STATISTICAL APPLICATIONS OF KALMAN FILTER: UPDATING A FUNCTION FOR SEQUENTIALLY OBTAINED OBSERVATIONS

by

Walter T. Federer

ABSTRACT

An application of Kalman filtering to update the estimated parameters as more observations are obtained is presented. The computational procedures are discussed in detail, and a computer program written in GAUSS for updating a function is given. The consequences of using bad guesses or assumptions for values of the parameters to start the procedure are discussed. Examples are utilized to illustrate the effects from using various values of the parameters. Suggestions are made for modifying original values of the parameters as the process is continued. The relationship of Kalman filtering to mixed model and Bayesian analyses is discussed.

Keywords: Regression, GAUSS, mixed model, prior distribution, posterior distribution.

TR-1571

~~BU--M~~ January 2001

In the Technical Report Series of the Department of Biometrics, Cornell University, Ithaca, New York 14853

INTRODUCTION

Following the Meinhold and Singpurwalla (1983) presentation and notation, let the observations, the data, be denoted by $Y_t, Y_{t-1}, \ldots, Y_1$, where the Y_t is the observation at time t and may be a scalar or vector. Let the parameter of interest be θ_t and the relationship between Y_t and θ_t be the observatioanal or response equation

$$Y_t = F_t\, \theta_t + v_t \qquad (1)$$

where F_t is a known design function and v_t are normally distributed with mean zero and variance V_t. Since the various parametric values may change over time, this is incorporated into the process by what is called the system equation, which is

$$\theta_t = G_t\, \theta_{t-1} + w_t \qquad (2)$$

where G_t is known and w_t is normally distributed with mean zero and variance W_t. The v_t and w_t may or may not be independent. Note the relationship to a Bayesian and mixedmodel presentation. The prior distribution is taken to be

$(\theta_t \mid Y_{t-1})$ which is distributed as

$$N(G_t\, \widehat{\theta}_{t-1}, R_t = G_t\, \Sigma_{t-1}\, G_t' + W_t) \qquad (3)$$

The posterior distribution is

$(\theta_t \mid e_t, Y_{t-1})$ which is distributed as

$$N[G_t\, \widehat{\theta}_{t-1} + R_t\, F_t'(V_t + F_t\, R_t\, F_t')^{-1} e_t,\;\; R_t - R_t\, F_t'(V_t + F_t\, R_t\, F_t')^{-1} F_t\, R_t]$$
$$= N[\,\widehat{\theta}_t, \Sigma_t\,], \qquad (4)$$

where $e_t = Y_t - F_t\, G_t\, \widehat{\theta}_{t-1}$

Meinhold and Singpurwalla (1983) give the following interpretation of this updating process. The mean of the posterior distribution of $(\theta_t \mid e_t, Y_{t-1})$ is the regression of θ_t on e_t. The mean (regression function) is the sum of the two items $G_t\, \theta_{t-1}$ and a multiple of a one step ahead forecast of error e_t. $G_t\, \theta_{t-1}$ is the mean of the prior distribution of θ_t and the multiplier of e_t, $R_t\, F_t'(V_t + F_t\, R_t\, F_t')^{-1}$ is the least squares regression of θ_t on e_t conditional on Y_{t-1}. We may view Kalman filtering as an updating procedure that consists of forming a prior or preliminary guess about the state of nature and then adding a correction to this guess. The correction is determined by how well the guess has performed in predicting the next observation.

The regression relationship is not estimated in the standard way since the pair θ_t and e_t constitutes a single observation. In sequential Bayesian estimation a new posterior distribution arises each time a new observation is obtained and this is what is happening here. At time zero, the regression of θ_1 on e_1 is mapped into $\widehat{\theta}_1$ through this regression function. This is then replaced by a new regression relationship based on θ_1, F_1, G_1, V_1, and W_1. This in turn is used to map e_2 into θ_2. The process continues in the usual Bayesian prior/posterior iterative manner. Kalman filtering can be viewed as the evolution of a series of regression functions of θ_t on e_t at times 0, 1, 2, ..., t - 1, t, each having a potentially different intercept and a different regression coeficient. The evolution stems from a learning process involving all of the data.

The updating procedure described above was the motivation for the original development of Kalman filter (Kalman, 1960; Kalman and Bucy, 1961). Its derivation followed least squares estimation theory. The Bayesian formulation described above yields the same result in an "elegant" manner and provides inference about θ_t through a probability distribution rather than only a point estimate.

PROCEDURE

In order to update a function as new observations, Y_t, are obtained, the following steps are used:

Step 1 - V_t, W_t, G_t, Y_t, and F_t are known or assumed. θ_0 and Σ_0 are guessed.

Step 2 - Compute $R_t = G_t \Sigma_{t-1} G_t' + W_t$ (matrix form)

$$= G_t^2 \Sigma_{t-1} + W_t \text{ (scalar form).} \tag{5}$$

Step 3 - Compute $\widehat{\theta}_t$ as $\widehat{\theta}_t = G_t \widehat{\theta}_{t-1} + R_tF_t'(V_t + F_t R_t F_t')^{-1}(Y_t - F_tG_t\widehat{\theta}_{t-1})$ (matrix form)

$$= G_t \widehat{\theta}_{t-1} + R_tF_t(Y_t - FtG_t\widehat{\theta}_{t-1})/ (V_t + R_tF_t^2) \text{ (scalar form)} \tag{6}$$

Step 4 - Compute Σ_t as $\Sigma_t = R_t - R_t F_t'(V_t + F_tR_tF_t')^{-1}F_tR_t$ (matrix form)

$$= R_t - R_t^2F_t^2/(V_t + R_tF_t^2) \text{ (scalar form).} \tag{7}$$

The Procedure Applied to an Example

Meinhold and Singpurwalla (1983) constructed their example in the following manner. They set $G_t = (-1)^t/2$, $W_t = 1$, and $V_t = 2$. They give values for F_t with no explanation of why such values were selected. Ordinarily, f_t would be the ordinary X_t values in regression, but why these f_t values? Their initial value for θ_0 was -0.353. Then a random normal deviate was selected from N(0, 2) and another random deviate from N(0, 1). These were v_1 = -0.376 and w_1 = 0.887. The value for f_1 is set equal to 1.3. Then, they calculated θ_1 and Y_1 as

$$\theta_1 = (-1)/2\ (-0.353) + 0.887 = 1.0635$$

and

$$Y_1 = f_1\ \theta_1 + w_1 = 1.3(1.0635) - 0.376 = 1.007\ .$$

For the second set of values, v_2 = 0.023, w_2 = -1.021, and f_2 = 0.8. Then θ_2 and Y_2 are

$$\theta_2 = ((-1)^2/2)(1.0635) - 1.021 = -0.489$$

and

$$Y_2 = 0.8(-0.489) + 0.023 = -0.368.$$

Continuing this procedure, they construct 25 Y_t values. They give no explanation as to why the Y_t values were obtained in this manner rather than from a real world data set. Presumably it was to demonstrate some characteristic of this Kalman filter updating process. Likewise, no explanation is given for using these particular f_t values other than "F_t is in the nature of the familiar variable of ordinary regression". They do not explain if there is or is not a relationship of Σ_t to W_t and V_t. They give a graph showing the corresponding 25 pairs of values for θ_t and $\widehat{\theta}_t$. There are large discrepancies between these values for the zeroth, the 20th, and the 25th pairs. No discussion is given other than θ_0 was a "bad guess". The mean of the 25 Y_t is -0.262, the mean of the 25 θ_t is 0.008, the mean of the 25 $\widehat{\theta}_t$ is -0.124, and the mean of the 25 Σ_t is 0.757.

The following illustrates the calculations involved in Steps 1 to 4. The values taken in Step 1 as the starting point are:

Step 1 - $V_t = 2$, $W_t = 1$, $\widehat{\theta}_0 = 4.183$ (This turns out to be a bad guess), $\Sigma_t = 1$, $G_t = (-1)^t/2$ which incorporates a cyclical behavior into θ_t, and the first pair of observations is $Y_1 = 1.007$ and $f_1 = 1.3$ as computed above.

Then the calculations for Steps 2 to 4 are:

Step 2 - $R_1 = (-1)/2(1)\,(-1)/2 + 1 = 1.25.$

Step 3 - $\widehat{\theta}_1 = (-1)(4.183)/2 + \{1.25(1.3)[1.007 - (-1/2)(1.3)(4.183)]\}/[2 + 1.25\,(1.3)^2]$
$= -2.0915 + 6.0547 / 4.1125 = -0.619.$

Step 4 - $\Sigma_1 = 1.25 - 1.25^2(1.3^2)/[2 + 1.25(1.3^2)] = 0.608.$

Let the second pair of values be $Y_2 = -0.368$ and $f_2 = 0.8$. Then we update the process as follows:

Step 2 - $R_2 = (1/4)(0.608) + 1 = 1.152.$

Step 3 - $\widehat{\theta}_2 = [(-1)^2/2](-0.619) + 1.152(0.8)[-0.368 - (1/2)(0.8)(-0.619)]/[2 + 0.8(1.152)]$
$= -0.3095 + 0.9216[-0.1204]/2.73 = -0.350.$

Step 4 - $\Sigma_2 = 1.152 - 1.1522(0.82) / [2 + 1.152(0.82)] = 0.842.$

Let the third pair of values obtained be $Y_3 = -1.764$ and $f_3 = 0.9$. Then the function is updated as follows:

Step 2 - $R_3 = (1/4)(0.842) + 1 = 1.2105$.

Step 3 - $\widehat{\theta}_3 = [(-1)^3/2](-0.350) + 0.9(1.2105)[-1.764 - 0.9(-0.350)/2]/[2 + 0.81(1.2105)]$
$= 0.175 - 0.7024 = -0.527$.

Step 4 - $\Sigma_3 = 1.2105 - 0.81(1.2105^2)/[\,2 + 0.81(1.2105)] = 1.2105 - 0.3980 = 0.812$.

For the fourth pair values, $Y_4 = 1.281$ and $f_4 = 1.1$. Then the updating of the function follows:

Step 2- $R_4 = (1/4)(0.812) + 1 = 1.203$.

Step 3 - $\widehat{\theta}_4 = [(-1)^4/2](-0.527) + 1.1(1.203)[1.281 - 1.1(-0.527)/2]/[2 + 1.21(1.203)]$
$= -0.2635 + 0.6016 = 0.338$.

Step 4 - $\Sigma_4 = 1.203 - 1.21(1.203^2)/[\,2 + 1.21(1.203)] = 1.203 - 0.507 = 0.696$.

Continuing these calculations, the results in Table 1 of Meinhold and Singpurwalla (1983). may be obtained for the 25 values for t.

A computer program for the calculations involved as written for GAUSS for the first set of observations is

```
let t = 1: let Y = 1.007; let F = 1.3; let G = -1 / 2;
let S = 1; let P = 4.183; let W = 1; let V = 2; t;
format(4,4);
R = 0.25*S + W; R;                                  ( R = 1.25)
PN = G*P + R*F*(Y - F*G*P) / (V + R*F*F); PN;       (PN =  -0.619)
SN = R - R*R*F*F / (V + R*F*F); SN;                 (SN = 0.608)
```

For ease of writing in GAUSS, SN replaces Σ_t and PN replaces $\widehat{\theta}_t$; the subscipt t is omitted.

For the second set of observations, change t to 2, change Y to -0.368, change F to 0.8, change G to 0.5, change S to 0.608, and change P to -0.619 to obtain R = 1.152, PN = -0.350, and SN = 0.842.

Effect of Different Values for W_t and V_t on the Results

In the above W_t and V_t were held constant throughout the computations. It is possible that the initial values were far away from where they should. To illustrate the effect of different values on the computations, consider the case where $W_t = 10$ and $V_t = 1$ and the case where $W_t = 1$ and $V_t = 10$. The resulting values for the first 15 values of the Meinhold and Singpurwalla (1983) example are presented in Table 1.

When $W_t = 10$ and $V_t = 1$, R_t values are considerably different than those for $W_t = 1$ and $V_t = 2$ as W_t is a large contributor to R_t. The P_t values for this case are lower than when $W_t = 1$ and $V_t = 2$. The P_{15} value for this case is -0.830 versus -0.324 for $W_t = 1$ and $V_t = 2$. The S_t values for this case average higher, 0.986 versus 0.753 and $S_{15} = 1.104$ versus 0.820 for $W_t = 1$ and $V_t = 2$.

When $W_t = 1$ and $V_t = 10$, $P_{15} = -0.119$ and $S_{15} = -1.185$ versus -0.324 and 0.820, respectively, for $W_t = 1$ and $V_t = 2$. The means for the 15 P_t and S_t are -0.205 and 1.150, respectively, versus -0.163 and 0.753 for $W_t = 1$ and $V_t = 2$. These examples demonstrate that bad assumptions about W_t and V_t have little effect on the solutions for P_t but do increase the values for S_t.

Table 1. Effect of varying values of W_t and V_t.

	$W_t = 1$ and $V_t = 2$					$W_t = 10$ and $V_t = 1$			$W_t = 1$ and $V_t = 10$		
t	Y_t	F_t	R_t	$\theta_t = P_t$	$\Sigma_t = S_t$	R_t	P_t	S_t	R_t	P_t	S_t
0				4.183	1		4.183	1		4.183	1
1	1.007	1.3	1.250	-0.619	0.608	10.25	0.618	0.559	1.250	-1.592	1.032
2	-0.368	0.8	1.152	-0.350	0.842	10.14	-0.357	1.354	1.258	-0.771	1.164
3	-1.764	0.9	1.210	-0.527	0.812	10.34	-1.732	1.103	1.291	0.163	1.169
4	1.281	1.1	1.203	0.338	0.696	10.28	1.103	0.765	1.292	0.228	1.118
5	-0.897	1.2	1.174	-0.434	0.636	10.19	-0.735	0.650	1.280	-0.213	1.080
6	0.109	1.0	1.184	-0.097	0.734	10.16	0.066	0.910	1.270	-0.082	1.127
7	-1.524	1.1	1.172	-0.550	0.690	10.23	-1.284	0.765	1.282	-0.736	1.110
8	-2.414	0.9	1.199	-1.050	0.795	10.19	-2.462	1.101	1.278	-0.585	1.158
9	1.042	0.9	1.202	0.732	0.807	10.28	1.166	1.102	1.289	0.374	1.168
10	0.366	1.0	1.188	0.366	0.751	10.28	0.385	0.911	1.292	0.208	1.144
11	-0.297	1.2	1.160	-0.213	0.640	10.23	-0.244	0.650	1.286	-0.126	1.085
12	-1.657	0.8	1.212	-0.638	0.846	10.16	-1.811	1.354	1.271	-0.214	1.176
13	2.037	1.1	1.175	0.967	0.699	10.34	1.782	0.765	1.294	0.343	1.294
14	-1.304	0.7	1.238	-0.041	0.912	10.19	-1.403	1.700	1.324	0.048	1.243
15	-0.915	0.9	1.210	-0.324	0.820	10.43	-0.830	1.104	1.311	-0.119	1.185
mean	-0.353	0.99	1.195	-0.163	0.753	10.25	-0.383	0.986	1.285	-0.205	1.150

W_t and V_t Unknown

It is suggested that the analyst consider changing values of W_t and V_t after some time n. For the Meinhold and Singpurwalla (1983) example, W_t and V_t were known to be correct but in the real world, knowledge of the correct values for W_t and V_t are usually unknown. A suggested procedure for this situation is the following. After a few, say n, values of Y_t have been obtained, compute an estimate of W_t as

$$(n - 1)\widehat{W}_t = \sum_{t=1}^{n}(Y_t - F_t\,\widehat{\theta}_t)^2 - [\sum_{t=1}^{n}(Yt - F_t\,\widehat{\theta}_t)\,]^2 \tag{8}$$

and of V_t as

$$(n - 1)\widehat{V}_t = \sum_{t=1}^{n}(\,\widehat{\theta}_t - G_t\,\widehat{\theta}_{t-1})^2 - [\,\sum_{t=1}^{n}(\,\widehat{\theta}_t - G_t\,\widehat{\theta}_{t-1})]^2 \tag{9}$$

In fact, there may be good reason to compute values for W_t and V_t for each values of t. In this way, the values of W_t and V_t will converge to their correct values for any data set under consideration. For the three pairs of values for W_t and V_t described above, let n = 5. The computations are given below.

<u>$W_t = 1$ and $V_t = 2$:</u>

$v_1 = 1.007 - 1.3(-0.618) = 1.8117$
$v_2 = -0.368 - 0.8(-0.350) = -0.0880$
$v_3 = -1.764 - 0.9(-0.527) = -1.2897$
$v_4 = 1.281 - 1.1(0.338) = 0.9092$
$v_5 = -0.897 - 1.2(-0.434) = -0.3762$

$\overline{v} = 0.19$ and $\widehat{V}_5 = 1.43$

$w_1 = -0.619 + 0.5(4.183) = 1.4725$
$w_2 = -0.350 - 0.5(-0.619) = -0.0405$
$w_3 = -0.527 + 0.5(-0.350) = -0.7020$
$w_4 = 0.338 - 0.5(-0.527) = 0.6015$
$w_5 = -0.434 + 0.5(0.338) = -0.2650$

$\overline{w} = 0.21$ and $\widehat{W}_5 = 0.72$

This would indicate that $V_t = 1.4$ and $W_t = 0.7$ are close to the correct values of 2 and 1.

<u>$W_t = 10$ and $V_t = 1$:</u>

$v_1 = 1.007 - 1.3(0.618) = 0.2036$
$v_2 = -0.368 - 0.8(-0.357) = -0.0824$
$v3 = -1.764 - 0.9(-1.732) = -0.2052$
$v_4 = 1.281 - 1.1(1.103) = 0.0677$
$v_5 = -0.897 - 1.2(-0.735) = -0.0150$

$\bar{v} = -0.01$ and $\widehat{V}_5 = 0.024$
$w_1 = 0.618 + 0.5(4.183) = 2.7095$
$w_2 = -0.357 - 0.5(0.618) = -0.6660$
$w_3 = -1.732 + 0.5(-0.357) = -1.9105$
$w_4 = 1.103 - 0.5(-1.732) = 1.9690$
$w_5 = -0.735 + 0.5(1.103) = -0.1835$

$\bar{w} = 0.38$ and $\widehat{W}_5 = 3.65$

The value for $W_t = 10$ is out of line with the data and should be reduced to 3, say, and the value for V_t left at one until further calculations.

$\underline{W_t = 1 \text{ and } V_t = 10:}$

$v_1 = 1.007 - 1.3(-1.592) = 3.0766$
$v_2 = -0.368 - 0.8(-0.771) = 0.2568$
$v_3 = -1.764 - 0.9(0.163) = -1.9107$
$v_4 = 1.281 - 1.1(0.228) = 1.0302$
$v_5 = -0.897 - 1.2(-0.213) = -0.6414$

$\bar{v} = -0.05$ and $\widehat{V}_5 = 3.66$

$w_1 = -1.592 + 0.5(4.183) = 0.4995$
$w_2 = -0.771 - 0.5(-1.592) = 0.0250$
$w_3 = 0.163 + 0.5(-0.771) = 0.2225$
$w_4 = 0.228 - 0.5(0.163) = 0.1465$
$w_5 = -0.213 + 0.5(0.228) = -0.0990$

$\bar{w} = 0.16$ and $\widehat{W}_5 = 0.05$

The value of $V_t = 10$ is too high and should be reduced to 3, say with $W_t = 1$. Further calculations will determine if other adjustments are needed.

It should be noted that the value $S_0 = 4.183$ was a guess but its effect diminishes quickly. The five obserations for t = 2- 6, say, should be used next to obtain refined estimates for V_t and W_t.

Discussion

The Kalman filter is robust to initial values of $\widehat{\theta}_t$ and Σ_t. If W_t and/or V_t are misspecified, suggestions are made for correcting these values in the updating process. Their effect can be nullified as more observations are obtained and as new estimates of these quantities are entered into the process. All in all, Kalman filtering appears to be a robust and self-correcting procedure. It is amazing that it has been around since 1960

(Kalman, 1960 and Kalman and Bucy, 1961) and still has not made its way into statistics textbooks and courses. By sequential use of the previous n measurements, the values used for the parameters can be filtered and refined until they become stable. If the values of parameters such as W_t and V_t change with time, this is easily incorporated into the process by the sequential use of the previous n values.

The gain in information using Kalman filter, mixed model, or Bayesian procedures may be seen from Equations (3) amd (4). The variance in (4) is smaller than in (3). The amount of recovery of information is apparent in these two equations.

Acknowledgement

Appreciation is expressed for the comments made by Professors Nozer D. Singpurwalla and B. R. Murty. The comments were helpful in clarifying certain consepts.

Literature Cited

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82:34-45.

Kalman, R. E. and R. S. Bucy (1961). New results in linear filtering and prediction theory. Journal of basic Engineering 83:95-108.

Meinhold, Richard J. and Nozer D. Singpurwalla (1983). Understanding the Kalman filter. The American Statistician 37(2):123-127.