\theta_1,\ldots,\theta_{K+1})$. In some change-point contexts, for example in reliability studies, one might expect monotonic relationships to hold (Smith, 1977), but for the purpose of this illustration we shall simply consider the (perhaps unreasonable) assignment of independent beta prior densities, so that

$$p\,(\theta_1,\ldots,\theta_{K+1}) = \Pi_{j=1}^{k+1} \theta_j^{\alpha_j-1} (1-\theta_j)^{\beta_j-1} \,/\, B\,(\alpha_j,\beta_j), \tag{6}$$

where B (.,.) denotes the usual beta function. Substituting (6) in (4), the required integration is immediate, and it is easily seen, for example, that

$$B_{0,(r_1,\ldots,r_k)} = \frac{B(\alpha_1+s(1,n),\beta_1+f(1,n))}{B(\alpha_1,\beta_1)} \Big/ \prod_{j=1}^{K+1} \frac{B(\alpha_j+s(r_{j-1}+1,r_j),\beta_j+f(r_{j-1}+1,r_j))}{B(\alpha_j,\beta_j)} \qquad (7)$$

For the particular choice $K = 1$, $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1$, and converting (7) into posterior probabilities on M_0, $M_1,\ldots,M_{13}$ (taking prior probability ½ on M_0, 1/26 on the others) we obtain the results shown in Table 2.

TABLE 2

Posterior probabilities assuming at most one change-point

M_0	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	M_{13}
-	-	-	-	.50	.41	.07	.02	-	-	-	-	-	-

If we go on to consider $K=2$, $\alpha_i=\beta_i=1$, $i=1,2,3$, and Pr $(K = 0) = Pr$ $(K = 1) = Pr$ $(K = 2) = 1/3$, with equal prior probabilities on all thirteen models, given that $K = 1$, and on all seventy-eight models, given that $K = 2$, we obtain the results shown in Tables 3 and 4.

TABLE 3

Posterior probabilities of up to two changes

no change	*one change*	*two changes*
0.00002	0.06856	0.93142

TABLE 4

Posterior probabilities of selected pairs of change-points

	5	6	7	8	9	10
1	.029	.029	-	-	-	-
2	.027	-	-	-	-	-
3	.037	.026	-	-	-	-
4	.287	.049	-	-	-	-
5	-	.042	.043	.048	.029	.029

The analysis so far would appear to indicate strongly that there was a change of scribe after section 4 and again after section 5. In fact, further analysis suggests that there is no strong evidence for further changes. As an example, we note that the Bayes factor for M (4,5) against M (4,5,6) is given, from (7), by

$$B_{0,(4,5,6)} / B_{0,(4,5)} = \frac{46.237}{282} \frac{\binom{45}{11}\binom{236}{41}}{\binom{281}{52}} \approx 3.29 \tag{8}$$

This is, of course, the same result as is obtained by taking sections 6-13 and testing for a change after section 6.

Finally, we note in passing that the calculations required in (7) can be greatly simplified by applying Stirling's approximation. For example, if we have $K = 1$ and define s_1, s_2, f_1, f_2 by (5), then (7) has the form

$$B_{0,r_1} = \frac{(s_1+f_1+1)\,(s_2+f_2+1)}{(n+1)} \cdot \frac{\binom{s_1+f_1}{s_1}\binom{s_2+f_2}{s_2}}{\binom{n}{s_1+s_2}}, \tag{9}$$

which can be shown to be well approximated by

$$B_{0,r_1} = \left[\frac{n\,(s_1+f_1)\,(s_2+f_2)}{2\pi(s_1+s_2)\,(f_1+f_2)}\right]^{1/2} \exp\{-1/2\,\chi^2\}, \tag{10}$$

where

$$\chi^2 = \frac{n\,(s_1 f_2 - s_2 f_1)^2}{(s_1+f_1)(s_2+f_2)(s_1+s_2)(f_1+f_2)} \quad , \tag{11}$$

the latter being the usual χ^2-statistic for testing the equality of the underlying propensities of two independent binomial samples. For (8), the approximation (10) gives 3.37.

At least in the case of a single change-point, the above approach has many points of contact with the Bayesian significance testing approaches of Jeffreys (1961) and Dickey and Lientz (1970).

3. CHANGE IN A REGRESSION RELATIONSHIP

We shall consider the problem of investigating the stability over time of the regression model

$$\tilde{y}_t = \mathbf{x}_t^T \beta^{(t)} + \tilde{\varepsilon}_t \ , \quad t = 1,\ldots,n, \tag{12}$$

where at time, t, $\tilde{y}_t$ is the observation on the dependent variable, $\mathbf{x}_t$ is the column vector of observations on p regressor variables (including, possibly, a constant), $\beta^{(t)}$ is the column vector of unknown regression coefficients and $\tilde{\varepsilon}_t$ is the error term, assumed normally distributed with mean zero and variance σ^2.

In this section, we shall work with independent, homoscedastic errors and non-stochastic regressor variables. In the next section, we shall show how to extend the approach to cover more general situations.

The regression structure defined by (12) will be said to have a change-point at r $(1 \leq r < n)$ if

$$\beta^{(1)} = \ldots = \beta^{(r)} = \beta, \ \beta^{(r+1)} = \ldots = \beta^{(n)} = \beta + \delta$$

with unknown $\delta \neq \mathbf{0}$. We shall denote this model by M_r. The model of no change, $\delta = \mathbf{0}$, will be denoted by M_0.

If we adopt the notation,

$$\tilde{\mathbf{y}}_r^T = (\tilde{y}_1,\ldots,\tilde{y}_r), \quad \tilde{\mathbf{y}}_{(n-r)}^T = (\tilde{y}_{r+1},\ldots,\tilde{y}_n),$$
$$\mathbf{X}_r^T = (\mathbf{x}_1,\ldots,\mathbf{x}_r), \quad \mathbf{X}_{(n-r)}^T = (\mathbf{x}_{r+1},\ldots,\mathbf{x}_n),$$

we see that model M_r $(1 \leq r < n)$ can be written in the form

$$\tilde{\mathbf{y}} \sim N(\mathbf{A}_r\theta, \sigma^2\mathbf{I}_n), \tag{13}$$

where $\tilde{\mathbf{y}} = \tilde{\mathbf{y}}_n$, $\mathbf{I}_n$ is the n x n identity matrix and

$$\mathbf{A}_r = \begin{bmatrix} \mathbf{X}_r & \mathbf{0} \\ \mathbf{X}_{(n-r)} & \mathbf{X}_{(n-r)} \end{bmatrix}, \ \theta = \begin{bmatrix} \beta \\ \delta \end{bmatrix} \tag{14}$$

In the case of M_0, (13) still holds, but with $\mathbf{A}_0 = \mathbf{X}_n$, $\theta = \beta$.

Again, inference about the change-point (is there one? and, if so, where?) reduces to consideration of the possible models M_r. To calculate (3) in this case, we require

$$p\,(\mathbf{y}\,|\,M_r) = \int \ldots \int p\,(\mathbf{y}\,|\,\mathbf{A}_r,\,\theta,\,\sigma)\,p\,(\theta,\,\sigma\,|\,\mathbf{A}_r)\;d\theta d\sigma, \tag{15}$$

and thus need to specify $p\,(\theta,\sigma\,|\,\mathbf{A}_r)$. This specification, and its relation to the whole question of significance tests and choice procedures among alternative linear models has been discussed at some length in the literature. A recent discussion is given by Smith and Spiegelhalter (1980).

In this paper, we shall examine the consequences of the specification,

$$p\,(\theta,\sigma\,|\,\mathbf{A}_r) = p\,(\theta\,|\,\mathbf{A}_r,\sigma)\,p\,(\sigma), \tag{16}$$

where $p\,(\sigma) \propto \sigma^{-1}$, and $p\,(\theta\,|\,\mathbf{A}_r,\sigma)$ corresponds, for $1 \leq r < n$, to a normal distribution with mean θ_0 and covariance matrix $\sigma^2\mathbf{V}_0$, where

$$\mathbf{V}_0 = \begin{bmatrix} \mathbf{V}_{0\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{0\delta} \end{bmatrix}, \ \theta_0 = \begin{bmatrix} \beta_0 \\ \delta_0 \end{bmatrix}; \tag{17}$$

in the case of M_0, we simply have $\mathbf{V}_0 = \mathbf{V}_{0\beta}$, $\theta_0 = \beta_0$.

With this prior specification, it is easily verified that, performing the integration with respect to θ in (15), $p\,(\mathbf{y}\,|\,M_r,\sigma)$ is equal to

$$(2\pi\sigma^2)^{-n/2}\,|\mathbf{V}_0|^{-1/2}\,|\mathbf{V}_0^{-1} + \mathbf{A}_r^T A_r|^{-1/2}$$
$$\exp\{-(1/2\sigma^2)[R_r + (\hat{\theta}-\theta_0)^T(\mathbf{V}_0 + (\mathbf{A}_r^T\mathbf{A}_r)^{-1})^{-1}\,(\hat{\theta}-\theta_0)]\}, \tag{18}$$

where $\hat{\theta}$ denotes the usual least-squares estimate of θ, and R_r the corresponding residual sum of squares.

If $\mathbf{V}_0^{-1}$ may be considered *small* in relation to $\mathbf{A}_r^T\mathbf{A}_r$, (18) may be simplified somewhat to give

$$p\,(\mathbf{y}\,|\,M_r,\sigma) \approx (2\pi\sigma^2)^{-n/2}|\mathbf{V}_{0\beta}|^{-1/2}|\mathbf{V}_{0\delta}|^{-1/2}|\mathbf{A}_r^T\mathbf{A}_r|^{-1/2}\exp\{-R_r/2\sigma^2\}, \qquad (19)$$

and

$$p\,(\mathbf{y}\,|\,M_0,\sigma) \approx (2\pi\sigma^2)^{-n/2}|\mathbf{V}_{0\beta}|^{-1/2}|\mathbf{A}_0^T\mathbf{A}_0|^{-1/2}\exp\{-R_0/2\sigma^2\}. \qquad (20)$$

Noting that $|\mathbf{A}_r^T\mathbf{A}_r| = |\mathbf{X}_r^T\mathbf{X}_r|\;|\mathbf{X}_{(n-r)}^T\mathbf{X}_{(n-r)}|$, the Bayes factor for M_0 against M_r, conditioned on known σ, is seen to be,

$$B_{0r}^{(\sigma)} = \left(\frac{|\mathbf{V}_{0\delta}|\;|\mathbf{X}_r^T\mathbf{X}_r|\;|\mathbf{X}_{(n-r)}^T\mathbf{X}_{(n-r)}|}{|\mathbf{X}_n^T\mathbf{X}_n|}\right)^{1/2}\exp\{-(1/2\sigma^2)(R_0-R_r)\}. \qquad (21)$$

Integrating (19) and (20) with respect to the assumed form for $p\,(\sigma)$, we obtain the unconditional Bayes factor

$$B_{0r} = \left(\frac{|\mathbf{V}_{0\delta}|\;|\mathbf{X}_r^T\mathbf{X}_r|\;|\mathbf{X}_{(n-r)}^T\mathbf{X}_{(n-r)}|}{|\mathbf{X}_n^T\mathbf{X}_n|}\right)^{1/2}\left(1 + \frac{p}{(n-2p)}F_r\right)^{-n/2}, \qquad (22)$$

where $F_r = [(R_0-R_r)/p\,]/[R_r/(n-2p\,)]$ is the usual F-statistic for testing M_0 versus M_r.

In the special case of a univariate normal distribution with prior variance $\lambda\sigma^2$ for δ, the Bayes factor (22) reduces to

$$B_{0r} = \left(\frac{\lambda r(n-r)}{n}\right)^{1/2}\left(1 + \frac{t_r^2}{n-2}\right)^{-n/2}, \qquad (23)$$

where t_r is the two-sample t-test statistic corresponding to the samples $y_1,\ldots,y_r$ and $y_{r+1},\ldots,y_n$. The form (23) is similar to that derived for the two-sample problem by Jeffreys (1961, see comments following (13) of Section 5.41).

Application of (22) to the case of switching straight-lines has been made by Smith and Cook (1980). In this case, if δ_1, δ_2 are the components of δ

representing possible changes in intercept and slope, respectively, then the parameter of interest is often $\gamma = -\delta_1/\delta_2$, the intersection point between the two straight-lines. Two cases are possible, according as a change at r necessarily implies $x_r \leq \gamma < x_{r+1}$, or not, where $x_1 < x_2 < \ldots < x_n$ denote the (time) ordered x-values. In the unconstrained case, we need to calculate

$$p(\gamma \mid \mathbf{y}) = \Sigma_r p(\gamma \mid r,\mathbf{y}) p(r \mid \mathbf{y}),$$

the latter term being calculated using an appropriate transformation. In the constrained case, denoted by c, say, we require

$$p(\gamma \mid c,\mathbf{y}) = \Sigma_r p(c \mid \gamma,r,\mathbf{y}) p(\gamma,r \mid \mathbf{y}) / p(c \mid \mathbf{y}), \quad (24)$$

where

$$P(c \mid \gamma,r,\mathbf{y}) = \begin{cases} 1 & \text{if } \gamma \in (x_r, x_{r+1}) \\ 0 & \phantom{\text{if }} \gamma \notin (x_r, x_{r+1}) \end{cases}$$

and

$$p(c \mid \mathbf{y}) = \Sigma_r \{\int p(c \mid \gamma,r,\mathbf{y}) p(\gamma \mid r,\mathbf{y})\} p(r \mid \mathbf{y}) = \Sigma_r \int_{x_r}^{x_{r+1}} p(\gamma,r \mid \mathbf{y}) d\gamma.$$

Similarly, we can obtain

$$p(r \mid c,\mathbf{y}) = \int_{x_r}^{x_{r+1}} p(\gamma,r \mid \mathbf{y}) d\gamma / \Sigma_r \int_{x_r}^{x_{r+1}} p(\gamma,r \mid \mathbf{y}) d\gamma. \quad (25)$$

These results were applied in Smith and Cook (1980) to data from kidney transplant patients, with the object of inferring the time of rejection of transplanted kidneys. It is thought that the constrained switching straight-line model provides a good model of the behaviour of reciprocal body-weight corrected serum-creatinine over the days following a transplant. Table 5 summarizes the data from a particular patient and the result from (25) when large prior variances are attached to the straight-line parameters and all change points are equally likely. The posterior density for γ given by (24) is symmetric and sharply peaked, with a mode at 4.15 and an approximate 95% credible interval is given by (3.71,4.59).

TABLE 5

Renal transplant data and posterior probabilities for r

r	1	2	3	4	5	6	7	8
y_r	48.4	58.3	62.3	73.1	68.3	55.3	49.1	43.9
$p(r\mid c,y)$	—	.012	.316	.657	.012	.003	—	—

Retrospective studies of this kind are proving valuable in identifying patterns in the time to rejection of transplants and seem to have removed a great deal of the arbitrariness arising from doctors' attempts to "eyeball" the data. On-line analysis of this kind of data will be considered in Section 5.

Related material on switching straight lines can be found in Ferreira (1975).

4. SHIFT OF LEVEL IN AN ARMA PROCESS

In order to illustrate a reasonably straightforward extension of the approach of Section 3 to cover more general linear time series models, we shall consider the problem of investigating a shift in level of an ARMA (1,1) process. The material in this and the previous section is a direct development of some preliminary ideas given in Smith (1976).

We shall consider the following representation of a stationary ARMA (1,1) process with unknown mean level λ, and a shift in mean level of unknown magnitude δ occurring between the r^{th} and $(r+1)^{th}$ observations, where r is unknown. Let

$$\tilde{z}_1 = \lambda + \tilde{\varepsilon}_1$$

$$\tilde{z}_t = \lambda + \tilde{\varepsilon}_t + (\varrho-\phi)\,\Sigma_{s=1}^{t-1}\,\varrho^{s-1}\,\tilde{\varepsilon}_{t-s}, \quad t = 2,\ldots,r, \tag{26}$$

$$\tilde{z}_t = \lambda + \delta + \tilde{\varepsilon}_t + (\varrho-\phi)\,\Sigma_{s=1}^{t-1}\,\varrho^{s-1}\,\tilde{\varepsilon}_{t-s}, \quad t = r+1,\ldots,n,$$

with λ, δ, ϱ, ϕ and r unknown, $|\varrho|<1$, $|\phi| < 1$ and $\tilde{\varepsilon}_t$ independently and normally distributed with mean 0 and variance σ^2, with σ^2 unknown.

In order to utilize the development of Section 3, we make, conditional on ϕ and ϱ, the transformations

$$\tilde{y}_1 = \tilde{z}_1, \quad \tilde{y}_t = \tilde{z}_t - (\varrho-\phi)\,\Sigma_{s=1}^{t-1}\,\phi^{s-1}\,\tilde{z}_{t-s}, \quad t = 2,\ldots,n. \tag{27}$$

It is then easily seen that the vector $\mathbf{y}^T = (y_1,\ldots,y_n)$ satisfies (13), where, for $r \neq 0$,

$$\mathbf{A}_r^T = \mathbf{A}_r^T(\varrho,\phi) = \begin{pmatrix} a_1 \ldots a_r & a_{r+1} \ldots a_n \\ 0 \ldots 0 & a_1 \ldots a_{n-r} \end{pmatrix}, \theta = \begin{pmatrix} \lambda \\ \delta \end{pmatrix} \quad , \tag{28}$$

with $a_1 = 1$, $a_t = 1 - (\varrho - \phi)\ \Sigma_{s=1}^{t-1}\phi^{s-1}$, $t = 2,\ldots,n$. If $r = 0$, the "no-change" model, then $\theta = \lambda$ and $\mathbf{A}_0^T$ consists of just the first row of the matrix in (28).

By considering appropriate limits corresponding to $\varrho = 1$, $\phi = 0$ and $\varrho = 0$, respectively, the above framework can be used to study the special cases of IMA(1), AR(1) and MA(1) models. Related material can be found in Box and Tiao (1965) and Smith (1976).

Noting that the Jacobian of the transformation from $\mathbf{z}$ to $\mathbf{y}$ is unity, and denoting by $p\ (M_r,\varrho,\phi)$ a prior specification for M_r, ϱ and ϕ, we see from the results of Section 3 that $p\ (M_r,\varrho,\phi\,|\,\mathbf{z})$ is proportional to

$$(V_\lambda V_\delta)^{1/2}\,|\,\mathbf{A}_r^T(\varrho,\phi)\mathbf{A}_r(\varrho,\phi)\,|^{-1/2}\,(R_r(\varrho,\phi))^{-n/2}\,p\ (M_r,\varrho,\phi), \tag{29}$$

where $V_\lambda\sigma^2$, $V_\delta\sigma^2$ are the prior variances (conditional on σ) for λ and δ, and $R_r(\varrho,\phi)$ denotes the residual sum of squares from a least squares fit of M_r, given ϱ and ϕ.

The matrix whose determinant is to be evaluated in (29) has elements $a_1^2 + \ldots + a_n^2$ and $a_1^2 + \ldots + a_{n-r}^2$ on the diagonal, and $a_1 a_{r+1} + \ldots + a_{n-r}\,a_n$ as off-diagonal entries. The determinant and inverse are thus easily calculated.

Assignment of the prior probabilities for M_r, ϱ and ϕ depends, of course, on the situation under study. In any case, it seems that perfectly adequate results can be obtained by the crude form of numerical integration resulting from a suitable discretization of the ranges of ϱ and ϕ, so that calculation of marginal posterior probabilities are simply obtained from (29) by summation over the remaining variables. Inferences about λ, δ or σ^2, or predictive distributions for future observations, are obtained by forming weighted averages, with weights given by $p\ (M_r\,|\,\mathbf{z})$, of the standard results obtained by conditioning on a particular model M_r.

The procedure outlined above has been applied to a series of daily measurements of the time (in seconds) taken by an individual performing a certain psychological test repeated on 33 successive days. The data are presented in Table 6.

TABLE 6

Psychological test data

4.09	3.52	3.72	4.43	3.97	3.85	3.65	3.31	3.55	3.47	4.32
3.77	3.77	3.90	4.05	3.97	3.64	4.28	3.83	3.91	3.44	3.77
3.40	3.29	3.21	2.95	3.13	2.97	3.25	2.95	4.18	3.65	3.03

The individual has already passed through a ''learning'' phase on this test and it is believed that the observations would follow a stationary process, except that during this period of 33 days there has been a switch in background treatment regime. It is thought that this could have the effect of causing a sudden shift in performance level. The data were originally given to us with no information about where the change in treatment regime occurred. In fact, the change occurred between the 20th and 21st days.

Preliminary exploration of similar, unchanged, sequences of observations suggested that either an ARMA(1,1) or an AR(1) model might be suitable, and two corresponding analyses of the data were made. The first analysis assumed an ARMA (1,1) model with uniform priors over the range of r, the range of ϱ from -0.95 to 0.95 and ϕ between 0.000 and 0.95, the latter two in steps of 0.05. The second analysis considered an AR(1) model with a uniform prior for ϱ over the range -0.95 to 0.95. A summary of the results obtained are given in Table 7. No specification for V_λ is required, and V_δ is taken equal to 3.

TABLE 7

Summary inferences from the psychological test data

Posterior Summary	ARMA(1,1)	AR(1)
mean	21	21
r mode	22	22
median	22	22
ϱ mode	-	0.17
ϱ joint	0.17	-
ϕ mode	0.00	-
λ mean	3.83	3.83
δ mean	-0.53	-0.52

5. ON-LINE DETECTION OF CHANGE

Detailed description of the use of a set of alternative Kalman filter models has been given by Harrison and Stevens (1976) in the context of adaptive Bayesian forecasting procedures, and by Smith and Makov (1980) in the context of jump detection and estimation in linear systems, as required, for example, in the tracking of manoeuvering targets.

A general formulation allowing for sudden perturbations in either or both of the system and observation equations is given by representing model M_i, at time t, by

$$\theta_t = \mathbf{G}_{t-1}\theta_{t-1} + \mathbf{B}^{(i)}(\triangle\theta)_t + \mathbf{H}_{t-1}(\delta\theta)_t, \tag{30}$$

$$\mathbf{y}_t = \mathbf{F}_t\theta_t + \mathbf{C}^{(i)}(\triangle\mathbf{y})_t + (\delta\mathbf{y})_t, \tag{31}$$

where $(\triangle\theta)_t$, $(\triangle\mathbf{y})_t$ represent possible abrupt changes in either the system or the measurement at time t, $\mathbf{B}^{(i)}$, $\mathbf{C}^{(i)}$ define the specific nature of these changes according to model M_i, and $(\delta\theta)_t$, $(\delta\mathbf{y})_t$ are the usual Gaussian "noise" inputs to the system and measurement equations. The matrices $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$ define the general characteristics of the system.

In the case of manoeuvering targets, θ_t represents position and velocity components in some chosen frame of reference and $\mathbf{y}_t$ usually consists of observed position components. If $(\triangle\theta)_t$ consists of a finite set of plausible manoeuvres available at time t, models corresponding to particular choices of manoeuvre are defined by appropriate choices of $\mathbf{B}^{(i)}$ (assuming here that $\mathbf{C}^{(i)} = \mathbf{0}$).

In the case of the very useful univariate Linear Growth model (Harrison and Stevens, 1976, 3.4), the case of no abrupt change is modelled by

$$y_t = \mu_t + (\delta y)_t$$
$$\mu_t = \mu_{t-1} + \beta_t + (\delta\mu)_t$$
$$\beta_t = \beta_{t-1} + (\delta\beta)_t,$$

which can be represented in terms of (30) and (31) by

$$\begin{pmatrix}\mu_t\\ \beta_t\end{pmatrix} = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1}\\ \beta_{t-1}\end{pmatrix} + \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}\begin{pmatrix}(\delta\mu)_t\\ (\delta\beta)_t\end{pmatrix}$$

$$y_t = (1 \quad 0)\begin{pmatrix}\mu_t\\ \beta_t\end{pmatrix} + (\delta y)_t$$

with $\mathbf{B}^{(i)} = \mathbf{C}^{(i)} = \mathbf{0}$. If we define this *no change* model to be M_0, and define M_1, M_2, M_3, by

$$\mathbf{B}^{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B}^{(2)} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{B}^{(3)} = \mathbf{0},$$

$$\mathbf{C}^{(1)} = 0, \quad \mathbf{C}^{(2)} = 0, \quad \mathbf{C}^{(3)} = 1,$$

we can represent "sudden change in level", "sudden change in slope" and "outlying observation", respectively, as well as "no change".

The recursive updating of the system, given a choice of M_i at time t, proceeds straightforwardly using the standard Kalman filter equations. Posterior weights on the individual models are also easily obtained using the appropriate modification of (3). In fact, of course, there is the problem of expanding mixture forms of posterior distribution, resulting from the unsupervised learning context, and practical use of this approach requires approximation of this mixture, at each stage, by a simple Gaussian distribution having the same mean and covariance structure as the mixture: see Harrison and Stevens (1976, 5.4) and Smith and Makov (1980) for further details.

This Linear Growth model, with the four model variants outlined above, has been used for on-line monitoring of kidney transplant patients, given data of the type shown in Table 5. For many patients, the series is considerably longer than the one shown, but we shall illustrate our procedure with this small data set. Table 8 shows, for each of the first six observations, the probability that it came from the situation modelled by M_0, M_1, M_2 or M_3. In addition, the table shows the same probabilities one-step back and two-steps back: thus, for example, $p\,(\tilde{y}_5 \epsilon M_2 | y_1,\ldots,y_7) = 0{\cdot}68$. By studying the changing pattern of these probabilities, the doctor can, hopefully, react to genuine changes fairly quickly, whilst avoiding over-hasty reactions to outlying measurements. Of course, the system depends on a number of prior inputs regarding reasonable variance levels and other features. These are assessed from knowledge of serum-creatinine measurement procedures and other background physiological information. Full details of this and other case studies will be reported elsewhere. The prior probabilities set on the four models for the first observation in this case were: 0.96, 0.01, 0.01, 0.02.

The results indicate that at observation 6 we suspect a slope change has occurred at observation 5. When we reach observation 7, we are fairly convinced that a slope change has occurred and that the patient is now in a new *steady state*. Posterior means of the slope parameter are positive up to and including observation 5 and then they suddenly switch to negative values, reinforcing the message of Table 8.

TABLE 8

On-line probabilities of M_0, M_1, M_2, M_3

Observation

	1				2				3			
	M_0	M_1	M_2	M_3	M_0	M_1	M_2	M_3	M_0	M_1	M_2	M_3
0-back	.99	-	-	-	.99	-	-	-	.99	-	-	-
1-back	.99	-	-	-	.99	-	-	-	.99	-	-	-
2-back	.99	-	-	-	.99	-	-	-	.99	-	-	-

	4				5				6			
	M_0	M_1	M_2	M_3	M_0	M_1	M_2	M_3	M_0	M_1	M_2	M_3
0-back	.99	-	-	-	.96	-	.01	-	.64	.09	.09	.17
1-back	.98	-	-	-	.56	.02	.41	.01	.84	.05	.10	-
2-back	.98	-	-	-	.29	.02	.68	-	.85	.04	.10	-

ACKNOWLEDGEMENTS

Study of kidney transplantation is in collaboration with Dr. M. Knapp of the Renal Unit, City Hospital, Nottingham. Study of schizophrenic test data is in collaboration with Dr. A. Elithorn of the Royal Free Hospital, London. The material of Section 4 has been developed in a Ph.D. thesis by N.B. Booth of the North London Polytechnic. The computations reported in Section 5 were carried out by M. West of Nottingham University.

REFERENCES

BOX, G.E.P. and TIAO, G.C. (1965). A change in level of a non-stationary time series. *Biometrika* **52**, 181-92.

DICKEY, J.M. and LIENTZ, B.P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *Ann. Math. Statist.* **41**, 214-26.

FERREIRA, P.E. (1975). A Bayesian analysis of a switching regression model. *J. Amer. Statist. Assoc.* **70**, 370-74

HARRISON, P.J. and STEVENS, C.F. (1976). Bayesian Forecasting (with discussion). *J. Roy. Statist. Soc. B* **38**, 205-47.

JEFFREYS, H. (1961). *Theory of Probability*. Oxford: University Press.

ROSS, A.S.C. (1950). Philological Probability Problems. *J. Roy. Statist. Soc. B* **12**, 19-40.

SILVEY, S.D. (1958). The Lindisfarne Scribes Problem. *J. Roy. Statist. Soc. B* **20**, 93-101.

SMITH, A.F.M. (1975). A Bayesian approach to inference about a change point in a sequence of random variables. *Biometrika* **63**, 407-16.

— (1976). A Bayesian analysis of some time-varying models; In *Recent Developments in Statistics* (ed. Barra *et al*). 257-67. Amsterdam: North-Holland.

— (1977). A Bayesian note on reliability growth during a development testing program. *IEEE Trans. on Reliability* **R-26**, 346-47.

SMITH, A.F.M. and COOK, D.G. (1980). Switching straight lines: a Bayesian analysis of some renal transplant data. *Appl. Statist.* **29**, 180-89.

SMITH, A.F.M. and MAKOV, U.E. (1980). Bayesian detection and estimation of jumps in linear systems. *Proceedings of the IMA Conference on "The Analysis and Optimization of Stochastic Systems"*, (O.R.L., Jacobs *et. al.* ed.) 333-346. London: Academic Press.

SMITH, A.F.M. and SPIEGELHALTER, D.J. (1980) Bayes factors and choice criteria for linear models. *J. Roy. Statistic. Soc* **B. 42**, 213-20.