Stochastic representation

We use Random variables (RVs) for representing our knowledge about certain variables of systems.

We use Probability Density Functions (PDF) for describing the statistical properties of RVs .

When a RV ξ is statistically described by a PDF $f(\xi)$ we usually express that fact as: $\xi \sim f(\xi)$

A PDF, in the particular case of a one-dimensional RV (here we name it ξ) is a function which satisfies the following characteristics

Stochastic representation

Given a scalar RV ξ which is statistically described by a PDF $p_\xi(\xi)$; that PDF satisfies the following:

$$\xi \sim p_\xi(\xi)$$

$$p_\xi(.) : \mathbb{R}^1 \to \mathbb{R}^1$$

$$p_\xi(\xi) \geq 0, \quad \forall \xi \in \mathbb{R}^1$$

$$\int_{-\infty}^{\infty} p_\xi(\xi) \cdot d\xi = 1$$

Stochastic representation

If, given a PDF $p_\xi(\xi)$ which models the statistical properties of the RV ξ $(\mathrm{i.e.}\ \xi \sim f(\xi))$, we want to know the probability of ξ having its value contained in the interval between A and B, then that probability is provided by the following operation,

$$p\left(\xi \in [a,b]\right) = \int_a^b p_\xi(\xi) \cdot d\xi$$

Stochastic representation

It is the probability of the RV having its actual value in that region $[a, b]$

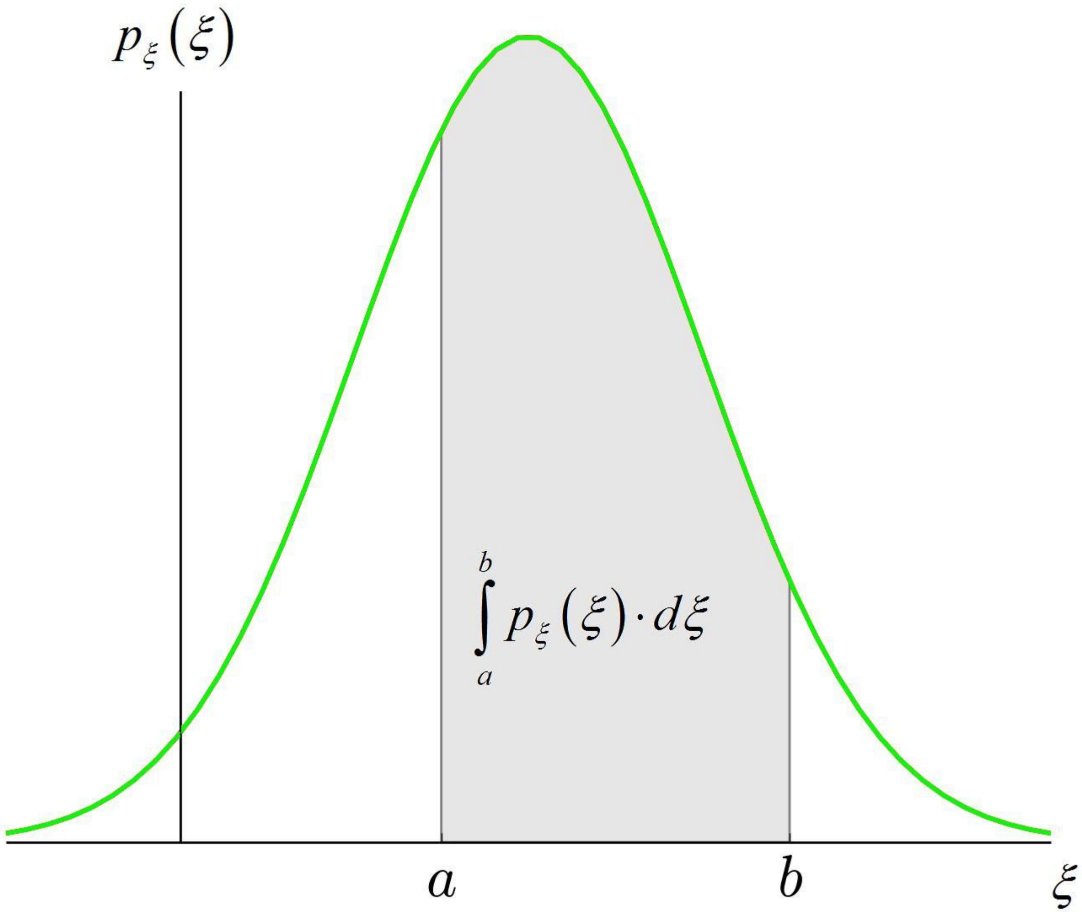
$$p\left(\xi \in [a,b]\right) = \int_a^b p_\xi\left(\xi\right) \cdot d\xi$$

Example of $p(\xi \in [\mathrm{a}, \mathrm{b}])$,

"The probability that the temperature of this room, tomorrow at 2PM will be in the range 25C<T<30C, is 0.3 (30%)."

It is the probability of the RV having its actual value in that region $[a, b]$

$$p\left(\xi \in [a,b]\right) = \int_a^b p_\xi(\xi) \cdot d\xi$$



consequently,

$$\int_{-\infty}^{+\infty} p_\xi(\xi)\cdot d\xi = 1$$

Or expressed this way:

$$\int_{\text{Domain of } \xi} p_\xi(\xi)\cdot d\xi = 1$$

Multivariate PDFs

PDFs are used for describing RVs.

Estimates of the states of a system can be treated as being RVs

We can express our belief about those estimates through a PDF.

An estimate of a state vector is usually a multidimensional variable (or a vector whose components are scalars, each of them being a scalar RV.)

These multiple RVs need to be jointly described.

A PDF that describes multiple scalar RVs is said to be a multivariate PDF.

This PDF is called the JOINT PDF of those scalar RVs.

“We can express our belief about those states estimates through a PDF”

From our perspective that hidden state is a RV,
because we do not know its exact value,
thus, we have “belief “about it,
which we model through a PDF

Multivariate PDFs

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T \in \mathbb{R}^n$$

$$\mathbf{x} \sim p_{\mathbf{x}}(\mathbf{x})$$

$$p_{\mathbf{x}}(.) : \mathbb{R}^n \to \mathbb{R}^1$$

We can also express it in this way:

$$p_{x_1, x_2, .., x_n}(x_1, x_2, .., x_n)$$

Multivariate PDFs

A Multivariate PDF does also satisfy:

$$p_{x_1,x_2,..,x_n}\left(x_1,x_2,..,x_n\right)\geq 0,\quad \forall\left(x_1,x_2,..,x_n\right)\in\mathbb{R}^n$$

$$\int_{-\infty}^{+\infty}..\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{x_1,x_2,..,x_n}\left(x_1,x_2,..,x_n\right)\cdot dx_1\cdot dx_2\cdot..\cdot dx_n = 1$$

$$p_{\mathbf{X}}\left(\mathbf{X}\right)\geq 0,\quad \forall\mathbf{X}\in\mathbb{R}^n$$

$$\int_{\substack{\text{domain}\\ \text{of } \mathbf{X}}} p_{\mathbf{X}}\left(\mathbf{X}\right)\cdot d\mathbf{X} = 1$$

Multivariate PDFs

Example?

Probability about the vehicle position, in 2D.

$$p_{x,y}\left(x, y\right)$$

Case: Probability about the vehicle pose, in 2D.

$$p_{x,y,\varphi}\left(x,y,\varphi\right)$$

What you are assuming as a “deterministic” variable in Project 1 would need to be represented as a RV and expressed via a PDF.

How to obtain that PDF?

That will be our task/goal in lecture 4.

Multivariate PDFs

$$p\left(\mathbf{x} \in \Omega\right) = \int_{\Omega} p_{\mathrm{X}}\left(\mathbf{x}\right) \cdot d\mathbf{x}$$

(This is a definition/concept.

We do not usually perform this calculation)

Discrete variables: Probability mass.

Discrete variable? $$\xi \in \Phi = \left\{ \xi_1, \xi_2, \ldots, \xi_N \right\}$$

examples of discrete variables,

The age, expressed in entire years, of the students in this class.

The gender of the students in this class.

The nationality of the international students, in this class.

Discrete variables: Probability mass.

Suppose we define the random variable x to describe the outcome of throwing a dice.

We have 6 possible results, $x \in \{1,2,\ldots,6\}$
Their probability masses are uniform (for a perfect dice and environment):

$$p_x(1) = p_x(2).. = p_x(6) = 1/6$$

Now, we decide to implement a "2-steps dice" in which we generate the variable z, which is randomly obtained as follows:
$z = x + y$, being x and y generated by throwing a dice, individually. Now z is a discrete RV, for which we propose a set of hypotheses $z \in \Phi = \{1,2,\ldots,12\}$
(we do not need considering z=1, but we include it in this example)

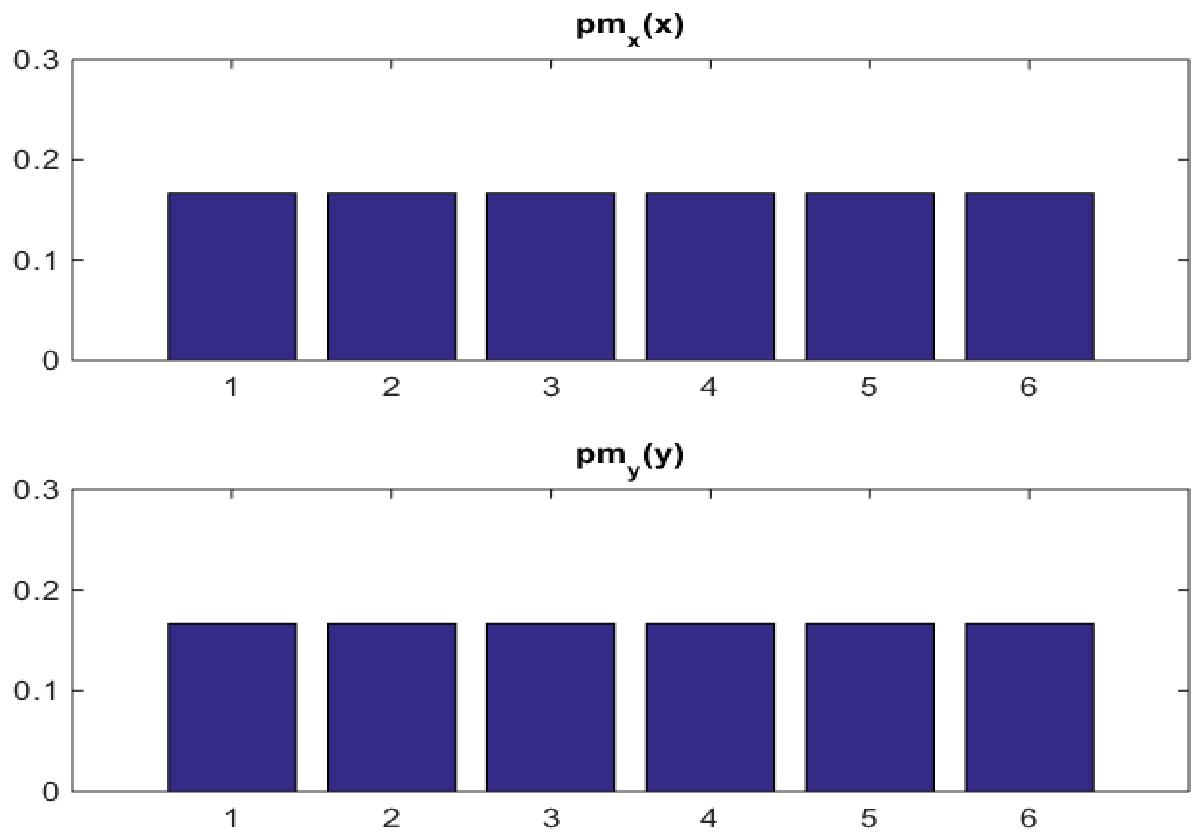
How would be the PM's of this RV?

Discrete variables: Probability mass.

Suppose we define the random variable x to describe the outcome of throwing a dice.

We have 6 possible results, $x \in \Phi = \{1,2,\ldots,6\}$
Their probability masses are uniform (for a perfect dice and environment):

$$p_x(1) = p_x(2)\ldots = p_x(6) = 1/6$$

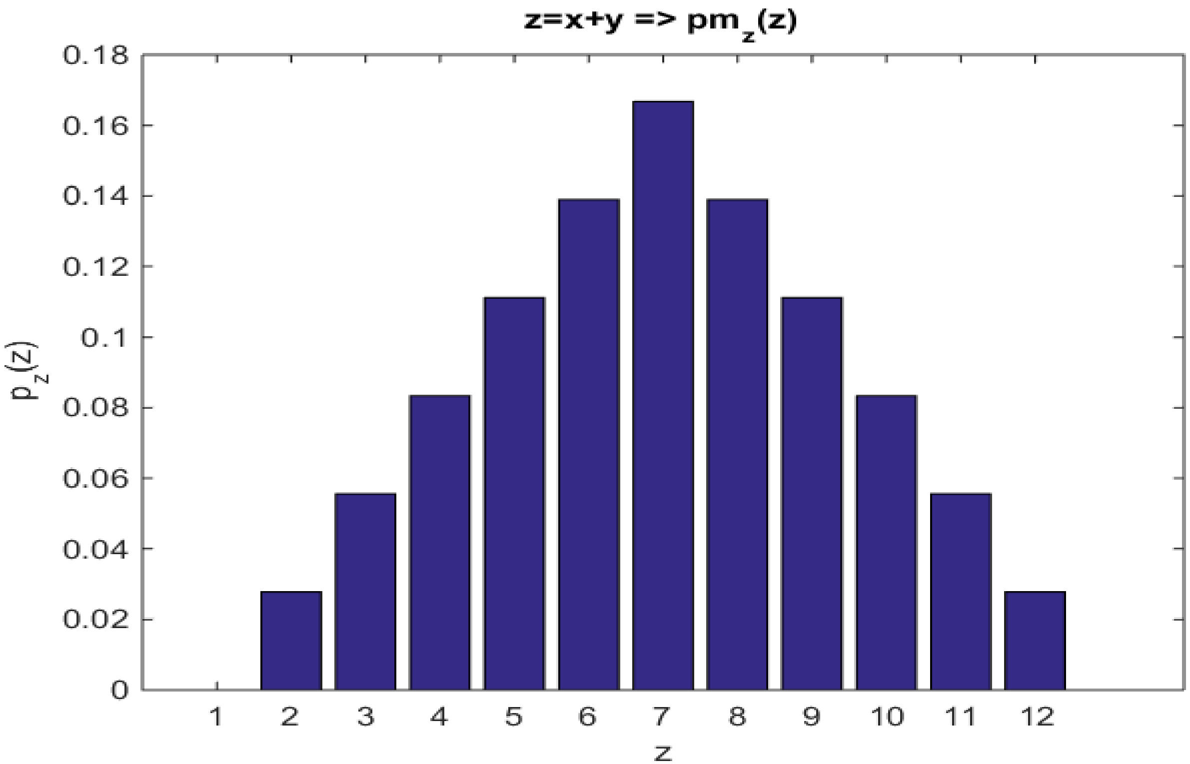


Discrete variables: Probability mass.

"2-steps dice"

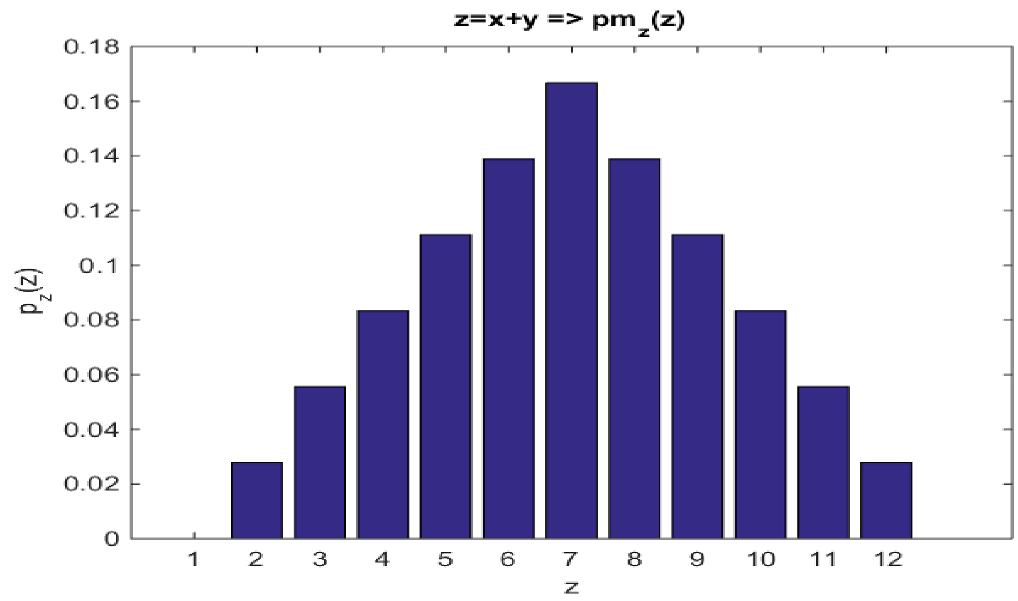
$z = x + y,$

(having been *x* and *y* generated by throwing a dice, individually.)



It is easier visualizing matters with PM functions, for instance the probability of one of hypotheses in a subset .

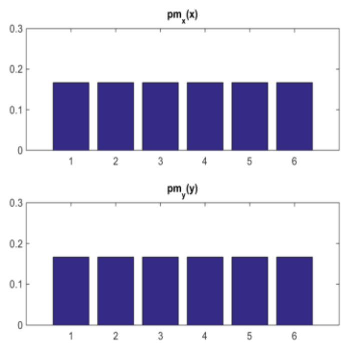
$p_z(z)$: How de we get it?

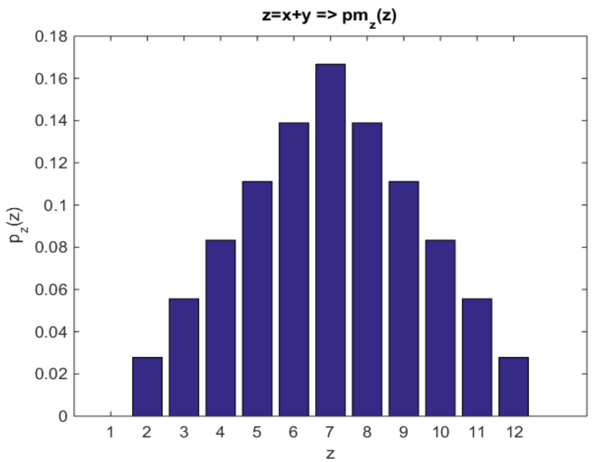


It is the result of "fusing" the PMs and the model "$z = x + y$"

$$p_x(x), p_y(y)$$

$$\left\{ \begin{array}{c} p_x(x) \\ p_y(y) \\ z = x + y \end{array} \right\}$$





$$p_z(z)$$

continuous RV

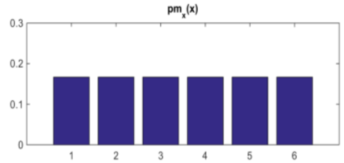
$$p\left(\mathbf{x} \in \Omega\right) = \int_{\Omega} p_{\mathrm{x}}\left(\mathbf{x}\right) \cdot d\mathbf{x}$$

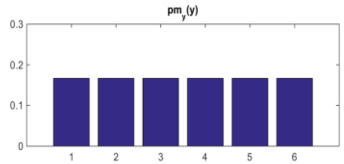
discrete RV

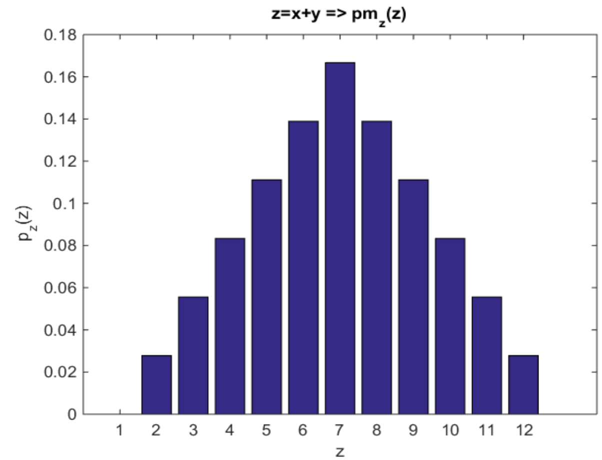
$$p\left(\xi \in \Omega\right) = \sum_{\xi_i \in \Omega} p_{\xi}\left(\xi_i\right)$$

(We end our discussion about PMs, here. We focus on continuous RVs)

$$\left\{ \begin{matrix} p_x(x) \\ p_y(y) \\ z = x + y \end{matrix} \right\}$$







$$p_z(z)$$

Why are we talking about this?

Comment: model, input variables, uncertainty. → estimate / belief /guess

Model like our imperfect kinematic model?
And variables such those inputs whose measurements are polluted by noise?
And that, with that, we need to estimate the platform's pose
(→ to see NEXT WEEK, not now)

Definition: **Marginal PDF** **(again continuous RV)**

Even if we have the joint PDF about a set of RVs, we may ask for the description of an individual variable or for certain subset of the variables. In that case we can *marginalize* the PDF.

For the "easy" case of having the joint PDF of 2 scalar RVs, the mathematical definition is shown by the following expressions:

$$p_x(x) = \int_{-\infty}^{+\infty} p_{x,y}(x,y) \cdot dy, \qquad p_y(y) = \int_{-\infty}^{+\infty} p_{x,y}(x,y) \cdot dx$$

…Marginal PDF

Higher dimensions can be considered as well; here we express it in a general way, using vectors (and "volume" differentials).

Given the joint PDF $p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{X},\boldsymbol{Y})$
We want to obtain the marginal PDF $p_{\boldsymbol{X}}(\boldsymbol{X})$

$$p_{\mathbf{X}}\left(\mathbf{X}\right)=\int\limits_{\substack{\text{domain}\\ \text{of } \mathbf{Y}}} p_{\mathbf{X},\mathbf{Y}}\left(\mathbf{X},\mathbf{Y}\right)\cdot d\mathbf{Y}$$

…Marginal PDF

Case in which we use scalar RVs, and have the joint PDF $p_{x,y,z}(x,y,z)$

But we have interest only on *x*.

So, we can obtain the PDF $p_x(x)$

By marginalization:

$$p_x(x) = \int\limits_{\substack{\text{domain}\\ \text{of } (y,z)}} p_{x,y,z}(x,y,z) \cdot dy \cdot dz$$

(we cannot infer the full joint PDF, from a marginal one)

(we **cannot** infer the full joint PDF just from its associated marginal ones)

$p_x(x), p_y(y), p_z(z)$

Definition: Conditional PDF

Given our joint belief about the RVs $(\mathbf{x}, \mathbf{y})$, $p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})$, what would be our belief about $\mathbf{x}$ if we knew that the actual value of $\mathbf{y}$ is $\mathbf{y}_0$?

We express it as follows:

$$p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x} \mid \mathbf{y} = \mathbf{y}_0\right)$$

Note that this PDF is about the RV $\mathbf{x}$

“Probability of x provided that y=y0” (← conditional)

Conditional PDF

In general, given a joint PDF about the RVs $(\mathbf{x}, \mathbf{y})$, $\mathrm{p}_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})$, we can obtain the conditional PDF $\mathrm{p}_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})$ as follows:

$$p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x} \mid \mathbf{y}\right) = \frac{p_{\mathbf{x},\mathbf{y}}\left(\mathbf{x}, \mathbf{y}\right)}{p_{\mathbf{y}}\left(\mathbf{y}\right)}$$

Similarly

$$p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y} \mid \mathbf{x}\right) = \frac{p_{\mathbf{x},\mathbf{y}}\left(\mathbf{x}, \mathbf{y}\right)}{p_{\mathbf{x}}\left(\mathbf{x}\right)}$$

$$p_{\mathbf{x},\mathbf{y}}\left(\mathbf{x},\mathbf{y}\right)$$

$$p_{\mathbf{x}}\left(\mathbf{x}\right)$$

$$p_{\mathbf{y}}\left(\mathbf{y}\right)$$

$$p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x}\mid\mathbf{y}\right)$$

$$p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}\mid\mathbf{x}\right)$$

Definition: statistical independence

We say that two RVs, $(\mathbf{x}, \mathbf{y})$ (which are statistically described by the joint PDF, $p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})$, are independent, if their joint PDF can be factorized as follows,

$$(\mathbf{x}, \mathbf{y}) \sim p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{x}}(\mathbf{x}) \cdot p_{\mathbf{y}}(\mathbf{y})$$

(independent RVs..)

That condition implies that their conditional PDFs are equal to their marginal ones, i.e.

$$p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x} \mid \mathbf{y}\right) = \frac{p_{\mathbf{x},\mathbf{y}}\left(\mathbf{x}, \mathbf{y}\right)}{p_{\mathbf{y}}\left(\mathbf{y}\right)} = p_{\mathbf{x}}\left(\mathbf{x}\right)$$

$$p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y} \mid \mathbf{x}\right) = \frac{p_{\mathbf{x},\mathbf{y}}\left(\mathbf{x}, \mathbf{y}\right)}{p_{\mathbf{x}}\left(\mathbf{x}\right)} = p_{\mathbf{y}}\left(\mathbf{y}\right)$$

(independent RVs..)

Any knowledge we may have about Y would not affect our "belief" about X.

A situation such this may be beneficial or not, depending on the cases in which the PDF is being used.

Any example about statistical dependency?

Example of dependent and Independent Random Variables

(see lecture notes, example about certain “rare lottery game”)

For a more natural understanding we consider the case of a discrete RV.

Definition: **“White noise”**

Consider a continuous time “signal” $w(t)$.

It has the characteristic that there is no statistical dependency between any couple $w(t1)$ and $w(t2)$, except for $t1=t2$.

(if we knew something about $w(t)$ at time $t1$, that information would not help, at all, to guess the value of $w(t2)$)

$w(t1)$ and $w(t2)$ are independent, for all $t1,t2$.

Certain noises behave in that way, or close to it.

Definition: **“White noise”**

Let’s generate a “white signal”, in discrete time and for a discrete variable.

We have a container with numbered/labeled balls, from 0 to 9.

we repeat 5 times this procedure :

{ Rotate/shake the container

Extract one ball randomly,

The number of the label is read and recorded (but kept hidden to us) , the ball put it again in the container.

}

So, we got 5 values, {x(k)}

Would we, knowing x(2) help us to guess x(4)? (i.e., refine our belief about x(4))

A: no. →que sequence x(k) is white

Definition: **“colored noise”**

Similar procedure, except we know that we usually forget to return the extracted ball to the container.

Would we, knowing x(2) help us to guess x(3)?

Belief about x(3) if we do not know x(2)

Belief about x(3) if we do know x(2)

Are those beliefs equal? NO.

Definition: **“White noise”**

Consider a discrete time “signal” $w(k)$ (sequence).

Which has the characteristic that there is no statistical dependency between any couple $w(k1)$ and $w(k2)$, except for $k1=k2$.

“White noise”

Example

Quantization error of a 12 bits ADC (Analog to Digital Converter)

Suppose we use an ADC for taking samples of a voltage which fluctuates in the range [0, 10]volts, and that the input range of the ADC is also that range.

$$p_e(e) = \begin{cases} a & \forall e \in [-W, +W] \\ 0 & \forall e \notin [-W, +W] \end{cases}, \qquad W = \frac{10}{2^{12}} \cdot \frac{1}{2} = \frac{10}{4096} \cdot \frac{1}{2},$$

$$a = (2 \cdot W)^{-1} = \frac{4096}{10}$$

This PDF is uniform in that small interval.

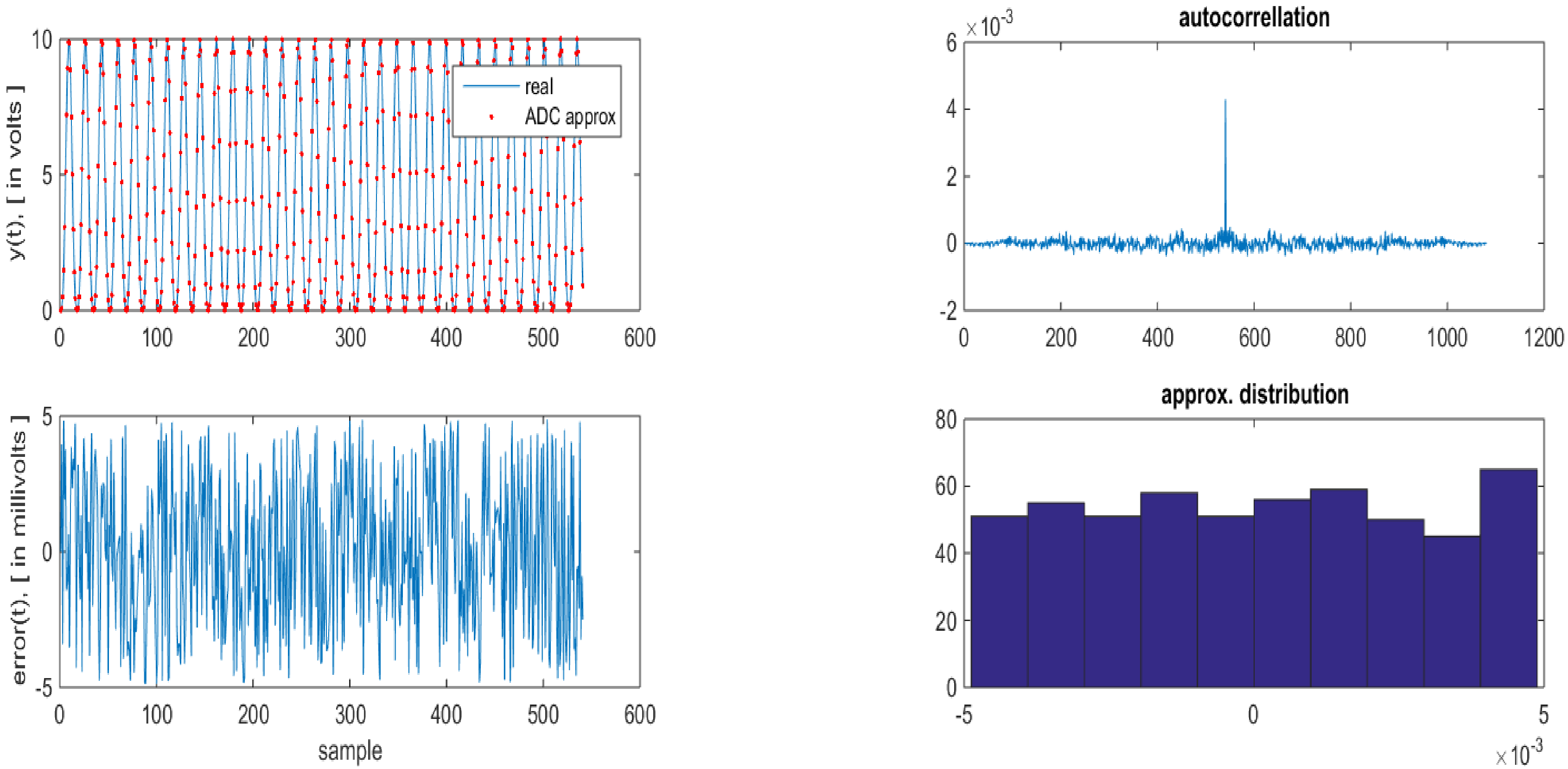
Each of the errors, e(k) follows that PDF

But, is the error signal white?

Strict answer : NO.

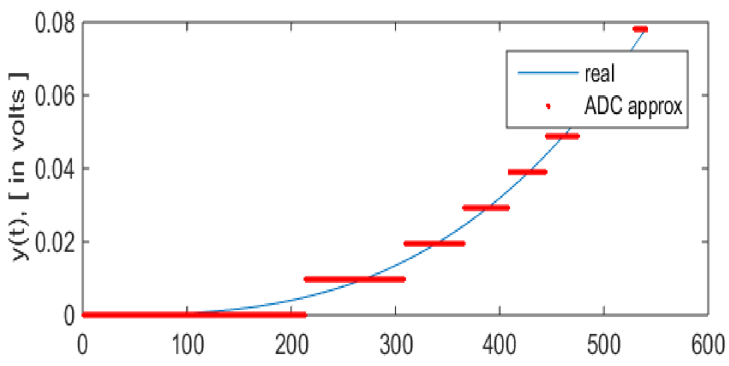
Quantization error . Is it white when we sample a signal?

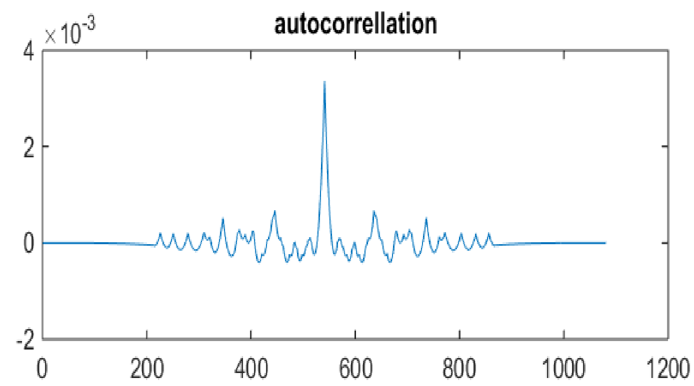
In many practical cases, we can assume it is white.

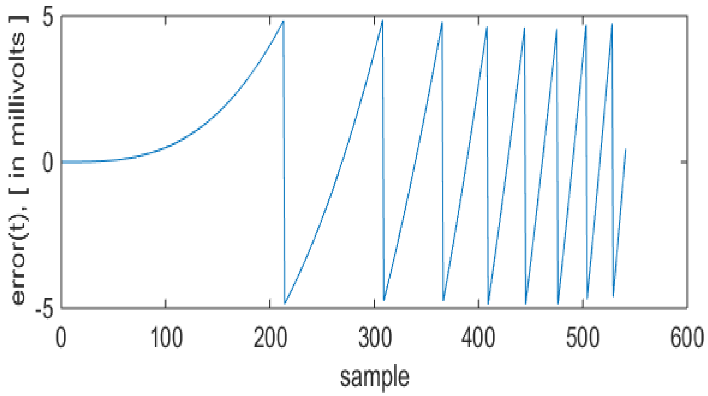


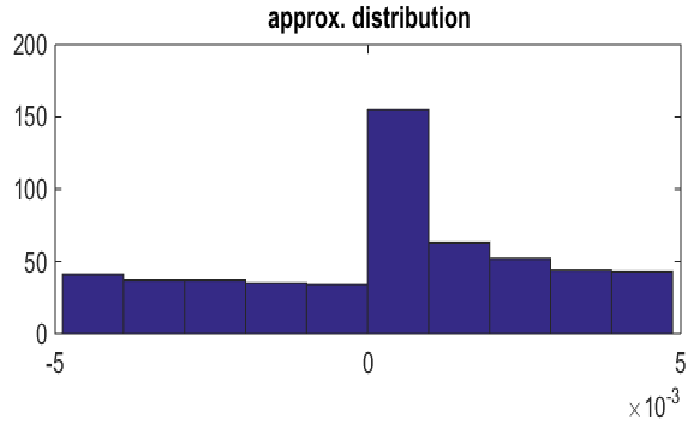
Quantization error . Is it white when we sample a signal?

Slow or constant signals being sampled by the ADC, may result in non-white quantization error.









Quantization error . Is it white when we sample a signal?

If the signal being sampled is constant or slowly time variant

$e(k)$ and $e(k+n)$ will be statistically dependent.

So, the sequence $\{e(k)\}$ is well “colored”.

Case 2:

Suppose we use the same ADC to read a voltage signal, which is previously amplified by an amplifier. The amplifier introduces some unknown bias (offset) which is not changing (or it is changing slowly with time)

2.1) Should we consider that bias to be an uncertainty polluting the measurement?

2.2) Should we model that uncertainty through a RV?

2.3) Should we consider error sequence as being white?

Answers:

2.1) Yes, we consider it to be uncertainty, because it introduces an unknown discrepancy between real and assumed model.

2.2) Yes, we can model it through a RV, we can even propose some PDF because we usually know the technical specifications of the amplifier.

2.3) No, it is not white, because it is strongly correlated in time.

Error in GPS position estimates (from a standard GPS receiver).

It is not white noise.

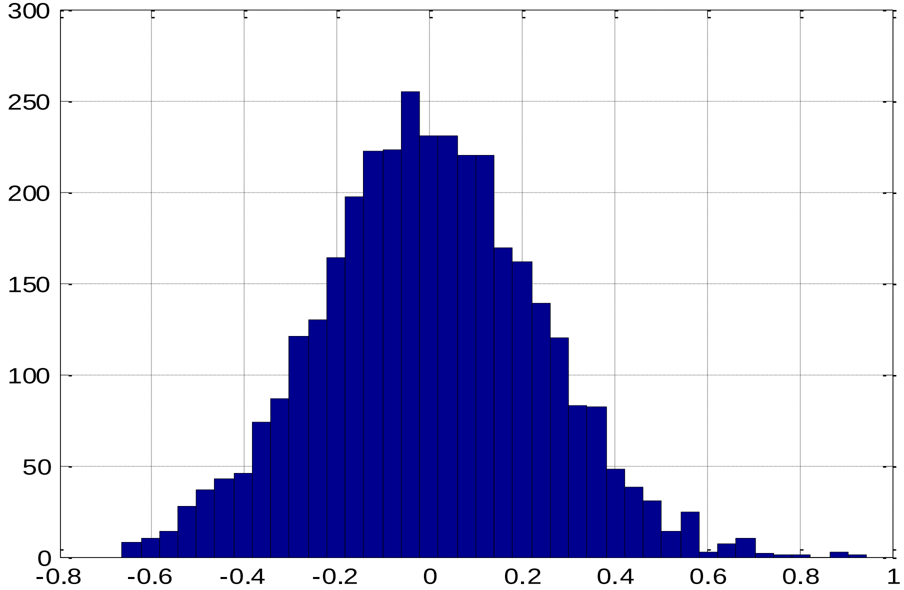
Why? the error in a position reading now, has usually strong correlation with the error which polluted other recent readings (e.g., due to slowly time varying atmospheric disturbances, and the slowly changing distribution of satellites visible to the GPS receiver, etc.)

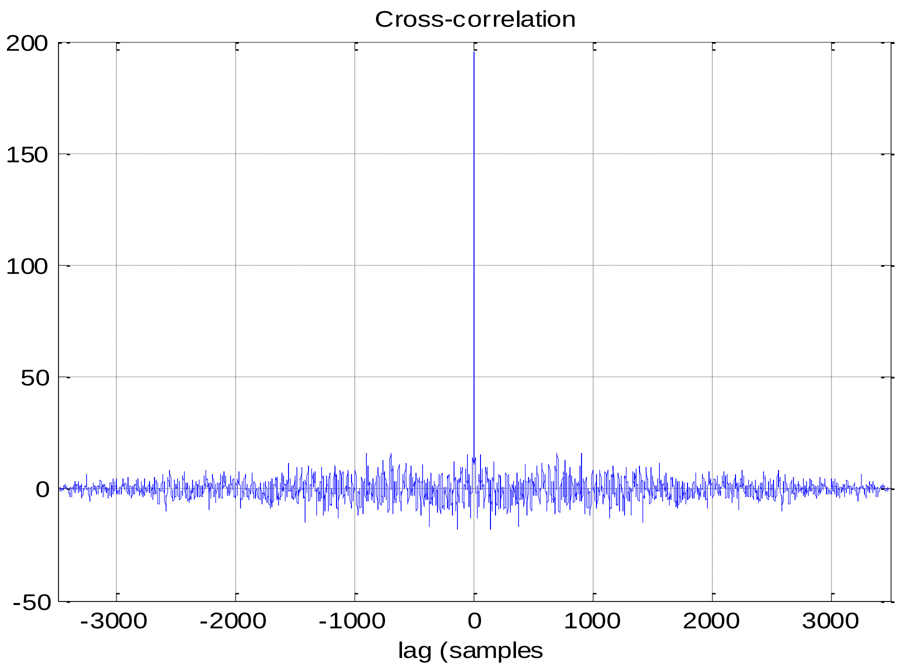
Consequence of noise not being white?

→ Averaging measurements would not help to remove error.

Case 3: Noise in gyroscope measurements

Bias + fluctuating noise (which is close to white, and it is also Gaussian.)





The Gaussian Representation

As in many other areas of engineering and sciences, where the models are simplified by using tractable approximations, we need, in certain cases, to approximate the PDF functions.

The Gaussian Distribution

Consider a scalar RV x. If it is said to be a Gaussian RV, then its PDF will have the following structure

$$p_x(x) = c \cdot e^{-\frac{(x-\hat{x})^2}{2\cdot\sigma_x^2}}$$

The Gaussian Representation

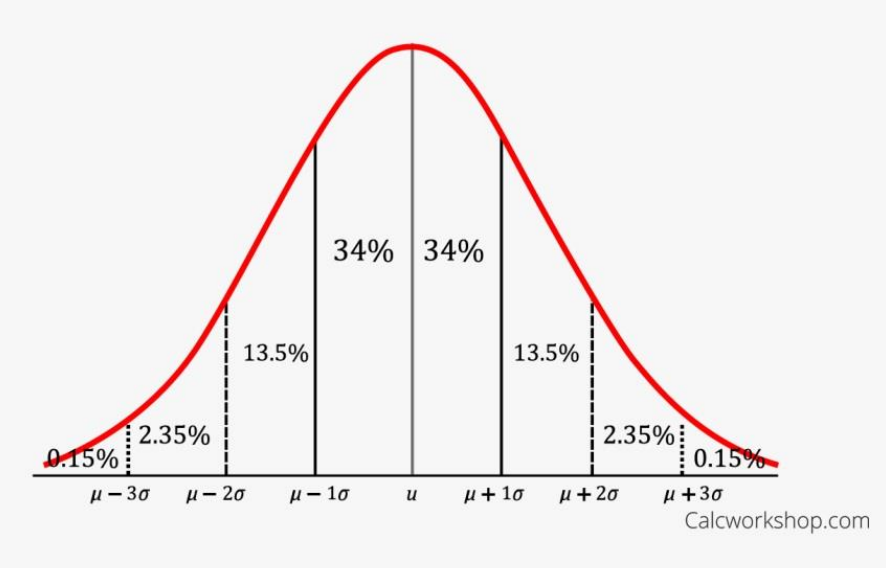
It has two parameters

$$p_x(x) = c \cdot e^{-\frac{(x-\hat{x})^2}{2\cdot\sigma_x^2}}$$

$\hat{x}$: expected value

σ_x^2 : variance

The Gaussian Representation



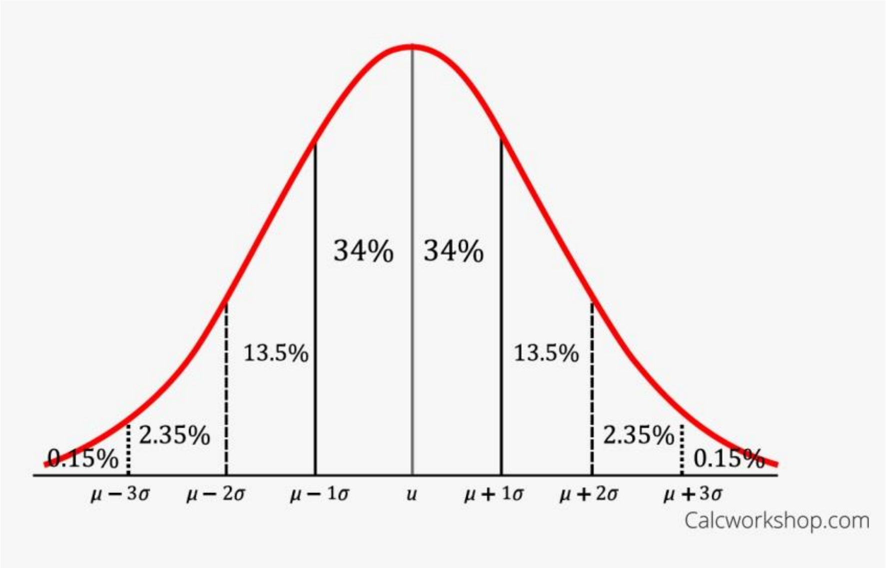
$$p_x(x) = c \cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}$$

μ : expected value

σ^2 : variance

c: : normalizing constant

The Gaussian Representation



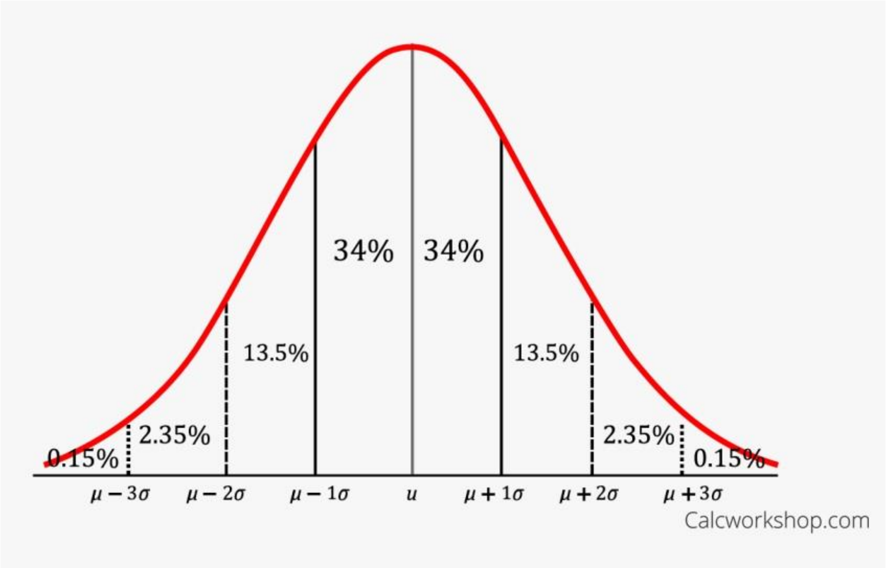
$$p_x(x) = c \cdot e^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}}$$

σ : standard deviation (what is telling us σ ?)

(confidence?)

The Gaussian Representation

$$p_x(x) = c \cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}$$



σ : what would mean $\sigma \to 0$?

$p_x(x) \to \delta(x-\mu)$ (Dirac Delta)

The Gaussian Representation

$$p_x(x) = c \cdot e^{-\frac{(x-\hat{x})^2}{2\cdot\sigma_x^2}}$$

It has two parameters

$\hat{x}$: expected value

σ_x^2 : variance

$$\hat{x} = \int_{-\infty}^{+\infty} x \cdot p_x(x) \cdot dx$$

$$\sigma_x^2 = \int_{-\infty}^{+\infty} (x-\hat{x})^2 \cdot p_x(x) \cdot dx$$

Multivariate Gaussians

2D case

$$p_{\mathbf{X}}\left(\mathbf{X}\right) = c \cdot e^{-\frac{1}{2}\cdot\left(\mathbf{X}-\hat{\mathbf{X}}\right)^T \cdot \mathbf{P}^{-1} \cdot \left(\mathbf{X}-\hat{\mathbf{X}}\right)}, \quad c = \frac{1}{\sqrt{\left(2 \cdot \pi\right)^n \cdot \left|\mathbf{P}\right|}}, n = 2$$

$$p_{\mathbf{X}}\left(.\right) : \mathbb{R}^2 \rightarrow \mathbb{R}^1$$

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \hat{\mathbf{X}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} P_{1,1} & P_{1,2} \\ P_{2,1} & P_{2,2} \end{bmatrix}$$

(covariance matrix)

$$p_{\mathrm{X}}(\mathbf{X}) = c \cdot e^{-\frac{1}{2}\cdot(\mathbf{X}-\hat{\mathbf{X}})^T \cdot \mathbf{P}^{-1} \cdot (\mathbf{X}-\hat{\mathbf{X}})}$$

in which

$(\mathbf{X}-\hat{\mathbf{X}})$: vector, 2×1

$(\mathbf{X}-\hat{\mathbf{X}})^T$: vector, 1×2

$\mathbf{P}$: matrix 2×2

$p_{\mathrm{X}}(\mathbf{X})$: scalar

$$p_{\mathrm{X}}(\mathbf{X}) = c \cdot e^{-\frac{1}{2}\cdot(\mathbf{X}-\hat{\mathbf{X}})^T \cdot \mathbf{P}^{-1} \cdot (\mathbf{X}-\hat{\mathbf{X}})}$$

in which

$(\mathbf{X}-\hat{\mathbf{X}})$: vector, $n \times 1$

$(\mathbf{X}-\hat{\mathbf{X}})^T$: vector, $1 \times n$

$\mathbf{P}$: matrix $n \times n$

$p_{\mathrm{X}}(\mathbf{X})$: scalar

The matrix ***P*** has special characteristics such as:

* It is a symmetric matrix

* It is a positive semi-definite matrix

Just these two parameters, expected value (a vector) and the covariance matrix fully define the PDF

So, we usually talk about these two parameters, implicitly knowing we are talking about a PDF.

Higher dimension cases?

Yes, and some of them of very high dimension

(thousands, or hundreds of thousands)

(not in our interest now)

We will try cases of 3 , 4 and 5D, in MTRN4010.

3D: 2D : pose!

4D: 2D Pose + longitudinal velocity

5D: 2D pose + speed + gyro bias

Higher dimension cases?

Flying vehicle:

9D (3D position + 3D attitude + 3D velocity vector)

Covariance matrix : 9x9

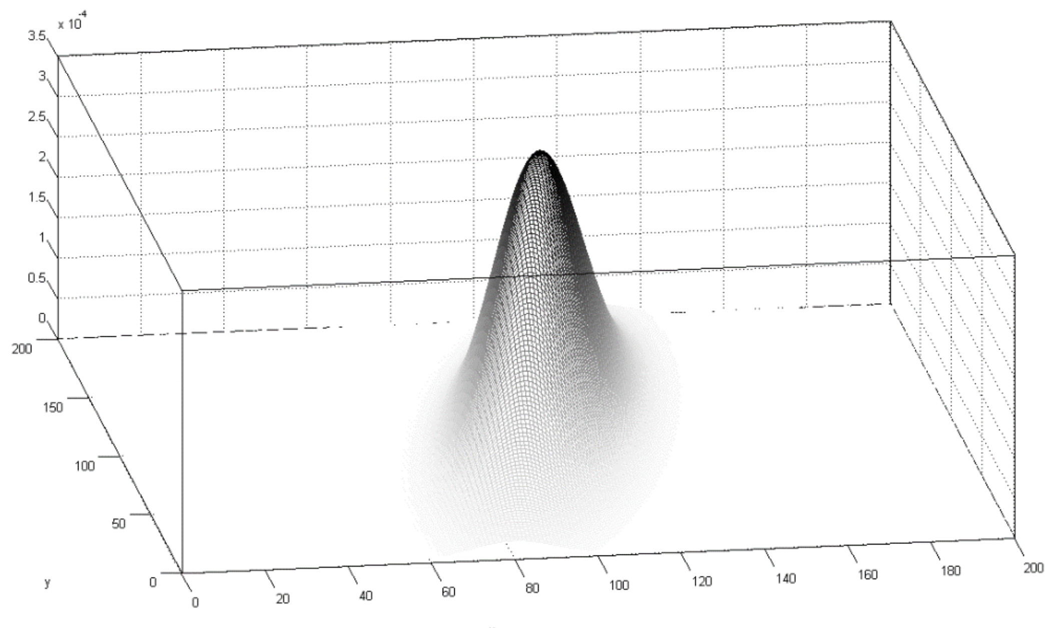
How do they look?

$$\hat{\mathbf{X}} = \begin{bmatrix} 1 \\ 3.5 \end{bmatrix}, \quad \mathbf{P}_{\mathrm{X}} = \begin{bmatrix} 2 & 0 \\ 0 & 9 \end{bmatrix}$$

How do they look?

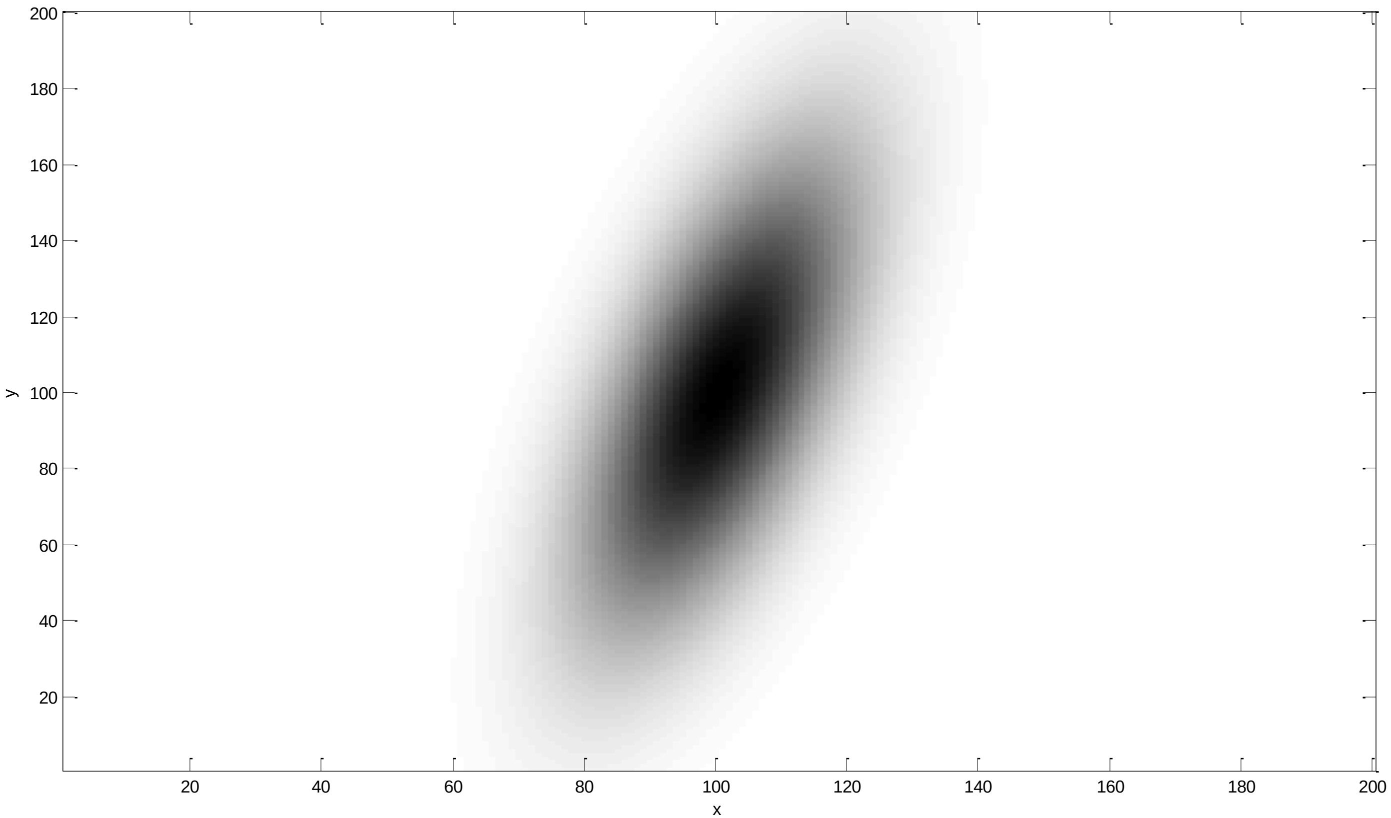
We can visualize 2D PDFs,

$$\hat{\mathbf{X}} = \begin{bmatrix} 100 \\ 100 \end{bmatrix}, \quad \mathbf{P}_{\mathbf{X}} = \begin{bmatrix} 400 & 752 \\ 752 & 3600 \end{bmatrix}$$



How do they look? / We can visualize 2D PDFs,

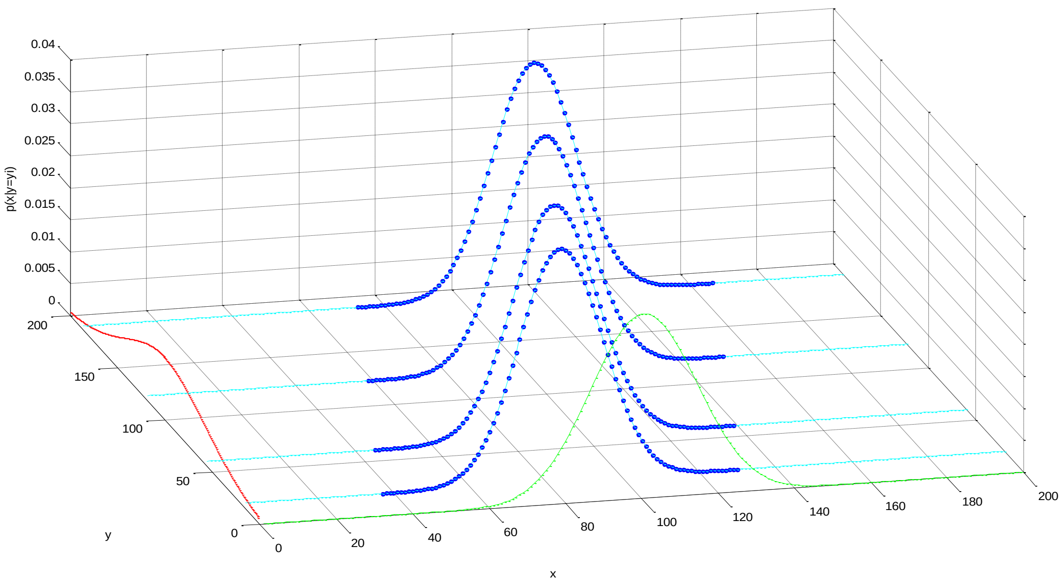
$$\hat{\mathbf{X}} = \begin{bmatrix} 100 \\ 100 \end{bmatrix}, \quad \mathbf{P_X} = \begin{bmatrix} 400 & 752 \\ 752 & 3600 \end{bmatrix}$$



Here I use grayscale : Dark regions correspond to high probability density

How do they look? / We can visualize 2D PDFs,

$$\hat{\mathbf{X}} = \begin{bmatrix} 100 \\ 100 \end{bmatrix}, \quad \mathbf{P}_X = \begin{bmatrix} 400 & 752 \\ 752 & 3600 \end{bmatrix}$$



Marginal PDFs, and some conditional ones.

Gaussian PDFs have properties which make them very efficient to be processed.

We will see an approach which exploits it.

(this is why we needed to refresh these concepts)

Next step in our work will be next week:

Bayesian Estimation

Gaussian Estimators.

KF / EKF (Kalman Filter / Extended Kalman Filter)