Advanced Autonomous Systems

Bayesian Estimation- Part 1

In this lecture we commence our study about Bayesian Estimation. Although the concepts which are discussed in this lecture can be applied to diverse problems, we will usually focus (in subsequent lectures) our attention in solving our problems; those are related to Advanced Autonomous Systems and Field Robotics. One of the problems we need to solve is the usually called "Localization", which, in our context, means the estimation of the position and attitude of a mobile platform.

To perform the estimation of the state of a system, we need to process information provided by multiple different sources, i.e. by the system's models, and by measurements of certain system's outputs, usually via sensing.
The uncertainty, which is present in the models and in the measurements (i.e. noise, uncertainty about model's structure and parameters, etc.), can be modeled through stochastic variables called random variables (random variable: RV). We do not know the values of those "contributions", but we usually know their statistical properties.

A dynamic system can be usually represented by a *process model*, such as the following one,

$$\frac{d\mathbf{x}(t)}{dt} = f\left(\mathbf{x}(t), \mathbf{u}(t)\right) \qquad \text{(E1)}$$
$$\mathbf{x}(t) \in R^n$$

in which the variable **X** is called the system's state vector, and the vector **u** represents the system's inputs. The model in (E1) describes the dynamics of the system, through a first order differential equation, usually of high dimension (see note 1). Other systems can be modeled through a discrete-time version of (E1).
Although the state vector is usually not accessible (note 2) we are able to measure certain system's outputs, at certain times. We express that situation as follows,

$$\mathbf{y}(t) = h\left(\mathbf{x}(t)\right), \quad t \in \{t_1, t_2, ..\} \qquad \text{(E2)}$$

Equation (E2) shows that the outputs of the system, $\mathbf{y}(t)$, are dependent on the system's states, $\mathbf{x}(t)$. We have access to (i.e. we can measure) those outputs at certain times ($t_i = t_1, t_2, ..$).
We call those measurements of the system's outputs: **observations**.
If there is no uncertainty in the models (*f(.)* and *h(.)*) and if the measurements are free of noise, and if the system is OSERVABLE (note 3), we are in condition to obtain estimates of $\mathbf{x}(t)$ through deterministic observers (e.g. Luenberger's observers or similar approaches, as those seen in MTRN3020)

Blank Space for drawing the block diagram of the system (to do in class).

Note 1: We assume you have knowledge about representing systems through state space representation (e.g. from courses about Control theory).
Note 2: Accessible in the sense of "able to be directly seen or measured"
Note 3: Observability, as you learned in Control theory.

In real implementations, we need to deal with certain complications to the previous formulation:

1) The models, ***f(.)*** and ***h(.)*** in (E1) and (E2), are not perfectly known.
2) The initial conditions (more exactly the ones we assume), $\boldsymbol{x}(t_0)$, necessary for initializing (E1), are also affected by uncertainty.
3) The measurements of the system's outputs, $\boldsymbol{y}(t)$, are polluted with noise.

This fact is mathematically expressed as follows,

$$\frac{d\mathbf{x}(t)}{dt} = f\left(\mathbf{x}(t),\mathbf{u}(t),\xi(t)\right)$$
$$\hat{\mathbf{x}}_0 = \mathbf{x}_0 + \boldsymbol{\mu} \qquad \text{(E3)},$$
$$\mathbf{y}(t_i) = h\left(\mathbf{x}(t_i),\boldsymbol{\eta}(t_i)\right)$$

in which the new members $\boldsymbol{\xi}(t), \boldsymbol{\eta}(t_i), \boldsymbol{\mu}$, which are included in the process and observation models, are random variables (RVs). The values of these RVs are unknown; they are included for representing the uncertainty which is present in this system (more exactly: in our knowledge about the system). If we were able to know the statistical properties of these RVs, then we would be able to obtain *estimates* of $\boldsymbol{x}(\boldsymbol{t})$; and those estimates would be expressed in a statistical way as well (i.e. through PDFs).

Some simplified version of (E3) expresses the uncertainty through additive terms in the process and observation models.

$$\frac{d\mathbf{x}(t)}{dt} = f\left(\mathbf{x}(t),\mathbf{u}(t)\right) + \xi(t) \qquad \text{(E4)}$$
$$\mathbf{y}(t_i) = h\left(\mathbf{x}(t_i)\right) + \boldsymbol{\eta}(t_i)$$

This is because we assume a process model (nominal model) $f\left(\mathbf{x}(t),\mathbf{u}(t)\right)$ which does not capture the real dynamics of the system, i.e. there is a discrepancy (error), $\xi(t) = \frac{d\mathbf{x}(t)}{dt} - f\left(\mathbf{x}(t),\mathbf{u}(t)\right)$, between the nominal and real dynamics.

Discrete time cases

Similarly, as in the continuous time case, in discrete time systems the uncertainty is modelled through RVs, as well.

$$\mathbf{x}[k+1] = F\left(\mathbf{x}[k],\mathbf{u}[k]\right) + \xi[k]$$
$$\mathbf{y}[k] = h\left(\mathbf{x}[k]\right) + \boldsymbol{\eta}[k]$$

This is the class of systems which we are going to discuss, as we have the intention to implement the estimation processes in digital processors.

Example about using observations

In this brief discussion, we consider a case, in which measurements of certain outputs of a system allow us to estimate the state of that system. We remark, again, that we consider a variable to be a system's output, if that variable is function of the state of the system $\mathbf{x}$, i.e. $\mathbf{y} = h(\mathbf{x})$. Each time we measure a system's output we say it is "an observation".

Suppose we have a static (it is not moving, in this example) robot, which is equipped with on-board sonar sensing capabilities. The sensor allows measuring its distance to certain landmarks whose absolute positions, in certain reference coordinate frame, are well known by us. Those landmarks are installed in the area of operation of the robot. The sonar reports range but it does not measure bearing (direction) information. The range measurements are contaminated with noise. Suppose we have just two (2) landmarks, installed in the area, which are detected (by the sonar) consecutively in a periodic fashion; i.e. we receive a range measurement associated to landmark-**a** and then one measurement associated to landmark-**b**, and, periodically, we may also repeat the sequence of measurements. In addition, we know that the robot is not moving, and that the landmarks, of course, do not move (are fixed to the terrain), and we know the position of both landmarks (someone previously surveyed their positions and provided a *map*, i.e. a table expressing their positions, in a reference coordinate frame.)

As we do know the position of the landmarks (assume we know them without any uncertainty) then we are in condition to estimate our position by a triangulation process.

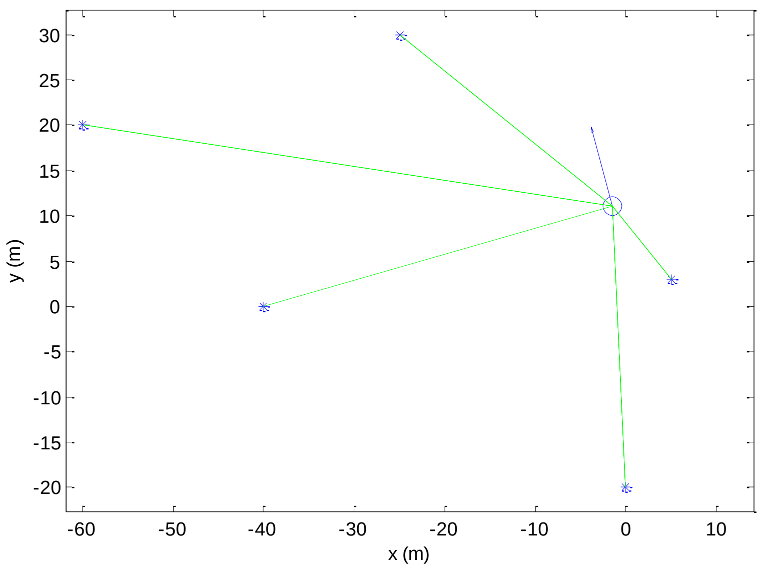


Figure: Platform with a sonar, able to measure its distance to individual beacons which are deployed in its area of operation (in this figure we show the case in which 5 landmarks are reachable to the robot's sonar). Each green segment represents an observation, i.e. a measurement of the distance between the sensor and one beacon.

For estimating the robot's position, (x, y), we propose a system of equations, to be solved as follows,

$$\begin{aligned} r_a &= \sqrt{(x - x_a)^2 + (y - y_a)^2} \\ r_b &= \sqrt{(x - x_b)^2 + (y - y_b)^2} \end{aligned} \qquad \text{(E5)},$$

In which r_a is the distance between the robot's sensor and landmark a; and r_b is the distance respect to landmark b. The centers of the landmarks #a and #b are (x_a, y_a) and (x_b, y_b) respectively. This problem is mathematically easy and well defined if landmarks ***a*** and ***b*** are located at different positions. The problem will have two candidate

solutions, i.e. the intersection of two circumferences, as we know from basic Math. One of the solution corresponds to the position of the robot.
if the sonar measurements are not perfect, (because there is noise polluting those measurements), then our equations (based on the same range measurements) become as follows,

$$\begin{aligned} r_a &= \sqrt{\left(x - x_a\right)^2 + \left(y - y_a\right)^2} + \eta_a \\ r_b &= \sqrt{\left(x - x_b\right)^2 + \left(y - y_b\right)^2} + \eta_b \end{aligned} \qquad \text{(E6)}$$

These equations express that each measurement of the distance is equal to the real distance (whose value is unknown to us) plus some additive error. The two error components are represented by the RVs η_a and η_b. If we had the statistical properties of η_a and η_b then we would be able to obtain a solution for $\left(x, y\right)$ in the form of estimates of $\left(x, y\right)$ expressed as a PDF of these variables, $p_{x,y}\left(x, y\right)$, i.e. they would be expressed as RVs as well. How is the shape of that PDF? It will depend on the PDFs of η_a, η_b, and on the functional relationship (between measurements and the states) given by (E6).
In addition, some average effect can be achieved by repeating the observations. In this case, we assume the robot and landmarks do not move, and that we are able to get more measurements, which are also polluted by random noise; and we also assume that the noise is “white” (i.e. noise components in different measurements are independent. We will discuss about this matter later). We define the following system of equations,

$$\begin{aligned} r_{a,1} &= \sqrt{\left(x - x_a\right)^2 + \left(y - y_a\right)^2} + \eta_{a,1} \\ r_{b,1} &= \sqrt{\left(x - x_b\right)^2 + \left(y - y_b\right)^2} + \eta_{b,1} \\ \\ r_{a,2} &= \sqrt{\left(x - x_a\right)^2 + \left(y - y_a\right)^2} + \eta_{a,2} \\ r_{b,2} &= \sqrt{\left(x - x_b\right)^2 + \left(y - y_b\right)^2} + \eta_{b,2} \end{aligned} \qquad \text{(E7)}$$

Although we are repeating the observations, these still contribute with extra information, because the noise included in them is *white*, which means, in practical terms, some benefit can be achieved by “averaging” them.

See the related videos; they show particular cases of this problem being solved through a Bayesian approach (the approach which we will explain later). In the experiment we assumed some realistic PDF for the error in the measurements and we tried different geometries (distribution of landmarks). Those different situations mean that the estimation processes evolve differently, i.e. mainly due to the fact they have different levels of observability, in this case due to the geometrical distribution of the landmarks.
The updating process is repeated up to 15 times (each time observing all the present landmarks).
Details about how to implement this estimator will be explained in subsequent lectures.
We will also run the estimator in class, to see how it behaves for different deployments of landmarks.

To watch the videos again try these links:
Based on geometries of 4 beacons: http://youtu.be/emJ-QJFp2bc
Based on geometries of 2 beacons: http://youtu.be/NQM9IlKMNls
These videos are also provided as: “VideoBayesianLocalizationUsing4Landmarks.mp4” and “VideoBayesianLocalizationUsing2Landmarks.mp4”
note: in class, we will see more examples and cases, through a computer program, simulating different cases.

Applying the Bayes' Rule for Estimation

Now, we consider the Bayes' rule

$$p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x}\,|\,\mathbf{y}\right)=\frac{p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}\,|\,\mathbf{x}\right)\cdot p_{\mathbf{x}}\left(\mathbf{x}\right)}{p_{\mathbf{y}}\left(\mathbf{y}\right)} \qquad \text{(E8)}$$

(we mentioned it, in Lecture 1, manipulating the expressions for the marginal and conditional PDFs, associated to a common joint PDF)
We can use expression (E8) in a recursive fashion in order to implement an estimator.

Note: we use the fact expressed (E8) in order to implement a recursive/iterative process. There exists a strong theory, which demonstrates the convergence of this estimation process. We do not discuss the demonstration of this approach in this course.

Recursive / Iterative Approach

Consider two RVs, $\left(\mathbf{x},\mathbf{y}\right)$; we know there exists a joint PDF $p_{\mathbf{x},\mathbf{y}}\left(\mathbf{x},\mathbf{y}\right)$ which describes them. We have interest in estimating ***x***, but we do not have measurements of it. We do not have special interest in knowing about y although we are able to measure (usually imperfectly) it.
We do not know how the joint PDF $p_{\mathbf{x},\mathbf{y}}\left(\mathbf{x},\mathbf{y}\right)$ is, and, what is very convenient, we do not need to know it.

This is our situation:

* We have interest in estimating the RV $\mathbf{x}$.
* We can get samples (e.g. measurements) of the second RV, $\mathbf{y}$.
* We know, approximately, the functional relationship between both variables.
* Both RVs, $\mathbf{x}$ and $\mathbf{y}$, are not independent and then there exists a joint PDF, $p_{\mathbf{x},\mathbf{y}}\left(\mathbf{x},\mathbf{y}\right)$, but we do not know it.

By assuming the existence of $p_{\mathbf{x},\mathbf{y}}\left(\mathbf{x},\mathbf{y}\right)$ we know that (E8) is a fact. We do not need and will never need to explicitly know $p_{\mathbf{x},\mathbf{y}}\left(\mathbf{x},\mathbf{y}\right)$. We will exploit (E8) in order to generate a belief about the RV ***x***, based on our previous belief about ***x*** and on subsequent observations, i.e. based on certain sources of information available to us (e.g. samples of ***y***).
A key source of information for applying (E8) is $p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}\,|\,\mathbf{x}\right)$. This factor is called Conditional Probability of ***y*** given ***x***. However, if we set the variable ***y*** to a value, that function turns to be a function of ***x***. This function is called LIKELIHOOD function about ***x***. This could be the result of having obtained a measurement or sample of ***y***. This likelihood function is based on assuming that ***y*** has certain specific value. The likelihood is a function of the RV ***x***, and it tells us *how likely* ***x*** is of taking the different values in its domain, just based on an individual source of information, such as the case of a measurement of ***y***. A likelihood function is not a PDF but is well related to it.

Intuitive example of a Likelihood Function

> Suppose I measure my distance to a landmark (whose ***global*** position is well known by me). I measure r =10 meters, exactly (free of noise, because I use an ideal perfect sensor). I also perfectly know that that landmark is located at position $\left(x_b, y_b\right)=\left(5,8\right)$ (expressed in meters in a reference coordinate frame). This measurement can provide a likelihood function about my position

(expressed in the same reference coordinate frame). The likelihood function will be equal to zero everywhere in its domain of (x, y), except at the points whose distance to the landmark is exactly 10 meters; for those values, I set the likelihood function be equal to one (1). I do this, for indicating which points are feasible, and which ones are not.
Now, in a more realistic case, suppose I know that the measurements provided by the range sensor have errors, which are bounded in the range of ± 1cm; in such a case I would define a likelihood that will be zero at points that do not belong to a ring of thickness 2cm, centered at point (5,8), i.e. as follows, (where both variables, ***(x,y)***, are expressed in meters)

$$L(x, y) = p_{r|y,x}(r = 10 \mid x, y) = \begin{cases} = 1 & \forall (x, y) \quad / \quad 9.99 \le \sqrt{(x-5)^2 + (y-8)^2} \le 10.01 \\ = 0 & otherwise \end{cases}$$

I produce this likelihood function based on the measurement of a RV, i.e. ***r***. I assume the existence of a joint PDF $p_{r,y,x}(r, y, x)$ and I also have (propose) $p_{r|y,x}(r \mid x, y)$, then I can use (E8) to improve my belief about (x, y).
I would repeat the same procedure for any other source of information that allows me to define a likelihood function about the variables (x, y). It may be due to measurements of my distance to other landmarks or to the same one in other measurement instances, or even because, for example, my mother also tells me something about my position: "you are in the kitchen".

Then the likelihood based on my mother's observation "you are in the kitchen" could be as follows:

$$L_{\substack{thanks \\ to \\ mama}}(x, y) = \begin{cases} = 1 & \forall (x, y) \in Kitchen \\ = 0 & \forall (x, y) \notin Kitchen \end{cases}$$

This is because I assumed my mother is a 100% safe source of information. As I suspect that, sometimes, she may be wrong, I may propose a more conservative likelihood

$$L_{\substack{thanks \\ to \\ mama}}(x, y) = \begin{cases} = 1 & \forall (x, y) \in Kitchen \\ = 0.01 & \forall (x, y) \notin Kitchen \end{cases}$$

Note: I do not request the integral of the likelihood function, over the domain of (x, y), to be =1; it is just a likelihood function not a PDF about (x, y)**.**
Now, we need a way for combining all the available likelihood functions, to generate an estimation about the RV of interest, i.e. (x, y), based on all the available information that we can collect.

Note that in this example the variable ***r*** had the role of the variable ***y*** and the 2D variable (x, y) had the role of the variable ***x***, which were used in equation (E8).

{ End of Example }

Now, we focus our attention on the general case, about estimating a RV ***x*** based on measurements of a RV ***y.***
Suppose we have the opportunity to obtain samples (e.g. measurements) of the RV ***y***, and that we also know the conditional PDF of ***y*** given ***x*** (i.e. $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x})$). Then we can apply (E8) recursively, each time we get a sample of ***y***, in order to improve our belief about ***x***, i.e. in order to obtain a PDF $p_{\mathbf{x}}(\mathbf{x})$. Each time we get a sample of ***y*** we produce a likelihood function about ***x***; then, we use it to refine our belief about ***x***, i.e. we *update* the function $p_{\mathbf{x}}(\mathbf{x})$, our previous belief.

Initially, we propose a PRIOR for $p_{\mathbf{x}}(\mathbf{x})$. It could be a completely uninformative one, e.g. a uniform PDF; let's call this prior PDF $p_{\mathbf{x},0}(\mathbf{x})$. We consider the first sample of ***y***; let's call it y_0, proceeding according to the following operation, based on (E8):

$$p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x} \mid \mathbf{y}=\mathbf{y_0}\right)=\frac{p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}=\mathbf{y_0} \mid \mathbf{x}\right) \cdot p_{\mathbf{x},0}\left(\mathbf{x}\right)}{p_{\mathbf{y}}\left(\mathbf{y_0}\right)} \qquad \text{(E9)}$$

In (E9), the components $p_{\mathbf{x},0}(\mathbf{x})$ and $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}=\mathbf{y_0} \mid \mathbf{x})$ are known. The factor $p_{\mathbf{y}}(\mathbf{y_0})$ is just a constant value (i.e. it is not a function of ***x***) and its mission is to normalize the resulting function (then the result of (E9) will be a PDF of ***x***, i.e. satisfying the condition of having "volume =1").
If required, that value can be obtained as follows,

$$p_{\mathbf{y}}\left(\mathbf{y_0}\right)=\int_{\text{domain of } \mathbf{x}} p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}=\mathbf{y_0} \mid \mathbf{x}\right) \cdot p_{\mathbf{x},0}\left(\mathbf{x}\right) \cdot d\mathbf{x}$$

This fact is very convenient, because we do not know the expression of $p_{\mathbf{y}}(\mathbf{y})$. We just know the expression of $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x})$ and a measurement of ***y***.
What is the factor $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}=\mathbf{y_0} \mid \mathbf{x})$? It is called "the likelihood function" of $\mathbf{x}$ given that we assume $\mathbf{y}=\mathbf{y_0}$. Note that it is the conditional probability of ***y*** given ***x*** evaluated at certain value of ***y***. How can we obtain it? We already had some discussion before, in an intuitive example. We will see, later, a formal way for obtaining it.
What is this new PDF $p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}=\mathbf{y_0})$? It is our new belief (the so called "*posterior*") about the RV ***x***, based on the combination of our previous belief $p_{\mathbf{x},0}(\mathbf{x})$ (the "*prior*"), and on the new source of information about $\mathbf{x}$, the likelihood function $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}=\mathbf{y_0} \mid \mathbf{x})$.
In practical terms we can exploit a more convenient operation:

$$p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x} \mid \mathbf{y}=\mathbf{y_0}\right) \propto p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}=\mathbf{y_0} \mid \mathbf{x}\right) \cdot p_{\mathbf{x},0}\left(\mathbf{x}\right) \qquad \text{(E10)},$$

because we can always normalize the resulting PDF when we need to really use it, and not when we are synthesizing it. (note: the operator $\propto$ means *proportional*)
As we usually have more samples of y (often, different values because we may be measuring , multiple times, some true value of y polluted with white noise), we repeat the process by processing these new samples,

$$p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x} \mid \mathbf{Y}=\left\{\mathbf{y_0},\mathbf{y_1}\right\}\right) \propto p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}=\mathbf{y_1} \mid \mathbf{x}\right) \cdot p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x} \mid \mathbf{y}=\mathbf{y_0}\right) \qquad \text{(E11)}$$

Where the member $p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x} \mid \mathbf{Y}=\left\{\mathbf{y_0},\mathbf{y_1}\right\}\right)$ express the probability of $\mathbf{x}$ (our belief about $\mathbf{x}$) given that we have sampled the RV $\mathbf{y}$ twice, obtaining the value $\mathbf{y}=\mathbf{y_0}$, and then $\mathbf{y}=\mathbf{y}_1$. We note that in this operation, for processing a new available likelihood function, we use as *prior* the previously obtained *posterior*.
In general, for a set of measurements, we recursively iterate as follows

$$p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x} \mid \mathbf{Y}=\left\{\mathbf{y}_i\right\}_{i=0}^{K}\right) \propto p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}=\mathbf{y}_K \mid \mathbf{x}\right) \cdot p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x} \mid \mathbf{Y}=\left\{\mathbf{y}_i\right\}_{i=0}^{K-1}\right) \qquad \text{(E12)}$$

Where the RV Y means certain set of measurements.

We have described how to process a sequence of samples of the same RV (e.g. certain output of a system). However, we can apply the same procedure for any sequence of likelihood functions we can obtain. These likelihood functions may come from measuring or sampling the same RV or different ones. This estimation process is able to accept any source of information, because it processes likelihood functions of the RV ***x***; it is not constrained to processing a specific measurement or source of information.
In some problems, certain sources of likelihood are not even produced by a measurement at all. In a previous example we mentioned that we could produce likelihood functions from measuring my distance to a landmark several times, and by measuring my distance to other landmarks, or by producing a likelihood based on a comment

from my mother. Each likelihood function can be used as in (E12). We could be more general by expressing it as follows:

$$p_{\mathbf{x}}^{k}\left(\mathbf{x}\right) \propto L_{k}\left(\mathbf{x}\right)\cdot p_{\mathbf{x}}^{k-1}\left(\mathbf{x}\right)$$

In which $p_{\mathbf{x}}^{i}\left(\mathbf{x}\right)$ would mean the belief about x based on the sequence of likelihood functions $\left\{L_{j}\left(\mathbf{x}\right)\right\}_{j=1}^{i}$

Note that this is a simplified case of estimation, as in this case our estimation process does not include the dynamics of the system (because the state being estimated is time invariant). The general case will be discussed later (i.e. for treating time varying states).

Note: The likelihood functions we produce are usually the result of observations which involve noise contributions, such as it was expressed in (E3), i.e. $\mathbf{y}\left(t_{i}\right)=h\left(\mathbf{x}\left(t_{i}\right),\boldsymbol{\eta}\left(t_{i}\right)\right)$, in which the noise $\boldsymbol{\eta}\left(t\right)$ is assumed to be white noise.

Applying the Estimation Approach

An estimation process, like the one previously discussed, was applied in the simulated experiment for the localization of the stationary robot, based on a "range-only" sensor. In that case, the position variables $\left(x,y\right)$ (a 2D vector) correspond to the variable $\mathbf{x}$ in (E12); the range measurements correspond to the vector $\boldsymbol{Y}$. (each $\mathbf{y}_{i}$ would be a range measurement).

The likelihood function is defined by the observation function $h\left(\ \right)$ and by the PDF that describes the noise, which is present in that observation. For the particular observation of the landmark #a, located at a well-known position, $\left(x_{a},y_{a}\right)$, we have that the range function is

$$r_{a}=h\left(x,y\right)+\eta_{a}=\sqrt{\left(x-x_{a}\right)^{2}+\left(y-y_{a}\right)^{2}}+\eta_{a} \qquad \text{(E13)}$$

where we model the noise η_{a} as a RV with PDF $p_{\eta_{a}}\left(\eta_{a}\right)$ (we can get an idea about this PDF by analyzing the statistics of the sensor error). Based on $h\left(\ \right)$ and $p_{\eta_{a}}\left(\eta_{a}\right)$, we are able to get the likelihood function $p_{r_{a}|x,y}\left(r_{a}\,|\,x,y\right)$.

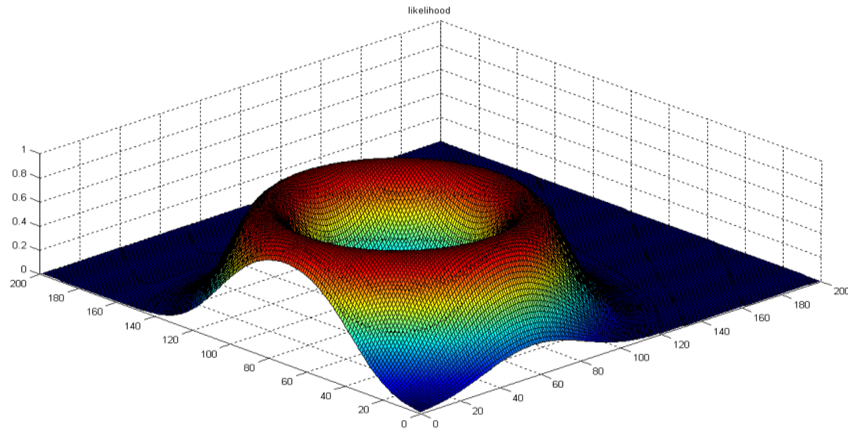
Note that, in this case, the observation is a scalar (i.e. 1D) and the estimated variables are in 2D. It can be demonstrated (we will discuss it later) that, in this case, the likelihood function takes the following expression:

$$p_{r_{a}|x,y}\left(r_{a}\,|\,x,y\right)=p_{\eta_{a}}\left(r_{a}-\sqrt{\left(x-x_{a}\right)^{2}+\left(y-y_{a}\right)^{2}}\right) \qquad \text{(E14)}$$

Suppose that we measure $r_{a}=5m$, and that we know that the landmark is located at $\left(x_{a},y_{a}\right)=\left(95{,}100\right)$, then we obtain the following associated likelihood function, about $\left(x,y\right)$,

$$p_{r_{a}|x,y}\left(r_{a}=5\,|\,x,y\right)=p_{\eta_{a}}\left(5-\sqrt{\left(x-95\right)^{2}+\left(y-100\right)^{2}}\right) \qquad \text{(E15)}$$

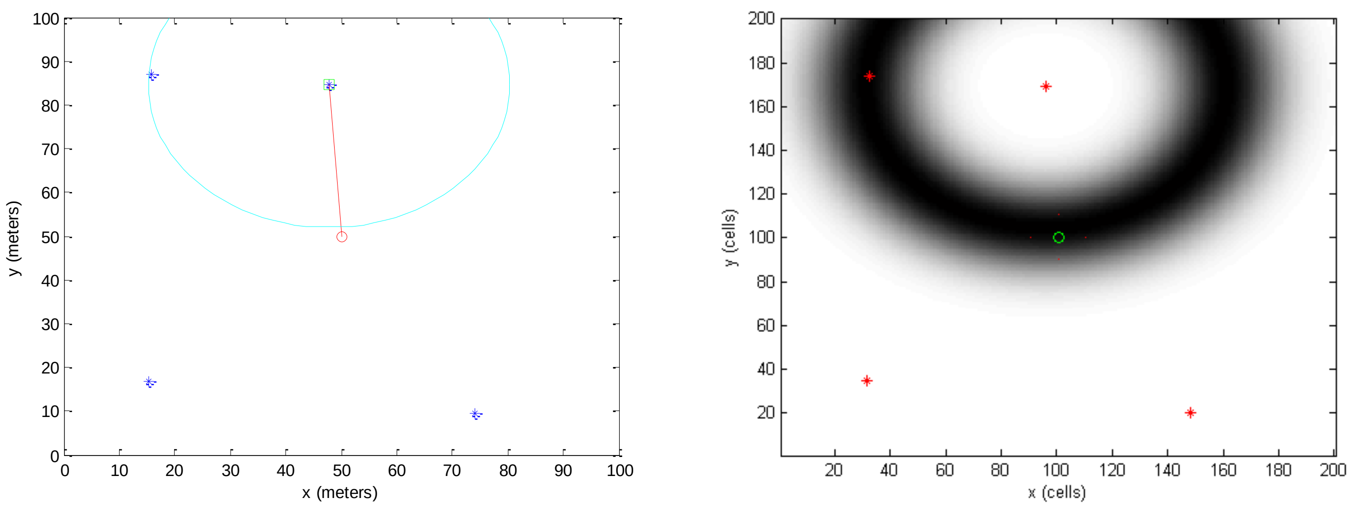
In this expression, all the components are known except the variables $\left(x,y\right)$, i.e. $p_{\eta_{a}}\left(.\right),x_{a},y_{a},r_{a}$ are known. The likelihood function in (E15) can be used for updating the prior PDF about $\left(x,y\right)$.



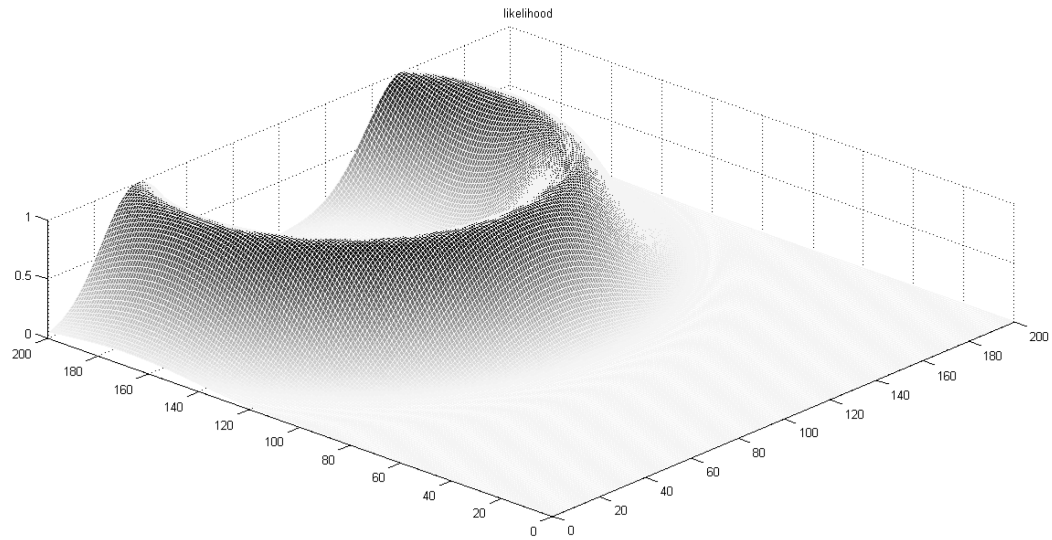
This function shows the likelihood of an individual range observation. The range observation reports that the robot is at certain distance from landmark #a. This observation defines a circular ridge (which surrounds the landmark), where the robot may likely be located. The shapes of the ridge sections are defined by the PDF of the measurement noise, $p_{\eta_a}(.)$, which, in this case, has a peak centered at zero (in this example we assumed the noise η_a has a zero mean Gaussian distribution, having an exaggerated standard deviation).

An Example of Applying a Sequence of Range Observations

In this simulated range-only localization process, we apply a sequence of polluted range measurements associated to four (4) known landmarks. Each time that we apply an observation, we see the associated likelihood function and the current PDF, which describes the robot position.
We will show it, in more detail, through a Matlab application during the lecture. We have included some pictures in this document; however, they show just part of the story (which is better shown in a video or a sequence of images).

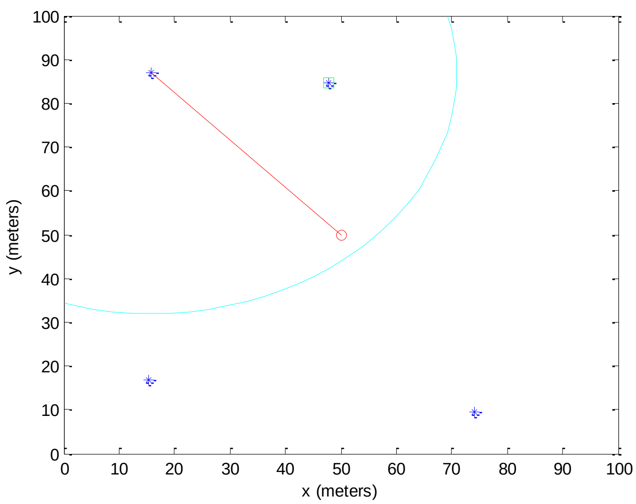


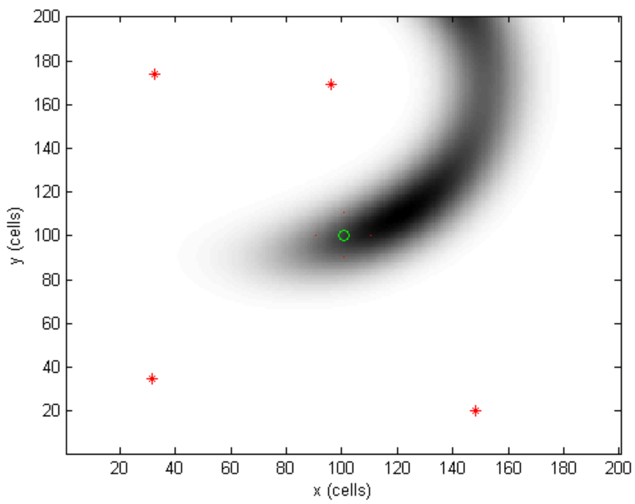
In the left figure, our platform (little red circle) detects a landmark, and then measures its distance to it. The real distance is shown by a red segment. However, due to the noise, the measured distance is, in this case, a bit shorter. Based on that measurement, the platform guess that its own position should be nominally located on the light blue circumference, an assumption which would be wrong. However, the likelihood function is able to consider that the measurement is polluted with a noise of certain statistical properties, consequently allowing to accept other feasible points nearby the nominal circumference.

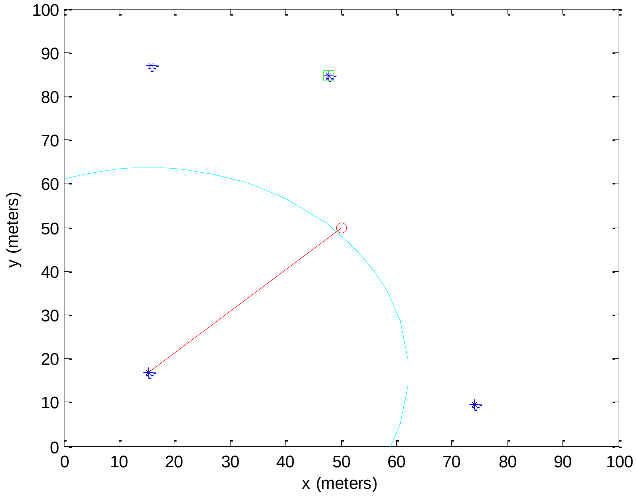


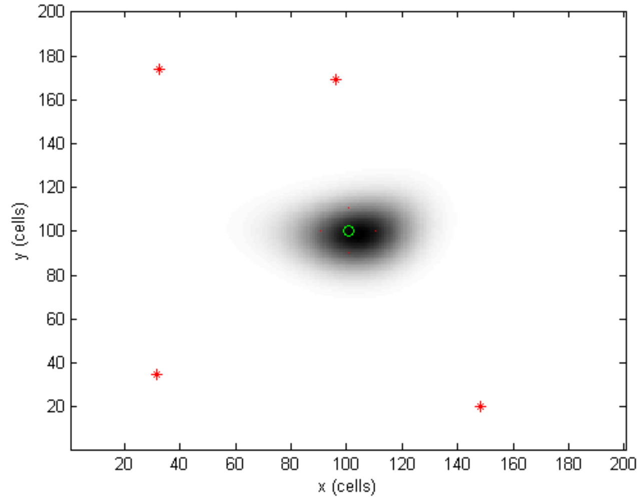
This is the PDF about the robot position, after we applied the first available likelihood. The Prior PDF was uniform in the ROI (constant value everywhere inside the Region of Interest (ROI)). After applying the first Bayesian update, it takes the shape of the truncated likelihood. The color convention: Dark means high density values, light colors (e.g. light gray or white) mean low values of probability density.
The measured distance, between the robot and one landmark, produces the likelihood on the top right figure. It should be a circular ridge. We have truncated it to be kept inside certain ROI. As it can be seen, in the left figure, the measured range defines a circle centered at the landmark's position. The circle does not intersect the true (but unknown) position of the sensor; this is due to the presence of noise polluting that measurement.

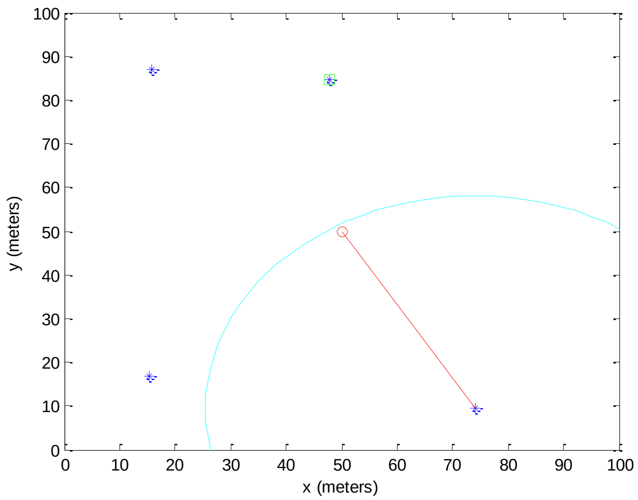
Now we apply more observations, the result can be seen in the following figures,

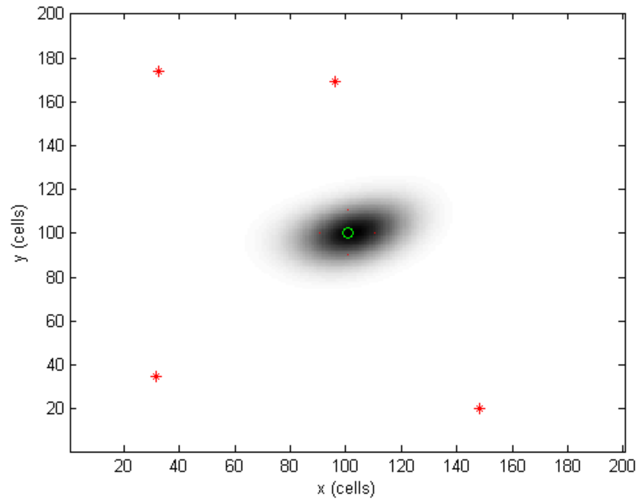




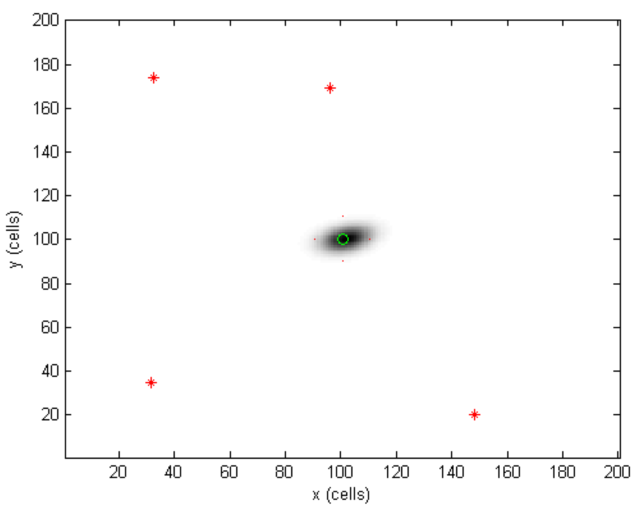


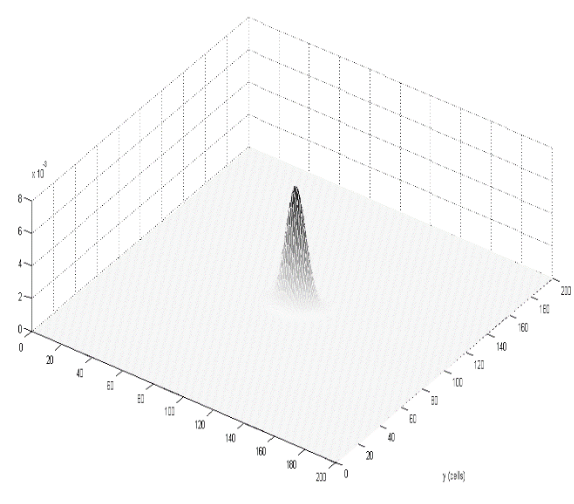




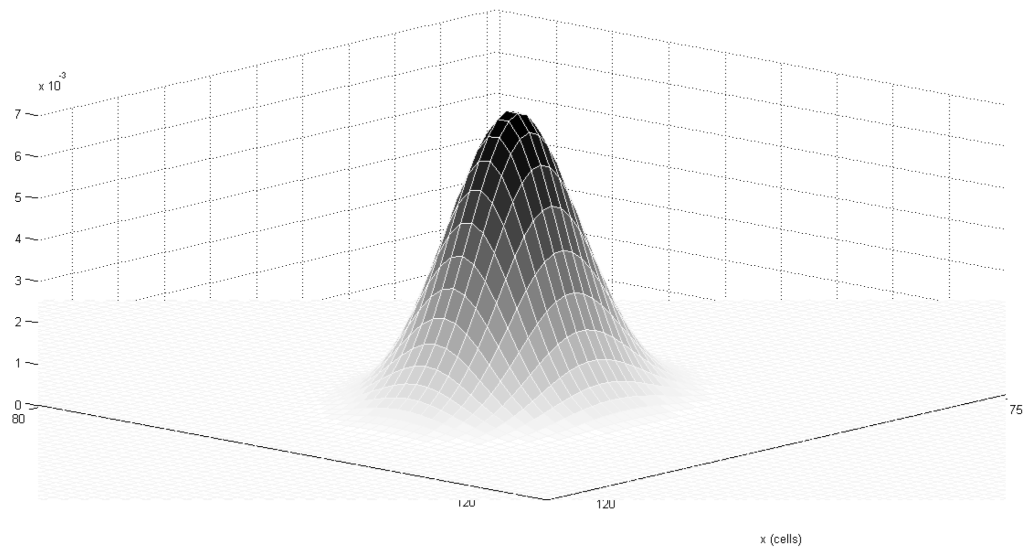


The previous figures show the evolution of the synthesized PDF, as result of fusing the first observations of the 2nd, 3rd and 4th landmarks. The left figures indicate the related landmark being observed, and, on the right, the resulting PDF after processing the observation. The associated likelihood functions are not shown for the sake of saving space in this document; however, their shapes would be like the one shown for the first update (but of different radii and centers).





After a sequence of three scans (each of them observing the four landmarks, totalizing 12 observations), we get an improved belief about the robot position. The real (but unknown) position of the robot is indicated by the small blue circle (a bit obscured by the black cloud, which represents the PDF values). The four red asterisks are the position of the map's landmarks.



Detail of the synthesized PDF, shown in the previous figure.

Videos to see: *"MTRN4010_2021_BayesLocaliza01.mp4"* (https://youtu.be/8jfGG4ZZ9C0)

To see in class: a more complex case.

Let's consider a similar localization problem, in which we have a stationary robot which does process range measurements to certain beacons; however, we add some extra difficulty: the sonar measurements do not include information about which landmark is producing the "ping". This situation is usual in cases in which we measure the ToF of reflected acoustic signals, in which the sonar does emit signals and measure the time of the detected reflections. For each emitted ping, the robot measures the times of the multiple reflections (multiple objects). Each peak would correspond to the reflection from certain object, which may be one of the known landmarks. However, we do not know which landmark is associated to each reflection. For a distance (i.e. measured time elapsed until certain reflected peak is detected), how should be its associated likelihood function? We will see the answer in class, running a simulation.
(a video can be watched here, for just a particular case: https://youtu.be/rTMRgNX-J_I)

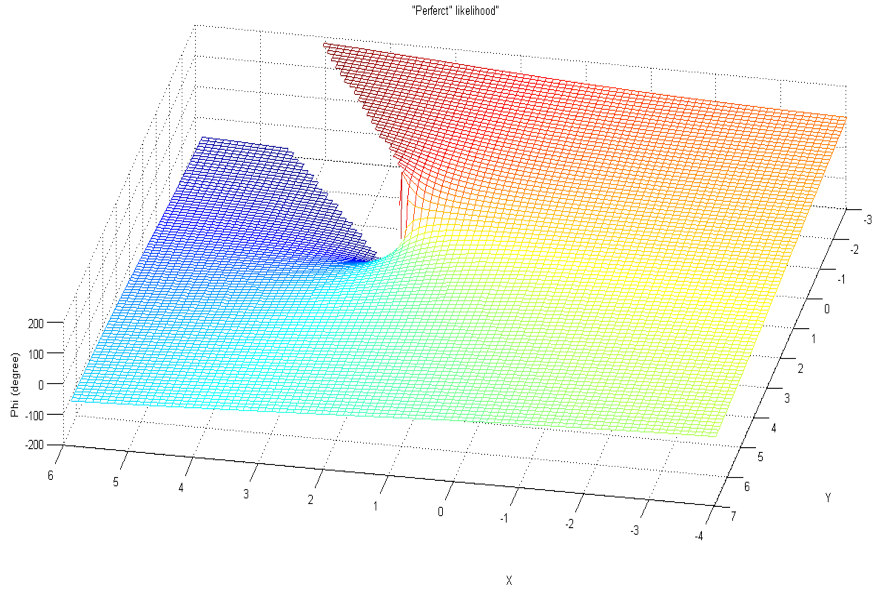
To Discuss in class

Relating this topic with previous Concepts (NOT TO BE DISCUSSED in MTRN4010)
As you have seen, the positions of the reference landmarks have a relevant effect on the resulting estimates of the robot's position. Does this fact have something in common with the DOP (Dilution of Precision) term, which is usually associated to GPS measurements?

Imagine a 3D likelihood function, associated to the following measurements:

a) Bearing Only Measurement

Discuss how should be the likelihood function for a perfect bearing observation. This observation involves three (3) states (i.e. 3DoF: x, y, heading).



This is how a likelihood function $p_{\alpha|(x,y,\phi)}(\alpha = \alpha_0 \mid x, y, \phi)$ would look if we had a perfect bearing observation (no noise). This means that $p_{\alpha|(x,y,\phi)}(\alpha = \alpha_0 \mid x, y, \phi)$ would be equal to zero at points not belonging to this surface. By intersecting 3 surfaces (like this one, but not identical to it, which should be obtained from observations of different landmarks) we can get a unique solution.

This example makes sense fora 3DoF localization problem, in which we estimate the pose of a platform (2D position and heading) .

If we had a bearing observation polluted with noise, then its individual likelihood would spread through the space x, y, ϕ. Think about a “sandwich” of layers generating a volume surrounding this surface.

---

b) GPS Pseudo Range Measurement (assuming a perfect receiver’s Clock and a perfectly known satellite position) ***b.1) Perfect range measurement***

b.2) Polluted range measurement

→To discuss in class.

---

c) Imagine a 4D likelihood function associated to a GPS Pseudo Range Measurement (not assuming a perfect receiver’s Clock) → To discuss in class.

Appendix 1: Definition of the likelihood function for the Localization Example

We obtained the following likelihood

$$p_{r_a|(x,y)}(r_a \mid x, y) = p_{\eta_a}\left(r_a - \sqrt{(x - x_a)^2 + (y - y_a)^2}\right)$$

by applying the following general rule

$$z = f(\mathbf{x}) + \eta$$
$$\mathbf{x} \in R^N, \quad \eta \in R^1, \quad z \in R^1$$
$$\mathbf{x}, \eta \text{ are independent}$$
$$\Downarrow$$
$$p_{z|\mathbf{x}}(z \mid \mathbf{x}) = p_\eta(z - f(\mathbf{x}))$$

This rule is demonstrated in ***Appendix 2***.

For our particular case, $r_a = \sqrt{(x - x_a)^2 + (y - y_a)^2} + \eta_a$, the random variables $\boldsymbol{z}$, $\boldsymbol{x}$ and η would correspond to the RVs r_a, $\begin{bmatrix} x \\ y \end{bmatrix}$ and η_a respectively.

Appendix 2: General Rule for defining Certain Likelihood Functions

We consider the following rule:

$$z = f(\boldsymbol{x}) + \eta$$
$$\boldsymbol{x} \in R^N, \quad z \in R^1, \eta \in R^1$$
$$\boldsymbol{x}, \boldsymbol{\eta} \text{ are independent RVs}$$
$$\text{then}$$

$$p_{z|\boldsymbol{x}}(z|\boldsymbol{x}) = p_\eta\big(z - f(\boldsymbol{x})\big)$$

We remark that this rule is valid for scalar observations.

Demonstration

We have that a measured scalar output is:

$$z = f(\boldsymbol{x}) + \eta$$

In which, the variable z is equal to a function of $\boldsymbol{x}$, plus certain independent noise η.
Consequently, the conditional PDF of z given $(\boldsymbol{x}, \eta)$ is

$$p_{z|\boldsymbol{X},\eta}(z|\boldsymbol{x}, \eta) = \delta(\boldsymbol{z} - f(\boldsymbol{x}) - \eta)$$

In which the function $\delta(.)$ is the Dirac's delta. This is because $\mathbf{z}$ is perfectly defined by $(\boldsymbol{x}, \eta)$.
We can know the joint PDF $p_{z,\boldsymbol{X},\eta}(\boldsymbol{z}, \boldsymbol{x}, \eta) = p_{z|\boldsymbol{X},\eta}(\boldsymbol{z}|\boldsymbol{x}, \eta) \cdot p_{\boldsymbol{X},\eta}(\boldsymbol{x}, \eta)$

As we simply want to obtain $p_{z|\boldsymbol{x}}(z|\boldsymbol{x})$, we can obtain it by proper marginalization.

$$p_{z|\boldsymbol{x}}(z|\boldsymbol{x}) = \int_{\substack{\text{full domain} \\ \text{of } \eta}} p_{z|\boldsymbol{X},\eta}(z|\boldsymbol{x}, \eta) \cdot p_\eta(\eta|\boldsymbol{x}) \cdot d\eta$$

$$\boldsymbol{x}, \eta \text{ are independent RVs} \Rightarrow p_{\boldsymbol{X},\eta}(\boldsymbol{x}, \eta) = p_\eta(\eta) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}) \Rightarrow p_{\eta|\boldsymbol{X}}(\eta|\boldsymbol{x}) = p_\eta(\eta)$$

$$p_{z|\boldsymbol{X}}(z|\boldsymbol{x}) = \int_{\substack{\text{all Domain} \\ \text{of } \eta}} p_{\boldsymbol{z}|x,\eta}(z|\boldsymbol{x}, \eta) \cdot p_\eta(\eta) \cdot d\eta =$$

$$= \int_{\substack{\text{full domain} \\ \text{of } \eta}} \delta(z - f(\boldsymbol{x}) - \eta) \cdot p_{\boldsymbol{\eta}}(\eta) \cdot d\eta =$$
$$= p_{\boldsymbol{\eta}}\big(z - f(\boldsymbol{x})\big)$$

(Also remember, from Mathematics: $\int_{\substack{\text{Full domain} \\ \text{of } \eta}} \delta(a - \eta) \cdot f(\eta) \cdot d\eta = f(a)$)

Note: This appendix is provided just for showing you how this rule is demonstrated. This demonstration will not be required in any exam, in MTRN4010.

Note 2: There are cases in which relationship between the measured variable and the hidden variable (the state which we want to know) may not be explicit but an implicit relationship, $f(\boldsymbol{x}, z, \eta) = 0$; that would still result in a likelihood function; however, we do not focus our attention on those cases.