Global Journal of Science Frontier Research: F
Mathematics and Decision Sciences
Volume 17 Issue 3 Version 1.0 Year 2017
Type : Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Deriving Kalman Filter - An Easy Algorithm

By Amaresh Das & Faisal Alkhateeb
Southern University At New Orleans, United States

Abstract- The Kalman filter may be easily understood by the econometricians, and forecasters if it is cast as a problem in Bayesian inference and if along the way some well-known results in multivariate statistics are employed. The aim is to motivate the readers by providing an exposition of the key notions of the predictive tool and by laying its derivation in a few easy steps. The paper does not deal with many other ad hoc techniques used in adaptive Kalman filtering.

Keywords: *bayes's theorem, state-space forecasting.*

GJSFR-F Classification: *MSC 2010: 11Y16*

DERIVINGKALMANFILTERANEASYALGORITHM

Notes

Deriving Kalman Filter - An Easy Algorithm

Amaresh Das [α] & Faisal Alkhateeb [σ]

Abstract- The Kalman filter may be easily understood by the econometricians, and forecasters if it is cast as a problem in Bayesian inference and if along the way some well-known results in multivariate statistics are employed. The aim is to motivate the readers by providing an exposition of the key notions of the predictive tool and by laying its derivation in a few easy steps. The paper does not deal with many other ad hoc techniques used in adaptive Kalman filtering.

Keywords: bayes's theorem, state-space forecasting.

I. Introduction

The Kalman filter wants to find at each iteration, the most likely cause of the measurement of Y_t given the approximation made by a flawed estimation the linear dynamics[1]. What is important here is not only that we have the measurement and the prediction, but knowledge of how each is flawed.[2] In the Kalman case, this knowledge is given by the covariance matrixes (essentially fully describing the distribution of the measurement and prediction for the Gaussian case). The main principle of forecasting is to find the model that will produce the best forecasts, not the best fit to the historical data. The model that explains the historical data best may not be best predictive model[3].The power of the Kalmancomes from its ability not only to perform this estimation once (a simple Bayesian task) but to use both estimates and knowledge of their distributions to a distribution for the updated estimate, thus iteratively calculating the best solution for state at each iteration[4].

Let Y_t , Y_{t+1}, ... Y_1, the data (which may be either scalar or vertical) denote the observed values of a variable of interest at times t, t-1, ... 1. We assume that Y_t

---

[1] The famous work by [1] was extension of Weiner's classical work. They focused attention upon aclass of linear minimum-error variance sequential error estimation algorithm. While the problem of linear minimum variance sequential filtering

[2] In the Kalman case, this knowledge is given by the covariance matrixes (essentially fully describing the distribution of the measurement and prediction for the Gaussian case.While many derivations of the Kalmanfilter are available, utilizing the orthogonalityprinciple orfnding iterative updates to the Best Linear Unbiased Estimator (BLUE), we will derive the Kalmanfilter here using a Bayesian approach, where 'best' is interpretedin the Maximum A-Posteriori (MAP) sense [2]instead of Gaussian innovations. This forecasting algorithm [5] is very flexible method that is particularly suitable in nonstationary time series. The Eq [7] used the method to forecast demand in the alcoholic drink industry over a period that included record demand followed by a drought and the imposition of a new excise duty.

[3] The future may not be described by the same probability as the past. Perhaps neither the past nor the future is a sample from any probability distribution. The time series could be nothing more than a non-recurrent historical record. •The model may involve too many parameters. Over fitted models could account for noise or other features in the data that are unlikely to extend into the future.The error involved in fitting a large number of parameters may be damaging to forecast accuracy, even when the model is correctly specified.

[4] It will be very convenient for the readers to remember the keywords used in the text:Filtering- When we estimate the *current* value given past and current observations, Smoothing: - When estimating *past* values given present and past measures, and Prediction - : When estimating a probable future value given the present and the past measures.

Author α: Professor, College of Business, Southern University at New Orleans & Department of Mathematics, University of New Orleans. e-mail: adas2@cox.net

Author σ: Assistant Professor, College of Business, Southern University at New Orleans.

depends on an unobservable be either a scalar or a vector whose dimension is independent of the dimensions of Y_t the relationship between Y_t and ϕ_t is linear and is specified by the observation equation

$$Y_t = \varpi_t \phi_t + \upsilon_t \tag{1}$$

where ϖ_t is a known quantity. The observation error υ_t is assumed to be normally distributed with mean zero and a known variance υ_t denoted as $\upsilon_t \to N(0, \upsilon_t)$

The essential difference between the Kalman filter and the conventional linear model representation is that in the former, the state of nature - analogous to the regression coefficients of the latter – is not assumed to be a constant but may change with time. This dynamic feature is incorporated via the system equation wherein

$$\phi_t = \Psi_t \phi_t + \zeta_t \tag{2}$$

Ψ_t being a known quantity and the system equation error $\zeta_t \to N(0, \zeta_t)$ with ζ_t known. Since there are physical systems for which the state of nature ϕ_t changes over time according to a relationship prescribed by engineering or scientific principles, the ability to include a knowledge of the system behavior in the statistical model is an apparent source of attractiveness of the Kalman filter. Note that the relationships (1) and (2) specified through ϖ_t and ψ_t may or may not change with time, as is also true of the variance υ_t and ζ_t we have subscripted these here for the sake of generality. In addition to the usual linear model assumptions regarding the error terms, we also postulate that υ_t is independent of time. The extension of the case of dependency is straightforward.

II. Extension of the Concept

To look at how the Kalman filter model might be employed in practice, we consider a situation in the context of statistical quality control. Here the observation Y_t is a simple (approximately normal) transform of the number of defectives observed in a sample obtained at time t, while ϕ_{1t} and ϕ_{2t}, represents, respectively, the true defective index of the process and the drift of the index. We have here as the observation equation

$$Y_1 = \phi_{1t} + \upsilon_{1t} \tag{3}$$

and as the system equations

$$\phi_{1,t} = \phi_{2,t} + \zeta_{2,t}$$

$$\phi_{2,t} = \phi_{2,t-1} + \zeta_{2,t}$$

In vector notation, the system of equation becomes $\phi_t = \psi\phi_{t-1} + \pi_t$
where

$$\phi_t = \begin{bmatrix} \phi_{1t} \\ \phi_{2t} \end{bmatrix} \text{and } \pi_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix}$$

$$\zeta = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Notes

Global Journal of Science Frontier Research (F) Volume XVII Issue III Version I Year 2017

does not change with time.

If we examine Y_t -.Yt-1. for this model, we observe that, under the assumptions of constant variance, $\upsilon_t = \upsilon$ and $\zeta_t = \zeta$ the autocorrelation structure of the difference is identical to that of ARIMA (0, 1, 1) process in the sense of [1]. Although with a correspondence is sometimes easily discernible, we should in general, not because of the discrepancies in the philosophies and methodologies involved, consider the two approaches to be equivalent.

Ref

1. Box and G E P and Jenkins G M (1970) *Time Series Analysis, Forecasting and Control*, San Francisco, Holden- Day

III. Recursive Procedure of the Filter

The Kalman filter refers to a recursive procedure for inference about the state of nature ϕ_t. The key notion here is that given the data $Y_t = (Y_t, \ldots Y_1)$ the inference about ϕ_t can be carried out through a direct application of a Bayes's theorem.

Prob {State of Nature | Data= Prob (Data | State of Nature}

which can be written as

$$P(\phi_t \mid Y_t) = P(Y_t \mid \phi_t, Y_{t-1}) \text{ x } P(\phi_t \mid Y_{t-1}) \quad (4)$$

Where the notion of P (A | B) denotes the probability of occurrence of even A given that (or conditional on) event B has occurred. Note that the expression on the left side of (4) denotes the *posterior* distribution for ϕ at time t, whereas the first and second expression on the right side denotes the likelihood and the prior distribution for ϕ, respectively.

The recursive procedure can best be explained if we focus attention on time point t-1, t = 1, 2, and the observed data until then, $(Y_{t-1}, Y_{t-2 \ldots} Y_1))$. In what follows, we use matrix manipulation in allowing for Y and / or, ϕ to be vectors without explicitly noting them as such.

At t-1, our state of knowledge without ϕ_{t-1} is embodied in the following probability statement for ϕ_{t-1}

$$(\phi_{t-1} \mid Y_{t-1} \rightarrow N(\hat{\phi}_{t-1}, \Sigma_{t-1}) \quad (5)$$

where $\hat{\phi}$ and Σ_{t-1} are the expectation and variance of $\phi_{t-1} \mid Y_{t-1}$ In effect (5) represents *posterior* distribution of ϕ_{t-1}; its evaluation will become clear in the subsequent text.

It is helpful to remark here that the recursive procedure is stated off at a time 0 by choosing $\hat{\phi}_0$ and Σ_0 to be our best guess about the mean and the variance of ϕ_0, respectively.

We now look forward to time t but in two stages[5]

1. Prior to observing Y_t and
2. After observing Y_t

[5] Kalman filters are ideal for systems which are continuously changing. They have the advantage that they are light on memory (they don't need to keep any history other than the previous state), and they are very fast, making them well suited for real time problems and embedded systems. For a Monte Carlo Sampling Method for Bayesian Filter see [3] Sequential Bayesian filtering is the extension of the Bayesian estimation for the case when the observed value changes in time. It is a method to estimate the real value of an observed variable that evolves in time. See [11]

Stage 1

Prior to choosing Y_t our best choice for ϕ_t is governed by the system equation (2) and is given by $\Psi \phi_{t-1} + \zeta_t$. Since ϕ_{t-1} is described by (5) and state of knowledge above ϕ_t is embodied in the probability statement

$$(\phi_{t-1} \mid Y_{t-1} \rightarrow N(\psi_t, \hat{\phi}_{t-1}, \ \Theta = \Psi_t \Sigma_{t-1} \psi' + \zeta_t) \tag{6}$$

This is our prior distribution.

In observing (6) which represents our prior for ϕ_t in the next cycle of (4), we use the well -known result that for any constant c

$$X \rightarrow N(\mu, \Sigma) = CX \rightarrow N(C\mu, \ C\Sigma C')$$

Where C denotes the transpose of C

Stage 2

On observing Y_t our goal is to complete the posterior of ϕ_t using(4). However, to do this, we need to know the likelihood $\Re(\phi_t \mid Y_t)$ or equivalently P (Y_t) the determination of which is undertaken via the following arguments.

Let e_t denote the error in predicting Y_t from the point t-1; thus

$$e_t = Y_t - \hat{Y}_t = Y_t - \omega_t \psi_t \hat{\phi}_{t-1} \tag{7}$$

Since ω_t, ψ_t and $\hat{\phi}_{t-1}$ are all known, observing Y_t is equivalent to observing e_t. Thus (4) can be written as

$$P(\phi_t \mid Y_t, Y_{t-1}) = P(\phi_t \mid e_{t-1}) = P(e_t \mid \phi_t, Y_{t-1}) \mathbf{X}$$

$$P(\phi_t \mid Y_{t-1}) \tag{8}$$

with P $(e_t \mid \phi_t, Y_{t-1})$ being the likelihood[6].

Using the fact that $Y_t = \omega_t \varphi_t + \upsilon_t$ (7) can be written as $e_t = \omega_t(\phi_t - \psi_t \phi_{t-1}) + \upsilon_t$ so that $\Sigma(e_t \mid \phi_t, Y_{t-1}) = \omega_t(\phi_t - \psi_t \hat{\phi}_{t-1})$

Since $\upsilon_t \rightarrow N(0, \upsilon_t)$ it follows that the likelihood function is described by

$$(e_t \mid \phi_t, Y_{t-1}) \rightarrow N(\upsilon_t(\phi_t - \psi_t \hat{\phi}_{t-1}), \upsilon_t \tag{9}$$

We can now use Bayes's theorem (eel 8) to obtain

$$P(\phi_t \mid Y_t, Y_{t-1}) = \frac{P(e_t \mid \phi_t, Y_{t-1}) \text{ x } P(\phi_t \mid Y_{t-1})}{\int_{\text{all } e_t} P(e_t, \phi_t \mid Y_{t-1}) \, d\phi_1} \tag{10}$$

and this best describes our state of knowledge about ϕ_t at time t. Once $P(\phi_t \mid Y_t, Y_{t-1})$ is continued, we can go back to (5) for the next cycle of the recursive procedure. Therefore, Kalman filter can be a very effective forecasting tool. It should be useful in a

[6] The opportunity exists to proclaim an inherent equivalence of the least square estimation and Kalman filter theory, See [3] See also [2]

Notes

wide variety of situations. [9]developeda complete numerical procedure called 'state-space forecasting' for predicting future values of a multivariate stationary process Y_t given past values. The procedure involves basically two main stages.

a. First fit a canonical state-space model to the given observation using Akaike's canonical correlations technique to determine the dimensions of the state vector and to provide estimates of the non-zero elements of the matrix ω. A multivariate AR model is also fitted to the observations(using AIC to determine the order) to provide estimates of Σ_t and the impulse response matrices.
b. Having filtered and estimated ω, ψ, Σ, the procedures are computed recursively using Kalman's algorithm[7]. Practical applications are given in the paper by Mehra.

IV. Conclusion

The note presents a mathematical theory of Kalman filtering. The filtering techniques is discussed as a problem in Bayesian inference in a series of elementary steps, enabling the optimality of the process to be understood. The style of the note is informal and the mathematics elementary but rigorous, making it accessible to all those with a minimal knowledge of linear algebra and systems theory Many other topics related to Kalman filtering are ignored (for example, Wavelet) although occasionally we referred to them inside the text.

References Références Referencias

1. Box and G E P and Jenkins G M (1970) *Time Series Analysis, Forecasting and Control,* San Francisco, Holden- Day.
2. Chui, C K and Chen G (1990) *Kalman Filtering with Real-Time Applications*, Second Edition, Springer- Verlag.
3. Doucet, A, Godsill S and Andrieu C (2000) "On Sequential Monte Carlo Sampling Method for Bayesian Filters, *Statistics and Computing,* 10, 197-208.
4. Duncan D B and Harvard S D (1972) "Linear Dynamic Recursive Estimation from the View Point of Regression Analysis" *Journal of the American Statistical Association*, 67, 815-821.
5. Harrisin F J and Stevens, C F (1971) 'A Bayesian Approach to Short-term Forecasting', *Operations Research Quarterly,* 22, 341-362.
6. Grossman, A and Morlet J (1984) "Decompositions of Hardy Functions into Square Intrgrable Wavelets of Constant Shape, SIAM J of Mathematics Annals, 15, 720-736.
7. Johnson, F R and Harrison, P J (1980) 'An Application of Forecasting in the Alcoholic Drinks Industry" *Journal of the Operations Research Society*, 31, 699-709.
8. Kalmam R E and BucyR S (1961) New Results in Linear Filtering and Predictions, Trans ASME Journal Basic Engineering, 83. 95-108.

[7] In addition to the Kalman filtering algorithms there are other time domain algorithms available in literature. Perhaps the most exciting ones are the so-called wavelet algorithms. Wavelets were first introduced by [6].Wavelets are based on translation W(x) $\rightarrow W(x+1)$ and above all on dilation (w (x) $\rightarrow (2x)$ The basic dilation is a two-scale difference equation $\Phi(x) = \Sigma c_k \Phi(2x-k)$ ... We look for a solution normalized by $\int \Phi \, dx = 1$ The first requirement on the coefficients c_k comes from multiplying by 2 and integrating $2\int \Phi \, dx = \Sigma c_k \int \Phi(2x-k)\, d(2x-k)$ yields Σc_k = 2. Uniqueness of Φ is ensured by $\Sigma c_k = 2$

Ref

9. Mehra, R K (1979) "Kalman filters and their Applications in Forecasting' TIMS Studies in Management Sciences Ed M K Starr, Amsterdam, North-Holland, 37, 207-213.

9. Mehra, R K (1979) "Kalman filters and their Applications in Forecasting' TIMS Studies in Management Sciences Ed M K Starr, Amsterdam, North-Holland, 37, 207-213
10. Mein hold, R and Singpurwalla, N D (1983) "Understanding the Kalman Filter' *The American Statistician,* Vol 37, Issue 2.
11. Sarakka, Simo (2013) *Bayesian Filtering and Smoothing*, Cambridge University Press (PDF).

Notes