KALMAN FILTERING ALGORITHM IN PRESENCE OF OUTLIERS

V.I. LOBACH
Belarusian State University
Minsk, BELARUS
e-mail: lobach@bsu.by

Abstract

A Kalman Filtering algorithm which is robust to observational outliers is developed by assuming that the measurement error may come from either one of two normal distributions and that transition between these distribution is governed by a Markov Chain. The state estimate is obtained as a weighted average of the estimates from the two parallel filters where the weights are the posterior probabilities. The impotents obtained by this Robust Kalman Filter in the presence of outliers is demonstrated with examples.

1 Introduction

The Kalman Filter is well known recursive estimator for the state of a linear system and has been used in the fields of forecasting and control. It has been derived as a least squares estimator, and also, under the assumption of normality as a Bayesian estimate. However, as with most least squares estimators, it is very sensitive to observational outliers. This sensitivity to outliers is a major draw back of the filter. Bad observations arising from periodic sensor problems can seriously bias the filter estimates for a considerable period of time thereafter.

In this paper a Bayesian approach is used to derive a recursive state estimator when occasional spurious observations will arise. The new state estimator is shown to have the structure of two parallel Kalman Filters in which the final state estimate is a weighted average of the estimators from the two filters.

2 The State Variable Model

Consider the state variable model given by

$$x_t = Fx_{t-1} + G\alpha_t, \tag{2.1}$$

$$z_t = Hx_t + \varepsilon_t, \tag{2.2}$$

where x_t is an $(n \times 1)$ vector of state variables defining the dynamic behavior of a system, F, G are known matrices (possibly time varying) and H is a known row vector of the appropriate dimensions. The measurement z_t taken at time t is related linearly to the states through the measurement equation (2.2). The state and observation noise sequences $\{\alpha_t\}$ and $\{\varepsilon_t\}$ are usually assumed to be i.i.d. Gaussian random variables with known variances σ_α^2 and σ_ε^2 respectively. Under the assumption that the system is observable the Kalman Filter provides a set of recursive equations for the one-step ahead predictions of the state $\hat{x}_{t+1|t}$ and the filtering state estimates $\hat{x}_{t|t}$ given information up to and including time t. However, the optimality of the estimator depends upon knowledge of the variances σ_α^2 and σ_ε^2 at each point in time. The occurrence of infrequent outliers can be characterized by an increase in the variance of α_t (an innovational outlier) which leads to a real change in the state of the system or in the variance of ε_t (an observational outlier) which has no effects on the true state (x_t), but will have a pronounced effect on the Kalman Filter estimates. In the following section we use a Bayesian approach to develop a robust filter to handle observational outliers.

3 Derivation of the Robust Filter

We characterize the observational outliers by replacing ε_t in equation (2.2) by $(1 - -i_t)\varepsilon_t + i_t\gamma_t$, where $\{i_t, t = 0, 1, \ldots\}$ is a sequence of $\{0,1\}$ random variables following a Markov Chain (MC). Suppose that the initial probability (π_0, π_1) of MC is known and that the transition probability matrix is $\{q_{i_{t-1},i_t}\}$. γ_t is a Gaussian random variable with zero mean and variance σ_γ^2 which is very large compared with σ_ε^2. We also assume that the conditional density $p(x_{t-1}|z_0^{t-1})$ is a Gaussian distribution with mean $\hat{x}_{t-1}$ and covariance matrix P_{t-1}. Then

$$p(x_t, i_t|z_0^t) = \sum_{i_{t-1}=0}^{1} \int p(x_{t-1}, i_{t-1}|z_0^{t-1}) \cdot p(x_t, i_t|x_{t-1}, i_{t-1}, z_0^{t-1}) dx_{t-1},$$

which on further simplification yields [1], [2]:

$$p(x_t, i_t|z_0^{t-1}) = (2\pi)^{-\frac{n}{2}} \sum_{i_{t-1}=0}^{1} q_{i_{t-1}i_t} k_{i_{t-1}} \left|P_{t|t-1}\right|^{-\frac{1}{2}} \times$$
$$\times \exp\left\{-\frac{1}{2}(x_t - \hat{x}_{t|t-1})^T P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1})\right\},$$

where $q_{i_{t-1}i_t} = P\left\{i_t|i_{t-1}, z_0^{t-1}\right\}$, $P_{t|t-1} = FP_{t-1}F^T + GG^T\sigma_\alpha^2$, $\hat{x}_{t|t-1} = F\hat{x}_{t-1}$, $k_{i_{t-1}} = P\left\{i_{t-1}|z_0^{t-1}\right\}$.

By Bayes theorem

$$p\left(x_t, i_t|z_0^t\right) = \frac{p(x_t, i_t|z_0^{t-1}) \cdot p(z_t|x_t, i_t, z_0^{t-1})}{\sum_{i_t=0}^{1} \int p(x_t, i_t|z_0^{t-1}) p(z_t|x_t, i_t, z_0^{t-1}) dx_t} =$$
$$= C \cdot (2\pi)^{-\frac{n}{2}} \left(\sum_{i_{t-1}=0}^{1} q_{i_{t-1},i_t} \cdot k_{i_{t-1}}\right) \left|P_{t|t-1}\right|^{-\frac{1}{2}} \times$$
$$\times \exp\left\{-\frac{1}{2}(x_t - \hat{x}_{t|t-1})^T P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1})\right\} \times$$
$$\times (\sqrt{2\pi}\sigma_{i_t})^{-1} \exp\left\{-\frac{1}{2}(z_t - Hx_t)^T \sigma_{i_t}^{-2}(z_t - Hx_t)\right\},$$

where

$$\sigma_{\imath_t}^2 = (1-\imath_t)\sigma_\varepsilon^2 + \imath_t \sigma_\gamma^2,$$

$$C^{-1} = \sum_{\imath_t=0}^{1} \int p(x_t, \imath_t | z_0^{t-1}) \cdot p(z_t | x_t, \imath_t, z_0^{t-1}) dx_t.$$

This can be simplified

$$p(x_t, \imath_t | z_0^t) = k_{\imath_t} (2\pi)^{-\frac{n}{2}} \left| P_{t|t}^{(\imath_t)} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left(x_t - \hat{x}_{t|t}^{(\imath_t)} \right) \cdot \left(P_{t|t}^{(\imath_t)} \right)^{-1} \cdot \left(x_t - \hat{x}_{t|t}^{(\imath_t)} \right) \right\},$$

where

$$\hat{x}_{t|t}^{(\imath_t)} = \hat{x}_{t|t-1} + R_{\imath_t}^{-1} P_{t|t-1} H^T (z_t - H\hat{x}_{t|t-1}), \tag{3.1}$$

$$P_{t|t}^{(\imath_t)} = P_{t|t-1} - R_{\imath_t}^{-1} P_{t|t-1} H^T H P_{t|t-1}, \tag{3.2}$$

$$R_{\imath_t} = \sigma_{\imath_t}^2 + H P_{t|t-1} H^T \tag{3.3}$$

and

$$\begin{aligned} k_{\imath_t} = C \cdot (\sqrt{2\pi}\sigma_{\imath t})^{-1} \left| P_{t|t}^{(\imath_t)} \right|^{\frac{1}{2}} \left(\sum_{\imath_{t-1}=0}^{1} q_{\imath_{t-1}\imath_t} k_{\imath_{t-1}} \right) \left| P_{t|t-1} \right|^{-\frac{1}{2}} \times \\ \times \exp\left\{ -\frac{1}{2} (z_t - H\hat{x}_{t|t-1})^T R_{\imath_t}^{-1} (z_t - H\hat{x}_{t|t-1}) \right\}. \end{aligned} \tag{3.4}$$

The $k_{\imath_t}$ are the posterior probabilities of the current observation at time t being an outlier ($i_t = 1$) or not being outlier ($\imath_t = 0$). C is a normalizing constant. Note that the term $\sum_{\imath_{t-1}=0}^{1} q_{\imath_{t-1}\imath_t} k_{\imath_{t-1}} = P_{\imath t_t}$ is the prior probability of being in state i_t at time t.

Thus

$$p(x_t | z_0^t) = \sum_{\imath_t=0}^{1} k_{\imath_t} N\left(\hat{x}_{t|t}^{\imath_t}, P_{t|t}^{(\imath_t)} \right), \tag{3.5}$$

where $N\left(\hat{x}_{t|t}^{(\imath_t)}, P_{t|t}^{(\imath_t)}\right)$ denotes a normal distribution with mean $\hat{x}_{t|t}^{(\imath_t)}$ and covariance matrix $P_{t|t}^{(\imath_t)}$. For computational simplicity we approximate the posterior density (3.5) by a single Gaussian distribution with matching moments. We write this posterior density as

$$p(x_t | z_0^t) = N(\hat{x}_t, P_t),$$

where $\hat{x}_t = \sum_{\imath_t=0}^{1} k_{\imath_t} \hat{x}_{t|t}^{(\imath_t)}$, $P_t = \sum_{\imath_t=0}^{1} k_{\imath_t} P_{t|t}^{(\imath_t)}$.

In the proposed robust filter (3.1) (3.4) the gain corresponding to i_t is given by

$$K_{\imath_t} = R_{\imath_t}^{-1} \hat{P}_{t|t-1} H^T,$$

where $R_{\imath_t}$ is given in (3.3). Thus the state estimate is computed as a linear combination of two parallel filters having a lower and a higher gain depending on the current observation z_t.

4 Simulation Examples

4.1 ARIMA(0,1,1) model with outliers

Consider the following special case of the system of equations given by (2.1), (2.2).

$$x_t = x_{t-1} + \alpha_t, \tag{4.1}$$

$$z_t = x_t + \varepsilon_t, \tag{4.2}$$

where $\{\alpha_t\}$ and $\{\varepsilon_t\}$ and are i.i.d. Gaussian random variables with mean zero and variances σ_α^2 and σ_ε^2 respectively. This is equivalent to an ARIMA(0,1,1) model:

$$\nabla z_t = (1 - \theta B) b_t \tag{4.3}$$

for $\theta > 0$, where θ is the moving average (MA) parameter, is a backshift operator such that $Bz_t = z_{t-1}$ and b_t is an i.i.d. sequence of Gaussian random variables with mean zero and variance σ_b^2. The relationships among the parameters in the state model form (4.1), (4.2) and difference equation (4.3) are given in [3] as $(1-\theta^2)/\theta = \sigma_\alpha^2/\sigma_\varepsilon^2$, $\sigma_b^2 = = \sigma_\alpha^2/(1-\theta)^2$. The procedure in the previous section is illustrated with time series (4.3). The subsequent 100 observations where then considered as the sample period and outliers were introduced at the 25th, 50th, 65th and 75th observations by adding $N(0, 1.75)$ random variables to them. Initially the ordinary Kalman filter (with no outlier protection) was run with $\hat{x}_0 = 17$ (the first observation in the sample period). $\hat{P}_0 = 0.009$, $\hat{\sigma}_\alpha^2 = 0.009$ and $\hat{\sigma}_\varepsilon^2 = 0.071$. The modeling results show that the ordinary Kalman filter estimates are sensitive to spurious observations. For example the predictions after spurious observation at $t = 50$ remain too high for a considerable period of time before aligning with the data again. A similar behavior can be noticed at time points subsequent to the other discrepant observation.

The robust filter was run assuming $\hat{x}_0 = 17$, $\hat{P}_0 = 0.009$, $\hat{\sigma}_\varepsilon^2 = 0.071$, $\hat{\sigma}_\gamma^2 = 1.75$, $\pi_0 = 0.9$, $\pi_1 = 0.1$. The modeling results show that this filter is robust to observational outliers. For example, the prediction at $t = 50$ and at the subsequent points do not seem to be affected by the aberrant observation at $t = 50$ un like in the previous case.

In addition, the robust filter also provides the posterior probability R_1 of any observation being an outlier.

4.2 ARIMA(1,1,0) model with outliers

Consider the model

$$(1 - \varphi B)(1 - B) y_t = \alpha_t, \tag{4.4}$$

where φ is the autoregressive parameter and $\{\alpha_t\}$ is a white noise sequence with mean zero and variance σ_α^2. Let the measurement process be defined by

$$z_t = y_t + \varepsilon_t, \tag{4.5}$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d. Gaussian random variables with mean zero and variance σ_ε^2. For $\varphi > 0$ this model represents a very slowly drifting process. Predictions for any lead time obtained at a particular time point $t = \tau$ in this series is a linear function of z_τ and $z_{\tau-1}$. Thus, if there is an observational outlier at τ it strongly influences the predictions. Thus robust Kalman Filtering and outlier detection are quite relevant in this case.

We shall now incorporate occasional outliers by replacing ε_t by $(1 - i_t)\varepsilon_t + i_t\gamma_t$, where $\{\gamma_t\}$ and $\{i_t\}$ are as defined before. A state space equivalent of this ARIMA(1,1,0) observational outlier model can be written as

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & \varphi \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \alpha_t, \tag{4.6}$$

$$z_t = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} + (1 - i_t)\varepsilon_t + i_t\gamma_t,$$

where x_t is a (2×1) state vector. Then the state and covariance predictions at time t can be given by

$$\hat{x}_{t|t-1} = \begin{pmatrix} 1 & 0 \\ 1 & \varphi \end{pmatrix} \begin{pmatrix} \hat{x}_{1t-1} \\ \hat{x}_{2t-1} \end{pmatrix} = (\hat{x}_{1t-1}, \hat{x}_{1t-1} + \varphi\hat{x}_{2t-1})^T,$$

$$\hat{P}_{t|t-1} = \begin{pmatrix} 1 & 0 \\ 1 & \varphi \end{pmatrix} \hat{P}_{t-1} \begin{pmatrix} 1 & 1 \\ 0 & \varphi \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \sigma_\alpha^2,$$

$$\hat{z}_{t|t-1} = \begin{pmatrix} 0 & 1 \end{pmatrix} \hat{x}_{t|t-1} = \hat{x}_{1t-1} + \varphi\hat{x}_{2t-1}.$$

A simulation study was conducted using 100 observations from the model (4.4) with $\varphi = 0.8$, $\sigma_\alpha = 1$, $\sigma_\varepsilon = 5$. As in the previous example the outliers were introduced at the points $t = 25, 50, 65, 75$ using a Gaussian distribution with $\sigma_\gamma = 25$. The initial values of the states which were used to generate the series were $\hat{x}_{10} = 20$ and $\hat{x}_{20} = 150$ respectively. The ordinary Kalman filter was initially used to estimate the state and to make predictions. The aberrant observations seem to have a strong influence on the predictions. Secondly, the robust Kalman filter was run on the assumption that $P\{i_t = 1\} = 0.1$. Unlike before, the predictions are not influenced in this case. Again the four large outliers have been clearly detected.

References

[1] Jazwinski A.H. (1970). *Stochastic Processes and Filtering Theory.* Academic Press, N.J.

[2] Anderson B., Moore J. (1979). *Optimal Filtering.* APrentice-Hall, N.J.

[3] Box G., Jenkins G. (1976). *Time Series Analysis Forecasting and Control.* Holden day, San Francisco.

[4] Priestly M. (1980) State-dependent models: A General Approach to Time Series Analysis. *J. Time Series Analysis*, No 1, pp. 47-71.

[5] Guttman J., Pena D. (1985)Comment on robust Filtering. *Journal of the American Statistical Association* , Vol. 80, pp. 91-92.