Advanced Autonomous Systems

Lecture 3.1 (Refreshing Concepts about Statistics)

Probability Distributions

We use Probability Density Functions (PDF) for describing the statistical properties of Random Variables (RV). 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,

$$\begin{gathered} \xi \sim p_{\xi}(\xi) \\ p_{\xi}(.) : \mathbb{R}^1 \rightarrow \mathbb{R}^1 \\ p_{\xi}(\xi) \geq 0, \quad \forall \xi \in \mathbb{R}^1 \\ \int_{-\infty}^{\infty} p_{\xi}(\xi) \cdot d\xi = 1 \end{gathered} \qquad \text{[E1]}$$

If, given a PDF $p_{\xi}(\xi)$ which models the statistical properties of the RV ξ, 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,

$$p(\xi \in [A, B]) = \int_{A}^{B} p_{\xi}(\xi) \cdot d\xi \qquad \text{[E2]}$$

(i.e. the density function is integrated in the interval of interest)

If we cover the complete domain of the RV, then the probability of the RV having its value contained in that range must be = 1 (i.e. 100%), as expressed in [E5] (i.e. it corresponds to the 1D (1 dimensional) case, i.e. $A \rightarrow -\infty, B \rightarrow +\infty$) .
Any RV, used for representing uncertainty in our models, will usually have a statistical description expressed through a PDF.

Note about interpretation: We usually use a RV for representing our belief about certain variable. This variable must actually have a value; the problem is that we do not know that value. We could have certain hypotheses about its value; then we represent our belief about it by using a PDF. We will discuss examples in class.

Example: We may ask a meteorologist: "What is the probability that tomorrow at 4PM, the temperature at this place will be in the range of values between 28C and 33C.?" His/her answer, after doing some calculations, would be given as a percentage or, alternatively, as a fraction of 1. Alternatively, we could have asked about the value of the temperature at that time and place, in that case the answer should have been given through a PDF $p_T(T)$.

To discuss in class
* We will see some basic examples of PDFs ("confident" PDFs, unimodal PDFs, typical PDFs, etc.), to get the concept.
* Probability Mass Functions (PMF: equivalent to PDF but for discrete RVs). Examples about Probability mass functions, in class.

Discrete variables: Probability Mass

For discrete variables, being those from a finite or infinite set.

$$\xi \in \Omega = \{\xi_1, \xi_2, ..., \xi_N\}$$

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

$$p_\xi(.) : \Omega \to \mathbb{R}^1$$
$$p_\xi(\xi) \geq 0, \quad \forall \xi \in \Omega$$
$$\sum_{\xi_i \subset \Omega} p_\xi(\xi_i) = 1$$

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.

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

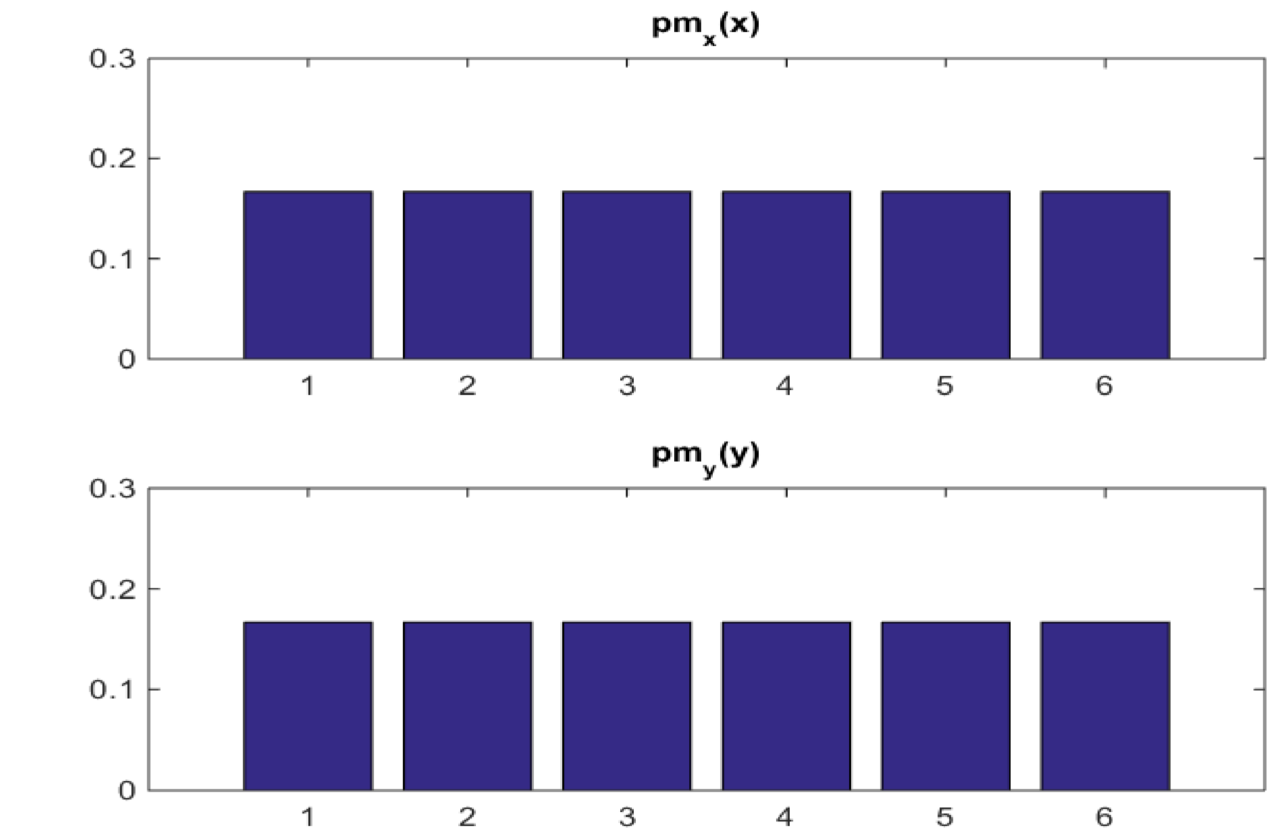
We have 6 possible results, $x \in \Omega = \{1, 2, ..., 6\}$

Their probability masses are uniform (for a perfect dice and environment):

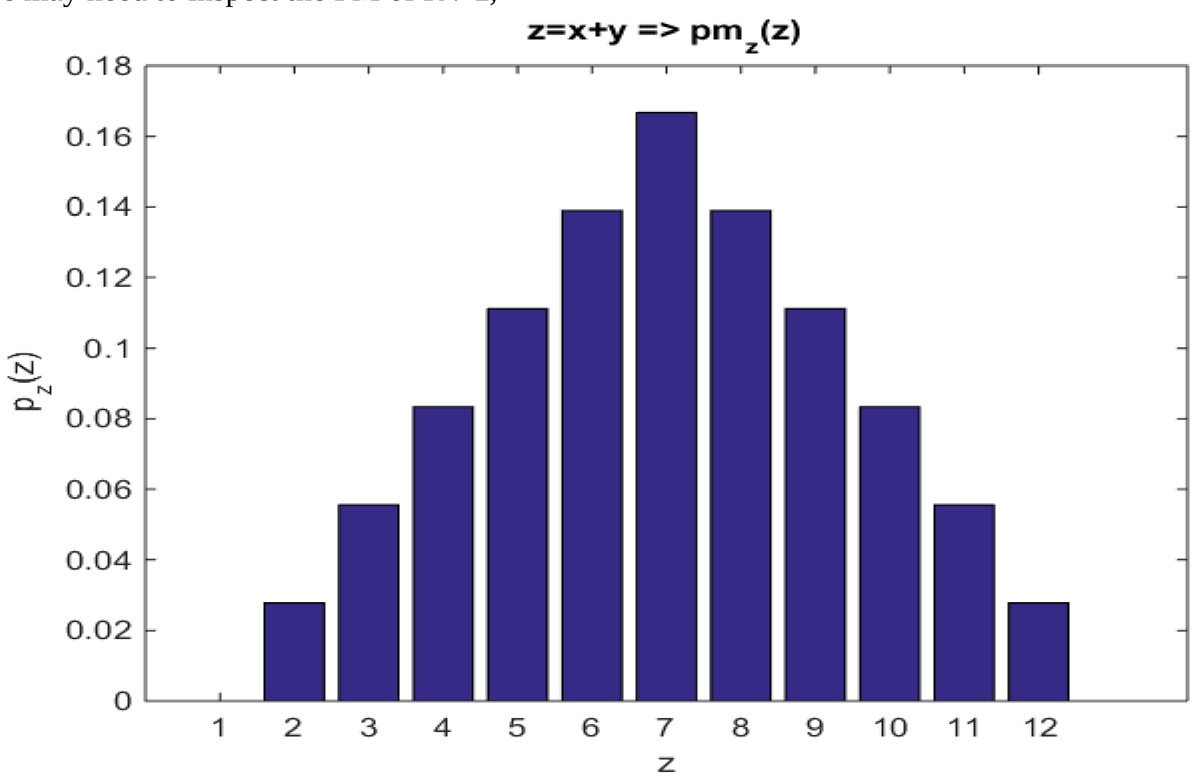
$p_x(1) = p_x(2).. = p_x(6) = 1/6$

We decide to implement an "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 $z \in \Omega = \{1, 2, ..., 12\}$ (we do not need considering z=1, but we include it in this example)



If we had to guess *z*, which hypothesis would we choose? (z from 1 to 12)
We may need to inspect the PM of RV z,



We do not have interest in discussing about how obtain it (exhaustive calculation manually, or by combinatorial analysis, or by a discrete convolution, etc). The purpose of this example is to show a PM in which the hypotheses have different probabilities.

Multivariate PDFs

PDFs are used for describing RVs. Estimates of the states of a system can be treated as being RVs; consequently, we usually 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 the RVs.

$$p_{\mathbf{X}}(\mathbf{X}) \geq 0, \quad \forall \mathbf{X} \in \mathbb{R}^n, \quad p_{\mathbf{X}}(.): \mathbb{R}^n \to \mathbb{R}^1$$
$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}^T \qquad \text{[E3]}$$

We can also express that PDF as a function of the set of scalar RVs,

$$p_{x_1,x_2,..,x_n}(x_1,x_2,..,x_n) \geq 0, \quad \forall (x_1,x_2,..,x_n) \in \mathbb{R}^n$$
$$\int_{-\infty}^{+\infty} .. \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{x_1,x_2,..,x_n}(x_1,x_2,..,x_n) \cdot dx_1 \cdot dx_2 \cdot .. \cdot dx_n = 1 \qquad \text{[E4]}$$

We may also express the integral as follows, to simplify the notation,

$$\int_{\substack{\text{domain}\\ \text{of } \mathbf{X}}} p_{\mathbf{X}}(\mathbf{X}) \cdot d\mathbf{X} = 1 \qquad \text{[E5]}$$

(in which *dX* is the “differential of volume” of the variable *X*)

The probability that the actual value of the variable X will fall in a given region ***Ω*** is:

$$p(\mathbf{X} \in \Omega) = \int_{\Omega} p_{\mathbf{X}}(\mathbf{X}) \cdot d\mathbf{X} \qquad \text{[E6]}$$

In the multivariate case (like in the 1D scalar case), the integral of the PDF over the full domain of the multidimensional variable must be =1 (i.e. 100%).
An example for a 2D case is the robot localization example (which will be discussed later, in lecture 4), in which the pair of scalar RVs (which represent the 2D position of the robot) will have an associated multivariate PDF, i.e. $p_{x,y}(x,y)$. Two dimensional cases are common: in general, when the states of a second order system are estimated stochastically, the estimates are represented through 2D RVs. For instance, a simple pendulum can be modelled through a second order system, then a PDF representing a belief about the state vector, at certain time, would be a PDF of the type $p_{x_1,x_2}(x_1,x_2)$ (or $p_{\theta,\omega}(\theta,\omega)$ if we consider that particular state representation for modeling that system.)
An example for a 3D case is the localization in 2D, including heading of the robot, which is a 3DoF system, *(x, y, heading)*. Any dynamic system which is modeled by a 3D state vector would imply a 3D PDF, if the states, at certain time t, are estimated stochastically.

Cases for 4 and 5 dimensions will be also considered in our lab projects. There are applications in which the state vector can have dimensionality of thousands and even millions of states (e.g. when solving, stochastically and numerically, Partial Differential Equations).

Definition: Marginal PDF

Even if we have the joint PDF of 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 a 2D PDF, its mathematical definition is shown in the following expression:

$$p_x(x) = \int_{-\infty}^{+\infty} p_{x,y}(x,y)\cdot dy, \qquad p_y(y) = \int_{-\infty}^{+\infty} p_{x,y}(x,y)\cdot dx \qquad \text{[E7]}$$

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

$$p_{\mathbf{X}}(\mathbf{X}) = \int_{\substack{\text{domain}\\ \text{of } \mathbf{Y}}} p_{\mathbf{X},\mathbf{Y}}(\mathbf{X},\mathbf{Y})\cdot d\mathbf{Y} \qquad \text{[E8]}$$

In [E8] we say that "we are marginalizing with respect to the random variable Y".
(we will see, visually, how it looks, in class, through some examples in 2D)

Definition: Conditional PDF

Given a joint PDF $p_{x,y}(x,y)$, the conditional PDFs are defined as follows,

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\,\mathbf{y}) = \frac{p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})}, \qquad p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\,\mathbf{x}) = \frac{p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})}{p_{\mathbf{x}}(\mathbf{x})} \qquad \text{[E9]}$$

The PDF $p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\,\mathbf{y})$ describes the Probability Distribution of the RV ***x*** assuming that the RV ***y*** would take a defined value. For instance if we have the expression of $p_{x|y}(x\,|\,y)$ and we know that the actual value of the RV ***y*** is ***11*** then we just evaluate $p_{x|y}(x\,|\,y=11)$ to know the probability of *x* (assuming that ***y=11***). Now, based on that assumption, our belief about ***x*** should be different (from the one that we would obtain by just marginalizing the joint PDF), because we know the value of ***y***. Question: Should the new PDF (of the individual RV) be modified due to the knowledge about the RV ***y***? Answer: If the RVs are not independent then the answer is YES; on the other side, if the RVs *(x,y)* are independent, then the conditional and marginal PDFs will be identical. (see definition of independence in the next subsection).
(We discuss in class a common-sense example: "where are my dogs?" cases for "1D and 2D farms")

If we get the weighted average of $p_{x|y}(x\,|\,y)$, for all the values of *y*, weighting it by the marginal probability of *y*, $p_y(y)$, we obtain the marginal PDF of the RV x, $p_x(x)$.

$$p_{\mathbf{x}}(\mathbf{x}) = \int_{\substack{\text{domain}\\ \text{of } \mathbf{y}}} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\,\mathbf{y})\cdot p_{\mathbf{y}}(\mathbf{y})\cdot d\mathbf{y} \qquad \text{[E10]}$$

Which is equivalent to marginalizing the joint PDF (as it was shown before),

$$p_{\mathbf{x}}(\mathbf{x}) = \int_{\substack{\text{domain}\\ \text{of } \mathbf{y}}} p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) \cdot d\mathbf{y}$$

Statistical Independence

We say that two RVs, $\boldsymbol{x}$ and $\boldsymbol{y}$ (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,

$$(\mathbf{x},\mathbf{y}) \sim p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) = p_{\mathbf{x}}(\mathbf{x}) \cdot p_{\mathbf{y}}(\mathbf{y}) \qquad \text{[E11]}$$

That condition implies that their conditional PDFs (whatever is the assumed value of the other RV) are equal to their marginal ones, i.e.

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\,\mathbf{y}) = \frac{p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})} = p_{\mathbf{x}}(\mathbf{x})$$
$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\,\mathbf{x}) = \frac{p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})}{p_{\mathbf{x}}(\mathbf{x})} = p_{\mathbf{y}}(\mathbf{y}) \qquad \text{[E12]}$$

(In many cases the RVs that we will try to estimate are not independent; we will take advantage of the RV' dependency for estimation purposes of the full state of a system)

Examples of Dependent and Independent Random Variables

Example 1

This example involves discrete RVs; consequently, in place of using a PDF we need to use a PMF (Probability Mass Function).

Suppose we have a lottery game composed by two numbers. We use two containers which have 10 balls each. In each container all the balls are of identical characteristics, except that they have different labels, numbered from 0 to 9. Our RV is a 2D variable $X = (x_{1,} x_2)$ in which each component is a scalar RV whose feasible hypotheses are defined by the set $\Omega = \{0,1,2,3,4,5,6,7,8,9\}$. In the process of generating the two numbers, we choose (randomly) one ball from the container 1 (we call this instance x_1), and then another ball from container 2 (we call this instance x_2.)

Question: Do you think that the value x_2 could be affected by the value we previously got for x_1?, or expressed in a different way: based on the value we got for x_1, is that information useful for guessing x_2?

Answer: NO, because these RVs seem to be independent (they are, if we assume those mentioned conditions) .

Now suppose a different way for implementing this lottery game. Now we know that in the process of generating the two numbers, we choose (randomly) one ball from container 1 (we call this instance x_1), then without returning that ball to the container, we get another ball from the same container 1 (we call this instance x_2).

Again, based on the value we got for x_1, is that information useful for guessing x_2? The answer is YES. This is because these RVs are, in this case, dependent. The variable x_2 is still random for us, but we know, at least, that it cannot be $x_2 = x_1$. So, if for x_1 we got ball#5 ($x_1 = 5$), we do know that x_2 will never be = 5 (consequently, we should not bet on that hypothesis!). In general, for this special class of lottery game, we would only try to guess couples (x_1, x_2) that satisfies $x_1 \neq x_2$.

We can obtain the JOINT PMF (Probability Mass Function) for both cases. In the first case we can represent the joint PMF as follows

$$p_{x_1,x_2}(x_{1,}x_2) = \begin{cases} a & \forall (x_{1,}x_2): \quad x_1 \in \Omega, \quad x_2 \in \Omega \\ 0 & otherwise \end{cases}$$

(where a is a positive constant, such as $p_{x_1,x_2}(x_{1,}x_2)$ has volume =1, i.e. $a = 1/(10 \cdot 10) = 1/100$)

We can see that integrating the joint PMF, over the full domain of the random variables, we verify that $\sum_{x_1}\sum_{x_2} p_{x_1,x_2}(x_{1,}x_2) = 1$. (note: We use a summation, in place of an integral, for dealing with PMF functions)

If we visualize this PMF we would see a flat plateau, on the region $0 \le x_1 \le 9$, $0 \le x_2 \le 9$, in the middle of a desert of value zero (any couple where $x_1 \notin \Omega$ or $x_2 \notin \Omega$).

In the second case, we would get

$$p_{x_1,x_2}(x_{1,}x_2) = \begin{cases} b & \forall (x_{1,}x_2): \quad x_1 \in \Omega, \quad x_2 \in \Omega, \quad x_{1,} \neq x_2 \\ 0 & otherwise \end{cases}$$

(where b is a positive constant, to guarantee that $p_{x_1,x_2}(x_{1,}x_2)$ has volume =1, i.e. $b = 1/(10 \cdot 10 - 10) = 1/90$)

If we visualize this PMF it would be a flat plateau (as in the previous case), but diagonally cut by a canyon having value =0.

Now we use the same example to see the conditional PDFs. We consider the conditional PDF of x_2 given x_1.

If someone ask you to predict the outcome of x_1 you would say, I do not know, i.e. you would propose a uniform estimate, $p_{x_1}(x_1) = 1/10, \quad \forall x_1 \in \Omega$

They play the ball, and without telling you the outcome of the first ball (x_1) they ask you to guess about x_2, so you would give the same guess $p_{x_2}(x_2) = 1/10, \quad \forall x_2 \in \Omega$. In both cases you have proposed the marginal PMFs.

A different situation is if they do tell you the outcome of x_1; in that case you would apply the conditional PMF of x_2 given x_1, that is $p_{x_2|x_1}(x_2 | x_1) = p_{x_1,x_2}(x_1, x_2) / p_{x_1}(x_1)$. Suppose they tell you that x_1 was = 5.

For case 1, you could infer that "that help does not actually help" (because both RVs are independent).

For case 2, ((x_1, x_2) are not independent) you may use the information for improving your belief about x_2.

You would say, x_2 may take any number (in Ω) except $x_2 = 5$!.

The expression of the conditional PDF $p_{x_2|x_1}(x_2 | x_1 = 5) = p_{x_1,x_2}(5, x_2) / p_{x_1}(5)$ tells you, mathematically, the same result. It would be expressed as

$$p_{x_2|x_1}(x_2 | x_1 = 5) = \begin{cases} 1/9 & \forall x_2: \quad x_2 \in \Omega, \quad x_2 \neq 5 \\ 0 & otherwise \end{cases}$$

Example 1b

Think about another rare lottery game, in which after extracting a ball from container 1, that ball is added to the ones in container 2, and then a ball is extracted from that 2nd container for obtaining x_2. How would be the joint PDF about (x_1, x_2)? (to think at home)

Example 2

In this example we use continuous variables. Suppose we have a hermetic container filled by 1 mole (1 mol) of an ideal gas. We do not have any information about the pressure and temperature of the gas (call those variables P,T) and about the volume of the container (V).

Now we have interest about knowing (P,T). From our perspective they are Random Variables, because we do not know their values. We represent our knowledge about (P,T) using a PDF (our "belief" about the state of the system). (Remember that we do not know the container's volume, V).
How is this PDF which we would propose? It is uniform, because we do not have any clues about P and T. Any combination of (P,T) is equally probable.
Suppose now that we are able to measure *T*. Could we guess (estimate or refine our belief) about the value of *P*? The answer is NO. Based on our current knowledge we cannot improve our belief about *P*.
After the measurement of T we can say, for example, *T=10C, P= any value*.

Our initial Belief: (T=any, P= any) ➔ *measure T=10C* ➔ *Our new Belief: (T=10C, P= any value)*

Suppose a different situation. Suppose we do have some information about the volume of the container. When they made it, they said its nominal volume was 1000 litres, but it could actually be between 900 and 1100 litres. We also know that there exists a constraint that governs the relation of *P* and *T* for an ideal gas, for a given volume. Something like this equation: $P \cdot V = k \cdot T$, where ***k*** is a well-known constant, defined by the mass and type of gas (both properties are implicitly known for us, because we said we have 1 mol of gas), *T* is the absolute temperature.
Now, even having a perfect knowledge about *T* we still cannot perfectly know P, however we can provide a range of possible values for P (we can even provide a sort of PDF for modeling our belief about P, $p_P(P)$.).
We are able to improve our belief about P because the RVs P and T, according to our belief (before we measured T), are NOT *independent*.

Our initial Belief: (T= any, P= any) ➔ *measure T=10C* ➔ *Our new Belief: (T=10C, P= certain* ***informative*** *belief about P)* (*)(**)

(*) How to obtain the new belief about P: we will learn proper approaches for doing it, later in AAS.
(**) Using our intuition: we could approximate it, by using the equation $P = k \cdot T / V$; where ***k*** is known, *V* is assumed to be close to its nominal value (which is known), and *T* was measured. We infer that P is close to its nominal value, obtained using the equation (OK! We, intuitively, got some sort of PDF about *P*, based on the condition that T=10C)

To be solved by the student (note: this is not a quiz or any type of assignment):

1) Propose one example of a system represented by a multivariate RV, where certain components are independent. Consider the case of continuous variables.

2) As in (1), but where certain components are dependent.

3) Consider the following "special" lottery game, where:
Container #1 contains 3 balls labeled=1 and 3 balls labeled [0,2,3] (totalizing 6 balls in the container). Container #2 contains 2 balls labeled=3 and 2 balls having labels 0,1. No balls labeled with other numbers.

3.1) Consider the case where x_1 is defined by taking one ball from container #1 and x_2 by taking one ball from container #2.

3.2) Consider the case where the ball extracted from container #1 is added to container #2, before the second ball (which gives x_2) is obtained. (I know, it is a "crazy lottery").

For each of the cases, 3.1 and 3.2, find the following PMF functions:

$p_{x_1,x_2}\left(x_1,x_2\right)$ (Joint PMF)

$p_{x_2|x_1}\left(x_2 \mid x_1 = 1\right)$ (Conditional PMF of x_2 given $x_1 = 1$)

$p_{x_2|x_1}\left(x_2 \mid x_1 = 2\right)$ (Conditional PMF of x_2 given $x_1 = 2$)

$p_{x_2}\left(x_2\right)$ (Marginal PMF for rv x_2)

$p_{x_1}\left(x_1\right)$ (Marginal PMF for rv x_1)

You may show, numerically in Matlab, a visualization of $p_{x_1,x_2}\left(x_1,x_2\right)$ (using image or surfaces or other graphical way, as shown in page 16, where I used the function *bar3*( ) to show the joint PMF)

Definition: WHITE NOISE

Any signal $\boldsymbol{\mu}(t)$ whose time values are instances of RVs, which are independent, i.e.

$$p_{\boldsymbol{\mu}(t),\boldsymbol{\mu}(t+T)}\left(\boldsymbol{\mu}(t),\boldsymbol{\mu}(t+T)\right) = p_{\boldsymbol{\mu}(t)}\left(\boldsymbol{\mu}(t)\right)\cdot p_{\boldsymbol{\mu}(t+T)}\left(\boldsymbol{\mu}(t+T)\right), \quad \forall t,T$$

(and in addition have zero mean); is said the be "white".
We are saying that we consider the signal μ at each time *t* to be a RV. Each instantaneous value of the signal, $\mu(t)$ is a RV and we also say that any couple of RVs $\left(\boldsymbol{\mu}(t_1),\boldsymbol{\mu}(t_2)\right)$ are independent, for any $\left(t_1,t_2\right)$.

Notes:

* The name "white" is due to the analogy with the white light, which has uniform spectrum. White noise signals have an *expected* Fourier spectrum which is uniform.

* The definition is also applied to discrete signals $\left\{\boldsymbol{\mu}(k)\right\}$ (we mention it in class)

The definition is also valid for higher dimensional cases. An example is a high dimensional discrete signal, such as a VGA grayscale image, $\left\{\boldsymbol{\mu}(u,v)\right\}_{\substack{1\le u\le 640\\ 1\le v\le 480}}, \boldsymbol{\mu}\in\mathbb{R}^1, 0\le\boldsymbol{\mu}\le 1$, where

$$p_{\boldsymbol{\mu}(u_1,v_1),\boldsymbol{\mu}(u_2,v_2)}\left(\boldsymbol{\mu}(u_1,v_1),\boldsymbol{\mu}(u_2,v_2)\right) = p_{\boldsymbol{\mu}(u_1,v_1)}\left(\boldsymbol{\mu}(u_1,v_1)\right)\cdot p_{\boldsymbol{\mu}(u_2,v_2)}\left(\boldsymbol{\mu}(u_2,v_2)\right)$$

What says that if we were able to know information about the value of pixel $\left(u_1,v_1\right)$, the information would be useless for improving our knowledge about the value at pixel $\left(u_2,v_2\right)$.
(this is a simple example; you should also consider it for continuous cases, where *u* and *v* are not discrete)

In class: We see samples of RGB random images; some of them made of "white noise", some others also random, but not "white noise".
(Note: remind me to show these examples, I may behave a bit randomly, and forget to do so.)

Example

Suppose we are periodically reading from a 12-bits Analog to Digital Converter (ADC), for measuring the voltage of a time varying signal, which always evolves in the valid range of the ADC input (e.g. 0 to 10 volts). The ADC conversion introduces quantization levels, of size $10/2^{12} = (10/4096)$ volts, which means that approximating the reading, by the nearest valid discrete level, will introduce an error, which is bounded, $-W < e < W$, in which $W = (10/4096/2)$ volts . From our perspective, this error is a RV, because we cannot know its value. However, we can easily define a uniform PDF, for describing this "noise",

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

All the instances of $e(k)$ will have the same PDF, which describe them.

We still have to decide: is this random sequence, $\{e(k)\}$, "white"?

Strictly talking, the answer is NO. However, in practical terms, under certain conditions we can say YES.
It depends on how frequently the sampling process is, and on the frequency content of the signal being sampled. If we sample at very high frequency, i.e. the signal being sampled does almost not change during consecutive samples, that would result in sampling almost the same voltage, consequently the quantization error would be almost the same, so there would be statistical dependence between the RVs $e(k)$ and $e(k+1)$.

However, if the signal being sampled does vary, then that quantization error will behave more "white".
Consequently, we usually assume the ADC's quantization error to be white noise.
(Btw: In addition to the ADC quantization error, other sources of noise may affect the ADC reading)

Example 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)

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 that RV to be 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.
(because $e(i) = e(j)$, or at least very similar for the case of a slowly varying bias)

Similar cases, where the noise is clearly non-white: Error in GPS position estimates (from a standard GPS receiver). 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.)

Example 3: In class, we will analyze the noise in Gyros measurements. We will see, it can be seen as being composed by a constant bias and by a fluctuating component which can be considered white noise (in addition Gaussian)

Definition: Expected Value

Given a PDF, we define the mean (or expected value), of the RV associated to that PDF, as:

$$\hat{\mathbf{x}} = E\{\mathbf{x}\} = \int_{\Omega_{\mathbf{x}}\text{ (domain of }\mathbf{x})} \mathbf{x} \cdot p_{\mathbf{x}}(\mathbf{x}) \cdot d\mathbf{x} \qquad \text{[E13]}$$

The interpretation of [E13]: Given a RV $\mathbf{x}$ whose statistical description is given by the PDF $p_{\mathbf{x}}(\mathbf{x})$ we can get its averaged value by [E13]. This average is called the *mean* or *Expected Value*. One consequence of this is that if we were able to take many samples of ***x*** and we performed an average of them, then this averaged value would converge to the mean of the random variable *x*.

We can also define the expected value of any function of the RV $\mathbf{x}$, e.g. for a given function $g(\mathbf{x})$,

$$E\{g(\mathbf{x})\} = \int_{\Omega_{\mathbf{x}}} g(\mathbf{x}) \cdot p_{\mathbf{x}}(\mathbf{x}) \cdot d\mathbf{x} \qquad \text{[E14]}$$

A particular case of [E14] is the well-known Covariance Matrix:

$$E\left\{(\mathbf{x}-\hat{\mathbf{x}})\cdot(\mathbf{x}-\hat{\mathbf{x}})^T\right\} = \int_{\Omega_{\mathbf{x}}} (\mathbf{x}-\hat{\mathbf{x}})\cdot(\mathbf{x}-\hat{\mathbf{x}})^T \cdot p_{\mathbf{x}}(\mathbf{x}) \cdot d\mathbf{x} \qquad \text{[E15]}$$

(where the vector *x* is expressed as a column vector, so the result of this operation it is a matrix)

This average is called the COVARIANCE matrix. If the RV *x* is a scalar, then it is called VARIANCE and expressed as $\sigma_x^2 = E\left\{(x - \hat{x})^2\right\}$.

If *x* is a vector, then the Covariance is a square matrix.

$$\mathbf{P}_{\mathbf{x}} = E\left\{(\mathbf{x}-\hat{\mathbf{x}})\cdot(\mathbf{x}-\hat{\mathbf{x}})^T\right\} \in \mathbb{R}^{N\times N}, \quad \hat{\mathbf{x}} \in \mathbb{R}^N \qquad \text{[E16]}$$

Definition: Cross-covariance

The cross-covariance of two RVs, described by a joint PDF $p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})$, is defined as:

$$\operatorname{cov}(\mathbf{x},\mathbf{y}) = P_{\mathbf{x},\mathbf{y}} = E\left\{(\mathbf{x}-\hat{\mathbf{x}})\cdot(\mathbf{y}-\hat{\mathbf{y}})^T\right\} = \int_{\substack{domain\\ of\ \mathbf{x}}} \int_{\substack{domain\\ of\ \mathbf{y}}} (\mathbf{x}-\hat{\mathbf{x}})\cdot(\mathbf{y}-\hat{\mathbf{y}})^T \cdot p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) \cdot d\mathbf{x} \cdot d\mathbf{y} \qquad \text{[E17]}$$

($\mathbf{P}_{\mathbf{x},\mathbf{y}}$ is a matrix of dimension $N_x \times N_y$, because $\mathbf{x} \in \mathbb{R}^{N_x}, \quad \mathbf{y} \in \mathbb{R}^{N_y}$)

(Note: remind me to mention what happens when the RVs x and y are independent.)

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 $\boldsymbol{x}$. If it is said to be a Gaussian RV, then its PDF will have the following structure

$$p_x(x) = c \cdot e^{-\frac{(x-\hat{x})^2}{2\cdot\sigma_x^2}} \qquad \text{[E18]}$$

The parameter $\hat{x}$ is called the **mean** (or **expected value**) and the parameter σ_x^2 is the **variance** of the RV $\boldsymbol{x}$. The parameter σ_x is called the standard deviation. The constant $\boldsymbol{c}$ is just needed for normalizing the PDF. For a Gaussian PDF the values of these two parameters, $\hat{x}$ and σ_x , fully define the PDF.
This distribution is also called NORMAL distribution and it is usually expressed as

$$x \sim N\left(x\,;\hat{x},\sigma_x^2\right)$$

It is also usually expressed as $x \sim N\left(\ \hat{x},\sigma_x^2\right)$ [E19]

Expression [E19] says that $\boldsymbol{x}$ is a RV, whose PDF has the expression in [E18], having the specified values of expected value and covariance, i.e. "a Gaussian PDF of the R.V. $\boldsymbol{x}$, that has expected value $\hat{x}$ and variance σ_x^2".
For instance, the voltage $v \sim N(240,25)$, implies a Gaussian RV with expected value =240volts, standard deviation 5 volts.
Gaussian PDFs have "*bell*" shapes, like the following examples (these are 1 dimensional examples):

The expected value and covariance are defined, for any PDF, by the following operations (for 1D cases):

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

In particular, for the Gaussian case, those values are the parameters of the functional expression of the PDF.

Multivariate Gaussians

RV's for representing higher order systems are generally multivariate, i.e. having dimensionality higher than 1. For instance, for representing the states of the pendulum, which is usually modeled by a second order ODE, we need two states. When we estimate the pendulum's states we model them as RV's, then we use a 2D PDF. We are saying that our belief about the system's state, at a certain time, is represented by a PDF of two RV's (e.g. angle and angular rate). If we assume that these variables are described as Gaussian then we need to use a 2D Gaussian PDF.

A 2D Gaussian PDF has the following structure:

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

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

[E20]

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

The matrix $\mathbf{P}$ is called COVARIANCE matrix. The diagonal elements of $\mathbf{P}$ are the variances of the individual RV's x_1 and x_2 (the variances of their *marginal PDFs*!). The elements $\mathbf{P}_{1,2}$ and $\mathbf{P}_{2,1}$ are the cross-covariance sub-matrixes. Those ones describe the statistical dependency between the estimates of x_1 and x_2.

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

1) It is a symmetric matrix (*)
2) It is a positive semi-definite matrix (*).

Note (*): both characteristics are mathematical definitions that you know from previous courses.

The expected value $\hat{\mathbf{X}}$ is a vector and its components are the expected values of the individual scalar variables which compose the vector.

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$$

(Note: see a picture of a 2D Gaussian PDF at the end of this lecture notes.)

If we consider cases of dimension *n*, then $\hat{\mathbf{X}} \in \mathbb{R}^n, \mathbf{P} \in \mathbb{R}^{n\times n}$.

(btw: The expected value and covariance matrix are defined for any PDF by the following operations:

$$\hat{\mathbf{X}} = \int_{\substack{Domain \\ of\ \mathbf{X}}} \mathbf{X} \cdot p_{\mathbf{X}}(\mathbf{X}) \cdot d\mathbf{X}, \quad \mathbf{P} = \int_{\substack{Domain \\ of\ \mathbf{X}}} (\mathbf{X}-\hat{\mathbf{X}}) \cdot p_{\mathbf{X}}(\mathbf{X}) \cdot (\mathbf{X}-\hat{\mathbf{X}})^T \cdot d\mathbf{X} \qquad \text{[E21]}$$

In particular, for the Gaussian case, those values are parameters of the functional expression of the PDF.)

High Dimensional Problems? There are estimation problems that can involve 3, 4, .., 20.. or even 1,000,000 dimensions! We will mention an example in class. Just remind me to do it.

Examples of Multivariate Gaussian PDFs

We say that a PDF is multivariate when the random variable is a vector (it has more than one dimension).
We saw that, in those cases, the expected value is a vector as well and it has the same dimension as the random variable. The Covariance matrix is a square matrix. If the RV *x* has *n* dimensions, then the covariance is a matrix of size *n* by *n.*
Let's analyze the case for *n=2* to understand this concept. Assume a random variable X

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$$

The components of the vector are the current state of our system. Examples can be the 2D position of a robot or the states of a second order system such as a pendulum (its angular position and its angular speed). Or any couple of random variables we would need to describe.

Case 1

Suppose the RV *X* is described by a Gaussian PDF whose expected value and Covariance are:

$$\hat{\mathbf{X}} = \begin{bmatrix} 1 \\ 3.5 \end{bmatrix}, \quad \mathbf{P}_{\mathbf{X}} = \begin{bmatrix} 2 & 0 \\ 0 & 9 \end{bmatrix} \qquad \text{[E22]}$$

In this case the non-diagonal elements are nil. This means that the RV x_1 and x_2 are independent. Although we describe them through a joint PDF they can be represented individually because, in this case, they are independent. In addition, the covariance matrix tell us that the variance of x_1 is =2 (standard deviation $\sigma_{x_1} = \sqrt{2}$) and that the one of x_2 is 9 ($\sigma_{x_2}^2 = 9 \Rightarrow \sigma_{x_2} = 3$).

If we plot the PDF in the domain (x_1, x_2) then we will see an inverted BELL with center (maximum peak) at the expected value ($(x_1, x_2) = (\hat{x}_1, \hat{x}_2) = (1,\ 3.5)$) and presenting an ellipsoidal section. This will be shown in the class.

Case 2:

Suppose we have the same expected value but with the following Covariance Matrix

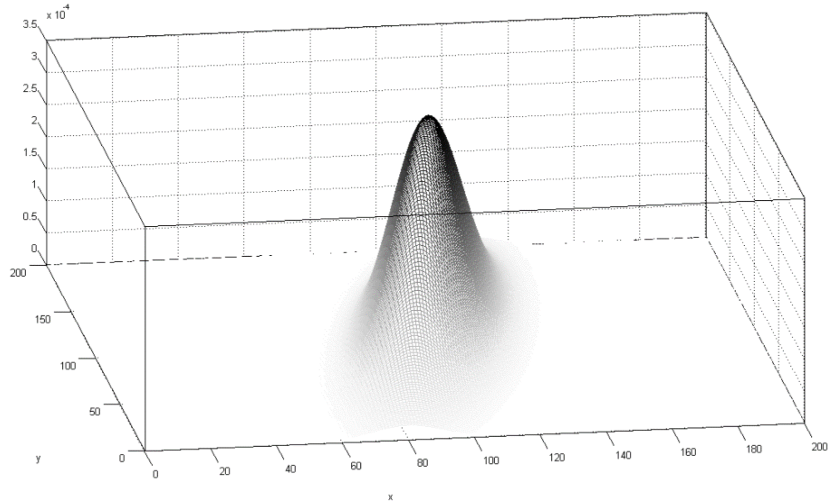
$$\mathbf{P}_x = \begin{bmatrix} 2 & 1 \\ 1 & 9 \end{bmatrix}$$

In this case the non-diagonal elements are non-zero. This tells us that there is some dependency between the components x_1 and x_2. Their marginal PDF are still defined by the diagonal values, i.e. the variable x_1 described by the variance ($\sigma_{x_1}^2 = 2$) and the variable x_2 by the covariance $\sigma_{x_2}^2 = 9$. The mission of the non-diagonal elements is to represent the dependency between the knowledge about the individual variables. This dependency means that information about one RV can also be shared with the other RV. For instance if we receive some information about x_1 (e.g. a measurement of it) then we will be able to refine our knowledge about x_1 and about x_2 as well! Note than in the **case 1** our knowledge about x_2 would be untouched.

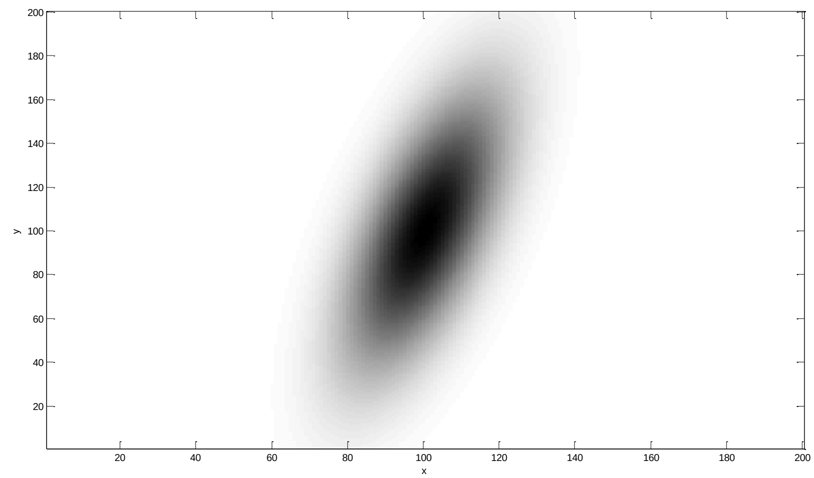
An Example of a 2D Gaussian PDF

Some pictures of a 2D PDF. A Normal Distribution with Covariance Matrix

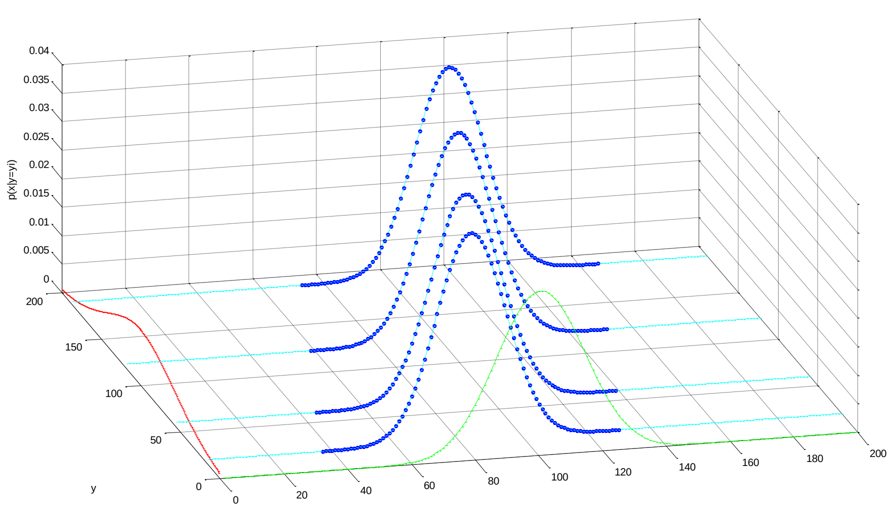
$\mathbf{P} = \begin{bmatrix} 400 & 752 \\ 752 & 3600 \end{bmatrix}$ and expected value at x=100, y=100.



Presented as a 2D function of (x,y) (the dark color means high density values).

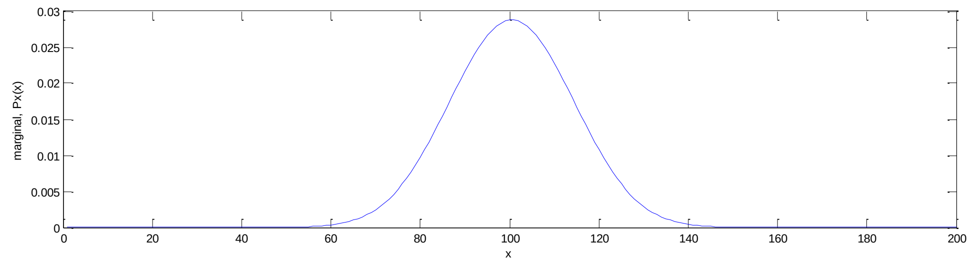


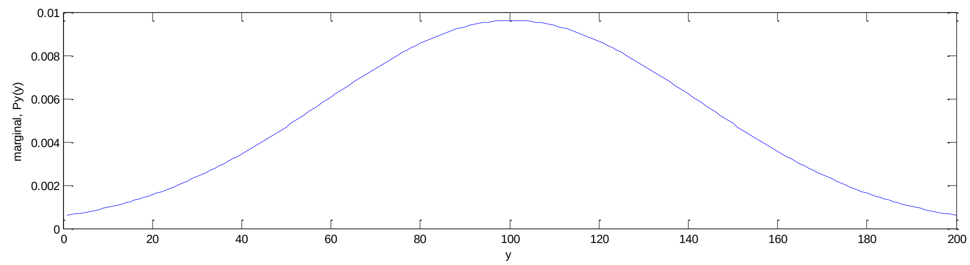
Same PDF, but presented as an image. Dark color means high values. White means very low value of the PDF at that "pixel" (point x,y).



Some instances of $p_{x|y}(x \mid y)$, for certain values of **y**.

In the previous figure, we could see some conditional PDFs. Probability Distributions of ***x*** given ***y***, $p_{x|y}(x\,|\,y)$, evaluated at certain values of ***y*** and represented in the ***x,y***, space. The marginal PDFs $p_x(x)$ and $p_y(y)$ are also presented. Interesting to note is that the variances and expected values of the conditionals $p_{x|y}(x\,|\,y)$ are function of the value of ***y***. If we observed some of the conditional PDFs, e.g $p_{x|y}(x\,|\,y=50)$ and $p_{x|y}(x\,|\,y=100)$, we would see that, seen as functions of x, would be well different Gaussian PDFs. That is because both RVs are not independent (according to the given joint PDF),





Corresponding Marginal PDFs $p_x(x)$ and $p_y(y)$.

Definition: Uncorrelated RVs

Two RVs are said to be uncorrelated if:

$$\operatorname{cov}(\mathbf{x},\mathbf{y}) = \mathbf{P}_{\mathbf{x},\mathbf{y}} = E\left\{(\mathbf{x}-\hat{\mathbf{x}})\cdot(\mathbf{y}-\hat{\mathbf{y}})^T\right\} = 0$$

If these RVs were vectors, that zero would be a zero matrix of proper dimension.

If the RVs are also Gaussian, then it means they are independent. Consequently, if jointly represented by a joint PDF, that PDF will have a block diagonal covariance matrix. If ***x*** and ***y*** were scalars, the 2x2 covariance matrix will be diagonal.

Independence for Gaussian Random Variables

Two Gaussian random variables, ***x*** and ***y***, are independent if they are zero mean and uncorrelated, i.e.

$$\mathbf{P}_{\mathbf{x},\mathbf{y}} = E\left\{(\mathbf{x}-\hat{\mathbf{x}})\cdot(\mathbf{y}-\hat{\mathbf{y}})^T\right\} = 0 \qquad \text{[E23]}$$

The matrix $\mathbf{P}_{\mathbf{x},\mathbf{y}}$ is called cross-covariance matrix. Note the term "zero" in [E23] is a ZERO matrix of adequate size (it depends on the dimension of *x* and *y*).

Definition: Gaussian White Noise (GWN)

It is similar to "white noise" but for the case in which the PDFs $p_{\boldsymbol{\mu}(t)}(\boldsymbol{\mu}(t))$ are Gaussian.
Note that the underlying PDFs $p_{\boldsymbol{\mu}(t)}(\boldsymbol{\mu}(t))$ can vary according to *t*, i.e. the associated covariance matrixes can depend on the time, i.e. $p_{\boldsymbol{\mu}(t)}(\boldsymbol{\mu}(t)) = N(\boldsymbol{\mu}(t);\bar{0},\mathbf{P}(t))$, although in many cases it is usually constant.
(for instance, for the noise that pollutes measurements from certain class of sensors we assume constant variance)

For zero mean Gaussian RVs, if those are uncorrelated it implies that are independent (and vice versa)

An equivalent way to express this characteristic (Gaussian White Noise, aka GWN) is to say that $\boldsymbol{\mu}(t)$ is zero mean, uncorrelated, Gaussian noise.

BAYES' rule.

This is an extremely relevant concept; it is the base of all the Bayesian estimations approaches (Kalman Filters, Particle Filters, etc.)

From Equation E16 we know that:

$$p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) = p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\,\mathbf{y})\cdot p_{\mathbf{y}}(\mathbf{y}) = p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\,\mathbf{x})\cdot p_{\mathbf{x}}(\mathbf{x}) \qquad \text{E19}$$

We focus our attention on the second and third members of E19,

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\,\mathbf{y})\cdot p_{\mathbf{y}}(\mathbf{y}) = p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\,\mathbf{x})\cdot p_{\mathbf{x}}(\mathbf{x}) \qquad \text{E20}$$

then we get,

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\,\mathbf{y}) = \frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\,|\,\mathbf{x})\cdot p_{\mathbf{x}}(\mathbf{x})}{p_{\mathbf{y}}(\mathbf{y})} \qquad \text{E21}$$

This equation represents a fact. Note: I am assuming that all the members in E21, marginal and conditional probabilities come from the same joint PDF.
If we were able to know all these functions, we would be able to verify E21 (or E20, E19). However, in estimation problems we usually do not know all of them (however we can still exploit E21, as we will see later, in a subsequent lecture)

Still pending to be seen.

PDF of transformed RV

PDF of transformed Gaussian RV

Understanding the covariance

Marginal PDF of a Gaussian PDF.

Q: Why do we need to know these concepts?

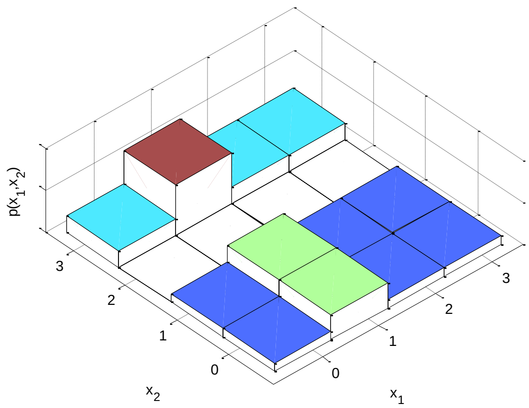
A: Because we exploit them to understand an approach we need to apply leter.

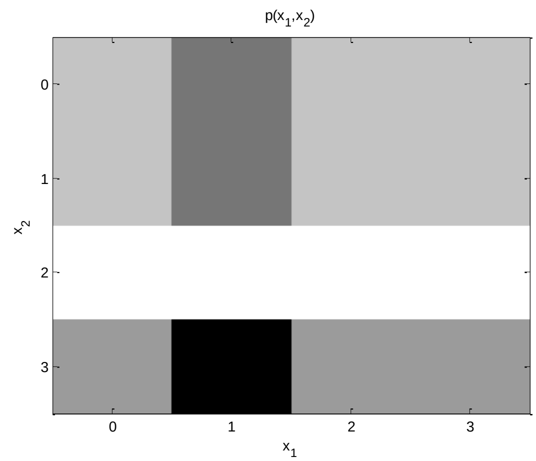
Appendix: Solutions for problem in page 6

Purpose: Just for visualizing a previous example about dependent RVs. We do not apply this type of calculations in MTRN4010, and we do not have assessment involving this type of calculation.

Mass Density Function, $p_{x_1,x_2}\left(x_1,x_2\right)$, for both "classes of Lottery", defined in page 10.

Case 1:





On the left: Probability Mass Function, $p_{x_1,x_2}\left(x_1,x_2\right)$, shown by bars. On the right: Probability Mass Function, $p_{x_1,x_2}\left(x_1,x_2\right)$, shown as a grayscale image. Dark colors mean high values of probability, light colors mean low probability (e.g. white means p=0).

Case2