1. Random Variables and Stochastic Process

**The following materials are also in Week\_3\_Chapter\_2 up to the end of Chapter 2.**

* **Prop. 2.29. Uncorrelated Gaussian random variables are independent**

Theorem 2.30. If is a Gaussian random vector with mean , and covariance, , and if , where is a Gaussian random vector with zero mean and covariance, , then is a Gaussian random vector with mean, , and covariance, .

* **Theorem 2.30**

**A R.V , another R.V. and they are independent**. Find mean and covariance of

%%% Kim’s comment : Characteristic function is difficult to remember. In the text book, using the characteristic method. In this case we may apply basic theory.

Sol: Let’s apply the basic definition.

Hence

* In general, independency implies the uncorrelated, not vice versa
* However, in Gaussian Does satisfy the opposite direction. %%%

%%% Kim’s comment: covariance matrix

Sometimes, but most case in this course, we may deal with a random vector whose components are random variables , i.e.

is a random vector, its components are random variables Then the covariance of random vector is defined ad

where

hence by definition

Therefore the matrix is a symmetric matrix, i.e.,

The diagonal terms of the covariance matrix are variance of each random variable

%%%

* The covariance of a uncorrelated (so independent) Gaussian is a diagonal matrix,

%%% Kim’s comment :Linear matrix theory: similar transform

For any semi-positive symmetric matrix , there is a **similar transform matrix** such that

Hence the covariance for any gaussian Random vectors (correlated), there exits a such that

* Any Gaussian Random vectors, we can find a transformed Random Vectors which is uncorrelated (independent).
* Independency is important to calculate the probability. You know the Gaussian probability table, but it is a scalar. So it you want to calculate the joint probability which may be correlated, first find a similar transform matrix to generate a diagonal covariance matrix. Then you may calculate the joint probability as a separate probability.

%%%

* **The central limit theorem**

Theorem 2.31. Let be i.i.d. random variables with finite mean and variance,

and denote their sum as . Then the distribution of the normalized sum

is a Gaussian distribution with mean 0 and variance 1 in the limit as

* Proof : textbook P.52
* Remarks:

1. See, the condition, that means   
   the mean and the variance is constant, but the experiment is many time processing. For example,
2. A die, which is fair or not, you roll the same die many times. Then the mean of the sum () is a Gaussian if .
3. Some RV has no mean, then it will not be applicable.
   1. Conditional Expectations and Conditional Probabilities

%%% Kim’s comment in 2021\_Week\_1

In 2021\_Week\_1, Definition of Conditional probability: the conditional probability of A given B

In 2021\_Week\_3. In the example of marginal probability, there is a diagram

Hence If a people has high pulse rate, what is the probability of temperature high?

Here the conditional probability as

To get the conditional probability, the marginal probability is calculated to

Hence

**What is important?**  %%%

* The conditional expectation
* Remarks
* is a constant, means it is not random variable.
* if is a constant, then is a constant
* if is a RV, then is a **Random Variable** of y
* **Iterated expectation** **(See the proof at p.57 and remember)**

Or if R.V. Y is discrete as

%%% Kim’s comment

Even if we do not know .

**I should say, this formula cannot emphasize too much!** This very simple fact uses diverse applications such as big data, machine learning, and dynamic system analysis. We should **remember** this.

%%%

%%% Example in <https://en.wikipedia.org/wiki/Law_of_total_expectation> (discrete case)

If you purchased a bulb, you would like to expect the life time of the bulb. Now you may collect the information as

1. Information:

-There are only two factories X and Y. You do not know the bulb’s supplier.

-The expectation of bul’s life time of X, (hr), and similarly

-The market share of X = 0.6, i.e., the probability of the bulb’s by X, and similarly

2) The solution

%%%%

%%% Kim’s Comment : continuous case

In 2021\_Week\_3\_Ch2 , the joint pdf

and the marginal pdf are

Hence

Hence

%%%%

* Lemma 2.34.
  1. Stochastic Process
* Def. 2.36. A stochastic process is a family of random variables, , indexed by a real parameter and defined on a common probability space .

%%% Kim’s comment

A stochastic process (or random process) is a time varying random variable, i.e., for any fixed , the process is a random variable.

%%%

* Ex. 2.37
* Def. 2.38.

1. A stochastic process is said to be continuous in probability at t if

for all

1. Skip: A stochastic process is said to be separable if there exists a countable, dense set such that for any closed set

differ by a set such that

* Skip: Theorem 2.40. The rational numbers in provide a separating set S.
* Def. 2.42. Let X be a random process defined on the time interval, T. Let

be a partition of the time interval, T. If the increments, are mutually independent for any partition of T, then X is said to be a process with **independent increments**.

* Def. 2.43 We say that a random process, X, is a Gaussian process if for every finite collection, the corresponding density function,

is a Gaussian density function.

* Def. 2.44 We say that a random process X is a Gaussian process if every finite linear combination of the form

is a Gaussian random variable

* Def 2.45. A random process, where T is a subset of the real line, is said to be a **Markov process** if for any increasing collection

or, equivalently

* 1. Gauss-Markov Processes – **The fundamental**

1. Dynamics

* State , is a known matrix, is a Gaussian Random sequence.

1. Given Conditions
2. Noise

where

1. The states

1. The correlation

which implies

1. The mean and covariance

* The mean
* The covariance
  1. Non-linear Stochastic Difference Equations 🡪 skip

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

a

b

d

c

Or you may plot then calculated.

%%%%%%

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

Let measurement , , are independent. Find the the minimum variance estimate

* a priori information,
* a noise,
* In order to find , so need to find

Since , need

1. First find

Since are independent and ,

* ,
* ,
* ,
* ,
* **What is**

1. Second find and

Since ,

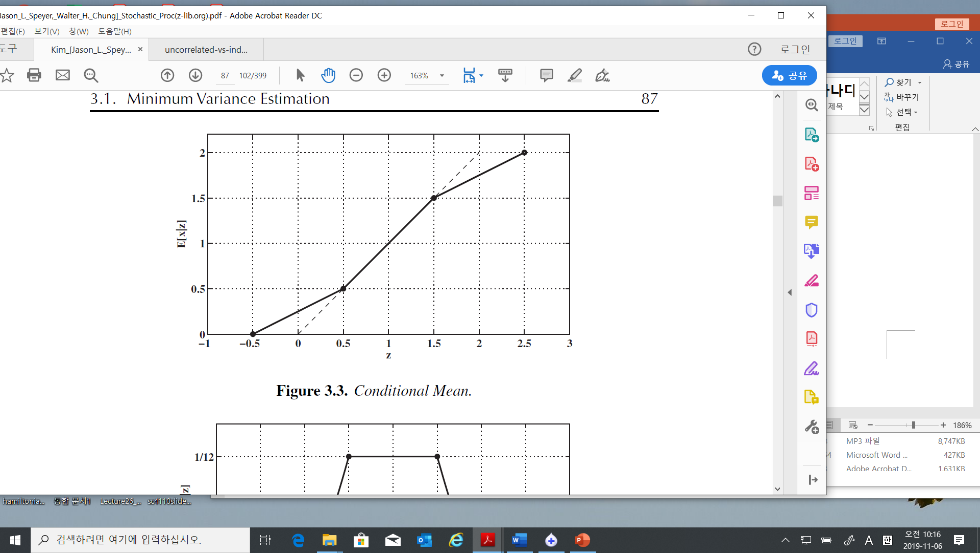
* For , from 1) and
* In addition,
* For , from 1) and
* In addition,
* For , from 1) and
* In addition,

1. Comparison between and ,

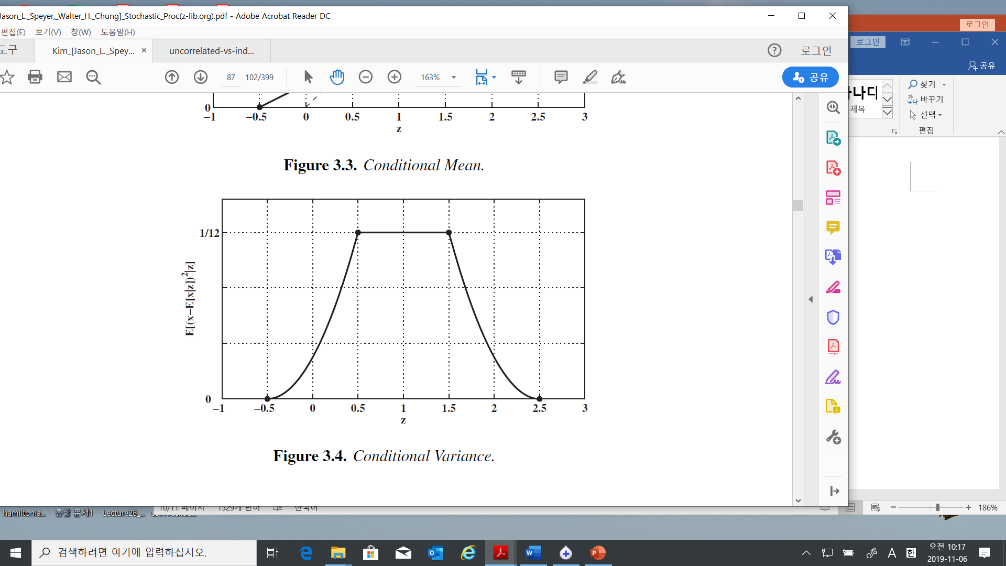
* First

1. Second and

* For ,
* For ,
* For ,
* Analysis



E[x] = 1



But

* 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.
* Scaler case

* 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
    3. 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