Kalman Filtering and Probability

1. A Brief Summary of Kalman Filtering in Relation to Random Variables

1.1 The State and Measurement Models

Let $Z_k = [Z_{k1}, \ldots, Z_{km}]^{tr}$ and let $X_k = [X_{k1}, \ldots, X_{kn}]^{tr}$. Assume the relationship between these random processes is:

$$X_{k+1} = F_k X_k + W_k \quad (\textit{state equation}) \quad ; \quad \hat{X}_0^- \, (\textit{initial condition}) \tag{1a}$$

$$Z_k = H_k X_k + V_k \quad (\textit{measurement equation}) \tag{1b}$$

where (1a) includes the known and non-random ($n \times n$ matrix-valued) parameters F_k and G_k, as well as the n-D non-measurable *white noise* random process w_k with

$$E(W_k W_{k+j}^{tr}) = Q \bullet \delta(j) \text{ and } E(V_k V_{k+j}^{tr}) = R \bullet \delta(j) . \tag{2}$$

The Random Variables involved in the state model- There are two ways in which random variables can enter into (1a). The first way is via the process $W_k = [W_{k1}, \ldots, W_{kn}]^{tr}$. For any k, this n-D random variable need not have all non-zero entries. In fact, it often does not. Furthermore, the non-zero random variables may have distinctly different distributions. For example, one component might be normally distributed, while another is uniformly distributed, while a third has a Bernoulli distribution. Moreover, at any time, k, the elements of W_k can be correlated. The only requirement is that the possibly *nonstationary random process*, $\{W_k\}$, be *zero-mean* and have mutually uncorrelated elements. The second way that randomness can enter into (1a) is via the initial condition; which may also entail zeros, as well as various types of random variables. It is possible that the state process (1a) has no driving noise input, and has a deterministic external input; but where the initial condition is a random variable. In this case, (1a) is an example of a *deterministic random process,* in that conditioned on knowledge of X_k for a given time, k, it is known for all future time.

Random Variables involved in the measurement model- Clearly, the randomness will arise, in part, from the state process. In addition to this, it may also have a contribution from the *measurement noise* k^{th} element of the random process $\{V_k\}$. This process has the same amount of allowed variety of random variables, and has the same constraint as the *state noise* process $\{W_k\}$.

1.2 The Kalman Filter Algorithm

From Figure 5.8 on p.219, we have the following KF algorithm:

For k=0 choose a value $\hat{X}_0^- = \hat{x}_0^-$ and compute the ***prediction error covariance***

$$P_0^- = E[(X_0 - \hat{X}_0^-)(X_0 - \hat{X}_0^-)^{tr}] . \tag{3}$$

[Notice that by choosing $\hat{x}_0^-$ =0, it follows that $P_0^- = E[X_0 X_0^{tr}]$.]

***Step* 1**: For k=0: Compute the Kalman gain: $K_k = P_k^- H_k^{tr} (H_k P_k^- H_k^{tr} + R_k)^{-1}$. (4a)

***Step* 2**: Compute the update estimate: $\hat{X}_k = \hat{X}_k^- + K_k(Z_k - H_k \hat{X}_k^-)$. (4b)

Step 3: Compute the update error variance: $P_k = (I - K_k H_k) P_k^-$. (4c)

***Step* 4**: Compute 1-step prediction: $\hat{X}_{k+1}^- = F_k \hat{X}_k$ (4d)

***Step* 5:** Compute the next state error covariance: $P_{k+1}^- = F_k P_k F_k^{tr} + Q_k$ (4e)

***Step* 6:** Return to *Step 1* and increment k by one.

Note that while steps 1-3 must precede steps 4 and 5, the order of the latter is irrelevant.

2. Probability Distributions Related to the State and Measurement Random Variables

2.1 Distributional knowledge related to X_k

Recall that the point of Kalman filtering is to recover an *estimate* of X_k from the measurements $\{Z_{j \le k}\}$ Hence, even though X_k is a random variable, we do not have access to it.

2.1.1 *The mean and covariance of* $X_k \mid \hat{X}_k^- = \hat{x}_k^-$:

Recall that $\hat{X}_k^-$ is the KF estimator of X_k prior to observing Z_k. Let $E_k^- \stackrel{\Delta}{=} X_k - \hat{X}_k^-$ denote the error associated with this estimator. This error has zero mean and covariance denoted as:

$$P_k^- = E[(X_k - \hat{X}_k^-)(X_k - \hat{X}_k^-)^{tr}] \stackrel{\Delta}{=} E(\mathrm{E}_k^- \mathrm{E}_k^{-\,tr}). \qquad (5)$$

Suppose that we are given $\hat{X}_k^- = \hat{x}_k^-$. Since $X_k = \hat{X}_k^- + (X_k - \hat{X}_k^-) \stackrel{\Delta}{=} \hat{X}_k^- + E_k^-$, it follows that

$$X_k \mid \hat{X}_k^- = \hat{x}_k^- \quad = \quad \hat{x}_k^- \quad + \quad \mathrm{E}_k^-. \qquad (6a)$$

Cleary, then

$$E(X_k \mid \hat{X}_k^- = \hat{x}_k^-) \;=\; \hat{x}_k^-, \qquad \text{and} \qquad Cov(X_k \mid \hat{X}_k^- = \hat{x}_k^-) \;=\; Cov(E_k^-) \;=\; P_k^-. \qquad (6b)$$

Thus, while we do not know anything *directly* in relation to X_k, we *do* know the *mean* and *covariance* of the *conditional* random variable $X_k \mid \hat{X}_k^- = \hat{x}_k^-$, are as is given in (6).

2.1.2 *The mean and covariance of* $X_k \mid \widehat{X}_k = \widehat{x}_k$ **:**

Recall that $\widehat{X}_k$ is the KF estimator of X_k that incorporates Z_k. Let $E_k \overset{\Delta}{=} X_k - \widehat{X}_k$ denote the error associated with this estimator. This error has zero mean and covariance denoted as:

$$P_k = E[(X_k - \widehat{X}_k)(X_k - \widehat{X}_k)^{tr}] \overset{\Delta}{=} E(\mathrm{E}_k \mathrm{E}_k^{tr}) . \tag{7}$$

Suppose that we are given $\widehat{X}_k^- = \widehat{x}_k$. Since $X_k = \widehat{X}_k + (X_k - \widehat{X}_k) \overset{\Delta}{=} \widehat{X}_k + E_k$, it follows that

$$X_k \mid \widehat{X}_k = \widehat{x}_k \quad = \quad \widehat{x}_k \quad + \quad \mathrm{E}_k . \tag{8a}$$

Cleary, then

$$E(X_k \mid \widehat{X}_k = \widehat{x}_k) \quad = \quad \widehat{x}_k , \qquad \text{and} \qquad Cov(X_k \mid \widehat{X}_k = \widehat{x}_k) \quad = \quad Cov(E_k) \quad = \quad P_k . \tag{8b}$$

From *Step* 2 of the *KF* algorithm, we have

and $$\widehat{X}_k \overset{\Delta}{=} \widehat{X}_k^- + K_k (Z_k - \widehat{Z}_k^-) \text{ with } \widehat{Z}_k^- \overset{\Delta}{=} H_k \widehat{X}_k^- . \tag{8c}$$

2.2 Distributional knowledge related to $Z_k \mid \widehat{X}_k^- = \widehat{x}_k^-$

Since Z_k depends on X_k, via (1b), we do not have any a priori knowledge of the distribution of $\{Z_k\}$. However, we may write (1b) as

$$Z_k = H_k X_k + V_k \quad = \quad H_k (\widehat{X}_k^- + \mathrm{E}_k^-) \quad + \quad V_k . \tag{9}$$

Define $\widehat{Z}_k^- \overset{\Delta}{=} H_k \widehat{X}_k^-$ and $\widehat{z}_k^- \overset{\Delta}{=} H_k \widehat{x}_k^-$. It follows immediately that

$$Z_k \mid \widehat{X}_k^- = \widehat{x}_k^- \quad = \quad H_k \widehat{x}_k^- + H_k \mathrm{E}_k^- + V_k \quad = \quad \widehat{z}_k^- + H_k \mathrm{E}_k^- + V_k . \tag{10a}$$

The mean and covariance of (10a) are given by

$$E(Z_k \mid \widehat{X}_k^- = \widehat{x}_k^-) = H_k \widehat{x}_k^- = \widehat{z}_k^- \quad \text{and} \quad Cov(Z_k \mid \widehat{X}_k^- = \widehat{x}_k^-) = + H_k P_k^- H_k^{tr} + R_k . \tag{10b}$$

Finally, notice that the variable $Z_k \mid \widehat{X}_k = \widehat{x}_k$ is not a random variable, but a number, z_k. This is because to be given $\widehat{X}_k = \widehat{x}_k$ requires that we are given $Z_k = z_k$.

2.3 Application of Baye's Theorem

In its simplest form, Baye's Theorem may be stated as

Baye's Theorem $$\Pr(A \mid B) = \Pr(B \mid A)\,\Pr(A)/\Pr(B). \tag{11}$$

Proof: The proof is a trivial consequence of the definition of the conditional probability $\Pr(A \mid B) \overset{\Delta}{=} \Pr(A \cap B)/\Pr(B)$. We may also write this as $\Pr(B \mid A) \overset{\Delta}{=} \Pr(B \cap A)/\Pr(A)$. Since $\Pr(A \cap B) = \Pr(B \cap A)$, the result (11) follows immediately. □

Now, let $Z_k \mid \widehat{X}_k^- = \widehat{x}_k^- \overset{\Delta}{=} W$, and let A be a random variable with sample space S_A. Then, from (11), we have, Notationally, $\Pr(A \mid W) = \Pr(W \mid A)\,\Pr(A)/\Pr(W)$. However, whereas (11) relates to *events* A and B, this expression relates to random variables. Hence, more formally, (11) becomes

$$f(a \mid w) = f(w \mid a)\, f(a)/f(w) \tag{12a}$$

where

$$f(w) = \int_{S_A} f(w,a)\,da \tag{12b}$$

Example. Consider the dynamical system

$$X_k = a\,X_{k-1} + U_k \tag{13a}$$

$$Z_k = X_k + V_k. \tag{13b}$$

Now, suppose that the number $a \in \{a_1\ , a_2\} = S_A$, with $\Pr[A = a_1] = p$. Then (12) becomes

$$\Pr[A = a_1 \mid W = w] = \left[\frac{f(w \mid a_1)}{f(w,a_1) + f(w,a_2)}\right] \Pr[A = a_1].$$

This can also be written as

$$\Pr[A = a_1 \mid W = w] = \left[\frac{f(w \mid a_1)}{f(w \mid a_1)\Pr[A = a_1] + f(w \mid a_2)\Pr[A = a_2]}\right] \Pr[A = a_1]. \tag{14a}$$

From (10b) we have

$$E(W \mid a) = \widehat{z}_k^- \quad \text{and} \quad Cov(W \mid a) = +P_k^-(a) + \sigma_V^2. \tag{14b}$$

Then for the event $[W = w = z_k - \widehat{z}_k^-]$, (14a) becomes

$$\Pr[A = a_1 \mid \widehat{z}_k^-] = \left[\frac{f(z_k - \widehat{z}_k^- \mid a_1)}{f(z_k - \widehat{z}_k^- \mid a_1)\Pr[A = a_1] + f(z_k - \widehat{z}_k^- \mid a_2)\Pr[A = a_2]}\right] \Pr[A = a_1]. \tag{15a}$$

Equation (15a) may be viewed as an updated specification of $\Pr[A = a_1]$, based on having z_k. With this viewpoint, we express this updated probability as

$$\Pr[A = a_1] \overset{\Delta}{=} \Pr[A = a_1 \mid \hat{z}_{k-1}^-]. \tag{15b}$$

Substituting (15b) into (15a) results in

$$\Pr[A = a_1 \mid \hat{z}_k^-] = \left[\frac{f(z_k - \hat{z}_k^- \mid a_1)}{f(z_k - \hat{z}_k^- \mid a_1)\Pr[A = a_1 \mid \hat{z}_{k-1}^-] + f(z_k - \hat{z}_k^- \mid a_2)\Pr[A = a_2 \mid \hat{z}_{k-1}^-]} \right] \Pr[A = a_1 \mid \hat{z}_{k-1}^-] \tag{16}$$

For convenience, assume that $W \mid a \sim N(\hat{z}_k^- \,,\, P_k^-(a) + \sigma_V^2)$.

We return to the first example of the semester.

Example. Consider

$$z_k = a_k z_{k-1} + v_k \quad ; \quad E(v_k^2) = \sigma_v^2(k) \tag{17a}$$

In the case where the AR(1) parameter a_k changes slowly in relation to the sampling interval. In this sense, the process (16a) is a locally *wss* process, and it is easy to show that the process variance (or power) is given by

$$\sigma_z^2(k) \cong \sigma_v^2(k)/(1 - a_k^2). \tag{17b}$$

The model (17) has the ability to capture the slow time variation in both the process power, and time-varying frequency content can be captured by the AR(1) parameter, a_k.

Suppose that the AR(1) parameter is changing in the manner described in Figure 1.

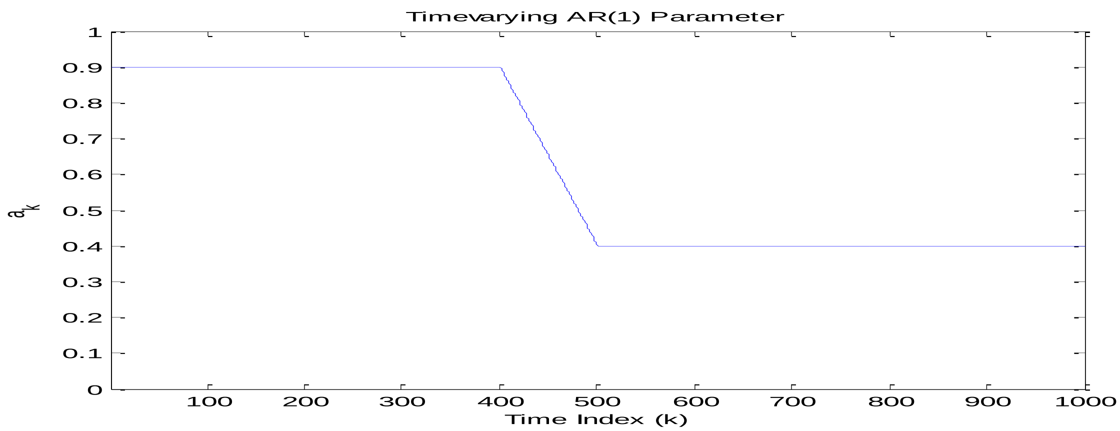


Figure 1. Time-varying AR(1) parameter, a_k.

Figure 2 shows a realization of this process. Note that the bandwidth is change, but the process power is not.

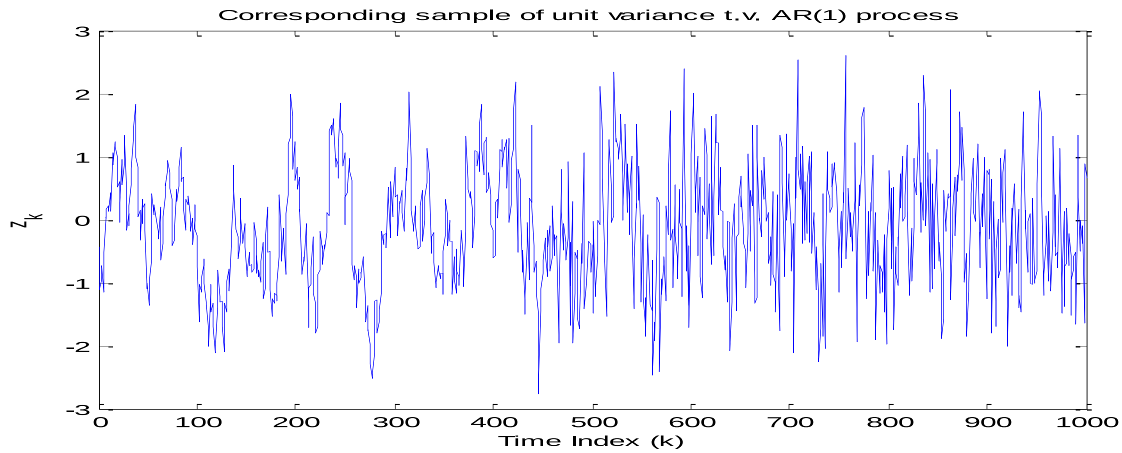


Figure 2. Measurement of z_k associated with (3) and Figure 1. **How was this generated?**

***Method 1*:** Here we will concentrate on tracking a_k. In this case we will use measurements of the process (16) to filter out estimates of the a_k's. The state model used here is a simple random walk model:

$$a_k = a_{k-1} + w_k \;\; ; \;\; Q_k = \sigma_w^2 \tag{18}$$

Hence, the quantities in the KF state/measurement model (1) are:

$$x_k = a_k ; F_k = G_k = 1; Q_k = \sigma_w^2 ; H_k = z_{k-1} ; R_k = \sigma_v^2(k) .$$

There are two items here that make this KF suboptimal in the sense of minimizing the mean squared error between the state process x_k and any estimator of it:

1.The parameter $H_k = z_{k-1}$ is not a non-random quantity. However, because at time k we have knowledge of z_{k-1}, it is *conditionally* non-random. In this sense, the KF is known as an *extended* KF.

2.We do not have knowledge of $R_k = \sigma_v^2(k)$, which is the driving white noise for the process (17a).

However, we do know that the process power (17b) is constant and known. Hence, with our estimate $\hat{a}_k$ we can estimate this white noise variance as $\hat{\sigma}_v^2(k) = \max\{0 \, , \, \sigma_z^2(1-\hat{a}_k^2)\}$.

An example of the EKF tracking for two specified values of $Q_k = \sigma_w^2$ is shown in the plots below.

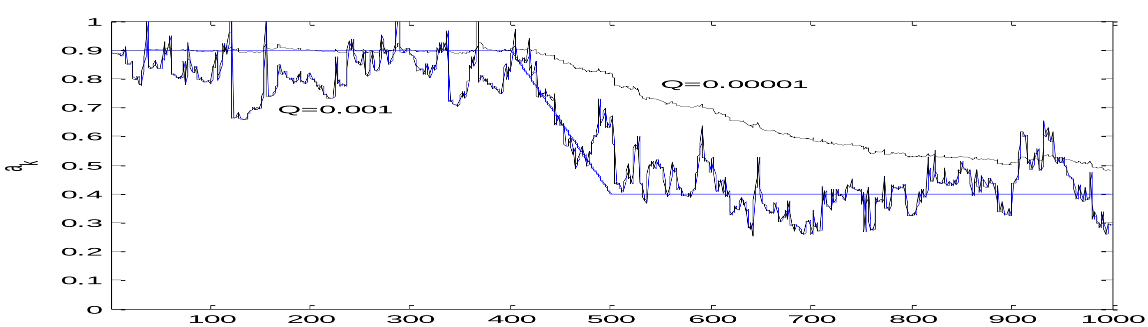


Figure 3. EKF estimates of a_k for $Q_k = \sigma_w^2$: 0.00001 and 0.001.

The trade-off between bias and variability is clear. What is not clear is how this translates into change detection. If we define change as change in relation to the value 0.9, then Figure 3 would include numerous false detections for a threshold value of, say, 0.8. If we desire to detect whether a_k is 0.9 or 0.4, with say a threshold value of 0.65, then the only incorrect detections occur in the transition region. The following method is designed to rapidly detect whether a_k has one of two specified values.

Method 2: For this method, the ramp was replaced by an instantaneous change in the AR(1) parameter

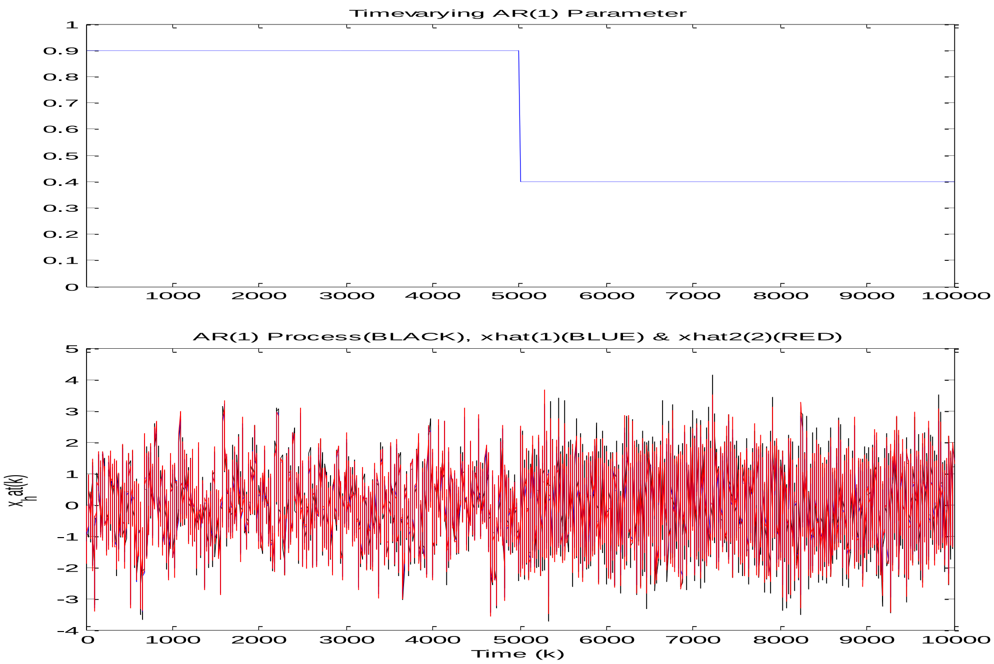
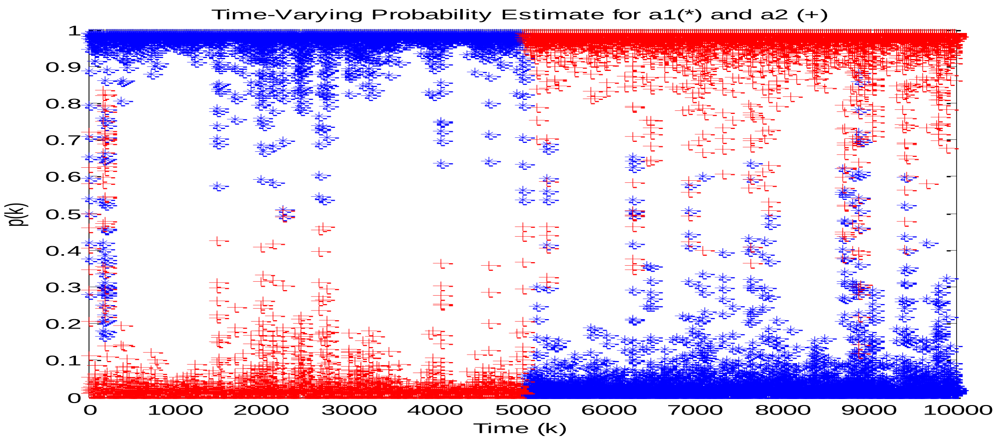
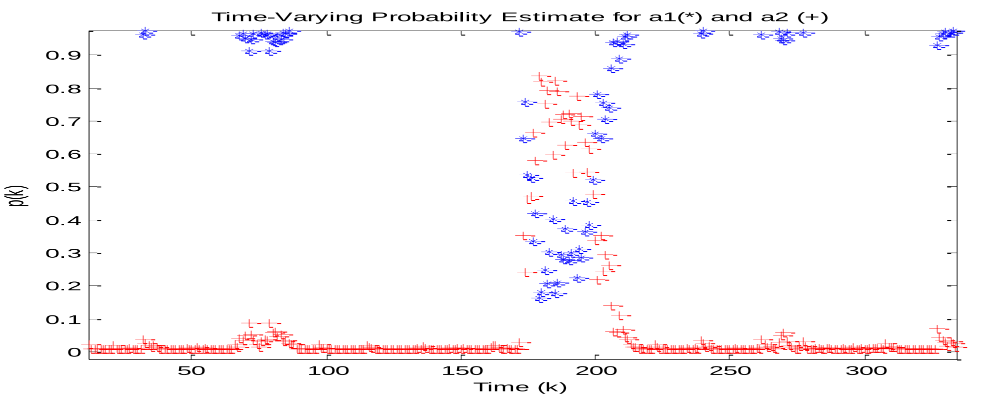


Figure 4. The top plot shows the behavior of the AR parameter. The bottom plot shows a partial realization of the measurement process.

In this method we will not track a_k. Instead, we will cast this as a problem of determining which of two *known* values of a_k the measurements are associated with. In this setting, we will use the probability evolution model (16) in relation to each value. In this way, we are actually running TWO EKFs, and using (16) to decide which one is related to the true value of a_k.

The results are shown in Figure 5.


Time-Varying Probability Estimate for a1(*) and a2 (+)
p(k)
Time (k)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000


Time-Varying Probability Estimate for a1(*) and a2 (+)
p(k)
Time (k)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
50
100
150
200
250
300

```
% PROGRAM NAME: mmae411.m
% Most Recent Modifications 4/16/2011
% =================================
% Multiple Model Adaptive Estimation- an Example
% AR(1) Parameter Probability Information
% =================================
% DATA GENERATION
n = 10^4;
% Construct t.v. AR(1) parameter
a11=.9; a22=.4;
a1=a11*ones(1,n/2);
a2=a22*ones(1,n/2);
 avec=[a11 ; a22]; % Vector of the two parameters being tested
 a=[a1 a2];
 figure(1)
   plot(a)
   axis([1,n,0,1])
   title('Timevarying AR(1) Parameter')
   pause
 % Compute driving noise variance for unity variance AR(1)
 sige2=(1 - a.^2);
 % Generate t.v. AR(1) Measurement Process
 x = zeros(1,n);
 x(1)=randn(1,1);
 for t=2:n
    x(t)=a(t)*x(t-1) + sige2(t)^(0.5)*randn(1,1);
 end
 figure(2)
   plot(x,'k')
   title('Sample of unit variance t.v. AR(1) process')
   pause
 R=0.1;              %  MEASUREMENT NOISE VARIANCE
v=sqrt(R)*randn(1,n);% WHITE NOISE
z=x+v;               % MEASUREMENT
%=============================================================================
%   Use of TWO Simultaneous Kalman Filters to Estimate Pr(A1) and Pr(A2) = 1-Pr(A1)
%=============================================================================
% PROBABILITY INITIAL CONDITIONS
pr=[];
pra=[0.5;0.5]; % Assign equal prior probabilities to a1 & a2
pnm1=pra;      % pdf for z(0)=[0 0]
%======================================================
% KALMAN FILTER INITIAL CONDITIONS
Xhat=[];
Pmat=[];
Kmat = [];
ehatm=[];
Q=1-avec.^2;  % white noise variances for sigma_x=1
xhatm=[0;0];
Pm=[1;1];     % Note that R has been specified above
%=====================================================
for k=1:n
 % COMPUTATION OF pr(k) = [pk;a1) p(k;a2)]'
  zk=z(k)*[1;1];
  p1coef=(2*pi*(Pm+R)).^(-0.5);
  p1=p1coef.*exp((-0.5*(zk-xhatm).^2)./(Pm+R));
  pd=p1'*pnm1;        %This is the sum of TWO probabilities
  p=(1/pd)*p1.*pnm1;  % NOTE that pden is scalar-valued
  [pmax,i]=max(p);
   if pmax > .99
     p(i)=.99;
     j=1+mod(i,2);
     p(j)=1-p(i);
   end
  pr(:,k)=p;
  pnm1=p;
  % KALMAN FILTER COMPUTATIONS
  K=Pm.*(Pm+R).^(-1);
```

```
  Kmat=[Kmat,K];
  P=([1;1]-K).*Pm;
  Pmat=[Pmat,P];
  xhat=xhatm+K.*(zk-xhatm);
  Xhat=[Xhat,xhat];
  Pm=(avec.^2).*P + Q;
  xhatm=avec.*xhat;
  em = x(k)*[1;1] - xhatm;
  ehatm = [ehatm em];
end
tvec=1:n;
figure(2)
  hold
  plot(tvec,Xhat(1,:),'b', tvec,Xhat(2,:),'r')
  xlabel('Time (k)')
  ylabel('x_hat(k)')
  title('AR(1) Process(BLACK), xhat(1)(BLUE) & xhat2(2)(RED)')
  pause
figure(3)
  dxhat=Xhat(1,:)-Xhat(2,:);
  plot(tvec,dxhat)
  xlabel('Time (k)')
  ylabel('del_xhat(k)')
  title('Difference in the AR(1) Process Predictions')
  pause
figure(4)
  plot(tvec,pr(1,:),'b*',tvec,pr(2,:),'r+')
  xlabel('Time (k)')
  ylabel('p(k)')
  title('Time-Varying Probability Estimate for a1(*) and a2 (+)')
  pause
figure(5)
  plot(tvec,pr(2,:))
  xlabel('Time (k)')
  title('Probability of detection')
```