1. Conditional Expectations and Discrete-Time Kalman Filtering

* Introduction
* Model: the measurement is modeled as the sum of two 2 RVs.
* Known fact
* Unknown fact (You may have to calculate )

* Given a measure of , What is an estimator of For example
* Kim: Example

If you measure your temperature, there are two R.V. Let two random variables,: temperature, measurement sensor noise.



You measured yours as .Which is the best estimator of your temperature?

1. The simple one may be the estimator is . So how to get ?

If you assume they are Gaussians, . Or you may statistical average.

Is this the best?

1. The answer is no. The best one is

How to get this? Later we will show if they are Gaussian, then

**Where this is the BASIC, FUNDAMENTAL LAW…**OK… Then the problem is how to find if they are not Gaussian. It is Example 3.8. It is not simple…



1. The **minimum variance estimator**
   1. Some textbooks, it is called as the **minimum mean square error**

<https://en.wikipedia.org/wiki/Minimum_mean_square_error#:~:targetText=In%20statistics%20and%20signal%20processing,values%20of%20a%20dependent%20variable.>

* 1. Identification : the **least (mean) square error**

1. The maximum a posteriori estimator

Bayesian probability

* : = the posteriori PDF
* : = the priori PDF
* : the likelihood

%% Kim

The posterior : after measurement, estimate the source

The priori : before measurement

%%%%

* 1. Minimum Variance Estimation
* Problem statement – static parameter estimation

where .

* The minimum variance estimator
* Remark 3.1.argmin
* (Skip): Define convex distance function

Is a non-negative and convex distance function.

-. Convex function: a function is convex if for any points, and some scalar such that

-. Example of a convex function

* (Skip): Define a loss function , such that

-.

-.

(Skip) Theorem 3.2(Sherman’s Theorem). Let x be a random vector with mean, , and density,. Let be a loss function as defined above. If . Is symmetric about and unimodal(i.e., has only one peak), then minimizes .

* Without observation, the minimum variance estimator.

Sol:

* Remark Sherman’s theorem is generalized of the above estimator.

%%%% Kim’s comment

Minimum variance estimator:

Minimum variance (conditional) estimator:

* There are two MVE. But is the MVE also, the expectation.
* Usually MVE is the conditional minimum variance estimator

And in Linear algebra

Least Square solution = least square error estimator:

* Solution : not to be confused with MVE

%%%

* Theorem 3.6. Given the equation(3.1). **if the estimate is a function of , then the minimum variance estimate is the conditional mean.**

(Skip) Proof: let be an estimator of ( that means given is a constant).

Then for

Since the cross term is

.

Then

Since are independent of and is positive semidefinite,

1. is equivalent to

* Remarks:

1. It means
2. It is **unbiased**

%% Kim, comment

It is important that

For example,

* The solution is

: which is a R.V. dependent of the initial point R.V.

What is MVE? Yap

Now what is

%%%%

**The following example by M.Idan should be carefully understood. The main goal is to verify**

%%% kim’s comment

Let tow random variables which are independent with pdf as Define

Find pdf of Z

Solution:

Now

dy

=

* This integral is called the convolution integral.

Examp.

Define Find

f

1. For 0, i.e., since (a.1), f ,
2. For , i.e., from (a.1)
3. For , from (a.1)
4. For from (a.1)
5. In conclusion

X

Y

z = x+y

Or you may plot then calculated.

f

1. Plot each pdf
2. Change the graph one of them

g(-x+z)

g(-x)

g(x)

-1+z z

1

-1

1. Increase from negative infinity to positive infinity to integrate the convolution integral

-1+z z

g(-x+z)

f(x)

1. If , there is no overlapped portion, so that f
2. If

There is overlapped portion so that

f

1. If , there is a overlapped portion so that

f

1. If If , there is no overlapped portion so that

f

-1+z z

f(x)

g(-x+z)

-1+z z

g(-x+z)

f(x)

%%%%%%

* **Ex.3.8: conditional mean and variance of the sum of two RV’s**

**In tutorial We well consider it.**

* 1. Conditional Estimate of a Gaussian Random Vector with Additive Gaussian Noise
* Modelling
* %% I will notify means is independent.
* Problem: find

To get the solution, we need

1. Since are gaussians, is a gaussian
2. , and are correlated.
3. How to find given ?

* Solution

where

* The minimum variance estimator of

= the conditional expectation given

~ mean of x + (scale) (measurement – mean of x)

* Uncertainty: if the variance of a RV is large, the uncertainty is large.

%%% Kim’s comment : Scaler case

In the scalar case, every thing is a scalar not a vector or a matrix. So it may help to understand the basic concept of the (3.14), which is called as a Kalman filter.

First of all, the conditional covariance of

* How to find ?

-first find : He use a trick. Define a new RV as

-second find : This is simple. As you know

* Tip: (3.5) Matrix inverse

Verify the inverse in (3.5) 🡺

And 🡨 You may try

* Tip.
  + 1. (Skip) Simplification of the Argument of the Exponential
    2. (Skip) Simplification of the Coefficient of the Exponential

%%% Kim’s comment:

Given a measurement R.V.’s which is the sum of two independent Gaussian R.V.’

1. The is a Gaussian as
2. The conditional mean of given is a Gaussian as

where

And the conditional variance of given

%%% Kim’s comment:

The filter (3.14) is called a Kalman filter in case of batch process. What is the batch process? Well

First of all we may understand the (3.14), then after that we may think about the batch process.

1. Scalar case (3.14)
   1. If the noise is very small, i.e.,

* The MV is only dependent of the measurement
  1. If the noise is very large, i.e.,
* The MV is only dependent of the state regardless of the measurement
  + 1. Processing Measurements Sequentially
* Batch process / real time process (sequential measurement)
* Extension of (3.2): Sequentially measuring of
* **Goal:** find sequential the conditional mean of given the sequential measurements

Where in (3.32) means corresponding to in (3.13) indexed n.

* Solution

With the sequential measurements, construct it as a batch model

Since

Hence

Substituting (3.21) and (3.20) into (3.14) gives

Multiplying (3.22) through by gives

where a new variable **defined** as

And **define** additional terms as

We get

For iteration calculation, define new variables at measurement

And



* Iteration with at time step

Multiplying by then



Or

* Important:

1. With the last measurement , the conditional mean of can be calculated by (3.14) or (3.32)



1. To get the , in the batch process you need more memory compared to the real time process
2. In general, .Hence we need recursively using (3.28)
3. Kalman gain:
   * 1. Statistical Independence(orthogonal) of the Error and the Estimate

* Orthogonal in Linear algebra

Let two vectors . Two vectors are orthogonal if the dot product of x and y,

1. Ex.

* Projection

<https://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/least-squares-determinants-and-eigenvalues/projections-onto-subspaces/MIT18_06SCF11_Ses2.2sum.pdf>

If we have a vector b and a line determined by a vector a, how do we find the point on the line that is closest to b?



The vector from the closest point vector on the line to the vector should be perpendicular to the line

Hence

The projection vector on is defined .

* The statistical orthogonal



* 1. Maximum likelihood estimator

(Skip-But Later will review this) Maximum likelihood estimator

* 1. The Discrete-Time Kalman Filter: Conditional Mean Estimator
* Problem: A dynamic system given as

1. We measure
2. The conditional estimator (the MV estimator) is
3. The propagated estimator is

* *The propagated estimator:*

= a Priori Estimator / a Predicted estimator

= measurements are previous compared to a posteriori estimator







* *In general text book notations are as follows*
* The Estimator
* The Predictor 🡪 the upper accents are different ~~

1. Propagating the conditional mean = a priori = prediction

* Remember , not

1. Propagating the variance of the a priori
2. Updating the Conditional Mean: a posteriori

* Remember not

1. Updating the covariance: a posteriori

* Define: **Residual 🡪 innovation sequence**

From (3.44)

* Define Kalman Gain: the residual gain :

Hence (3.44.1) is

* Orthogonality Properties of the Conditional Mean Estimator

1. The Innovation Sequence is independent for each

Since, the mean of ,

And,

with these facts, therefore the covariance is

which implies the gaussian are independent.

1. Error is orthogonal to ,

* Example 3.9 (estimating the speed of a car)
* Problem: estimate traveled distance by car.

1. The speed of car = 55 mi/hr, for 1 hour, implies the distance is 55 mi
2. The trip meter shown 55.3 mi

What is the best estimated distance with two information?

* Method

1. The system is modelled as
2. We may guess
3. The variance by the measurement: the quantization of the trip meter is = 0.1(mi) 🡪 The variance of the trip meter ,
4. The variance by the process: the velocity varies +/- 1 (mi/hr), the variance of the process
5. The best estimator of the distance in mean square sense,
6. Comments:  
   - the variance of a uniform RV in (a,b) is

-. Compare