Let's use PDFs

Consider this case: 1D localization of a vehicle.

(moving in a 1D universe),

Its initial 1D position is x=x0, which is known by us.

We have speed measurements.

τ

Nominally, we would implement predictions as follows:

$$x(t+\tau) \cong x(t) + \tau \cdot \dot{x}(t)$$

$$\dot{x}(t) = v(t) = v_{measured}(t)$$

In which the sampling period τ is small enough, so that we can assume the discrete time model is free of error.

Let's use PDFs

Some complication: speed measurements are polluted with noise.

$$x(t_0+\tau)=x(t_0)+\tau\cdot v(t_0)$$

$$v_{measured}(t_0)=v(t_0)+\xi_v(t_0) \qquad \left(\text{measurement is polluted with noise } \xi_v(t_0)\right)$$

$$\hat{x}(t_0)=x(t_0) \qquad \left(\text{we know the initial condition}\right)$$

$$\Downarrow$$

$$\hat{x}(t_0+\tau)=\hat{x}(t_0)+\tau\cdot v_{measured}(t_0) \qquad \left(\text{prediction, 1 step}\right)$$

$$\hat{x}(t_0+\tau)=x(t_0+\tau)+\tau\cdot\xi_v(t_0)$$

Now, in a 2-steps prediction,

$$\hat{x}(t_0+2\tau)=\hat{x}(t_0+\tau)+\tau\cdot v_{measured}(t_0+\tau)$$

$$\left(\; v_{measured}(t_0+\tau)=v(t_0+\tau)+\xi_v(t_0+\tau) \;\right)$$

$$\hat{x}(t_0+2\tau)=x(t_0+2\tau)+\tau\cdot\xi_v(t_0)+\tau\cdot\xi_v(t_0+\tau)$$

$$\hat{x}(t_0+2\tau)=x(t_0+2\tau)+\xi_x(t_0+2\tau)$$

$$\hat{x}(t_0+2\tau)=x(t_0+2\tau)+\xi_x(t_0+2\tau)$$

(estimate=real value+"discrepancy")

$$\xi_x(t_0+2\tau)=\tau\cdot\xi_v(t_0)+\tau\cdot\xi_v(t_0+\tau)$$

(discrepancy $\xi_x(t)$: we do not know its actual value, but we may have a PDF describing it)

If we keep predicting,

$$\widehat{x}(t) = x(t) + \xi_x(t)$$

$$\xi_x(t) = \tau \cdot \sum_{i=1}^{n} \xi_v(t_i)$$

Our error at time t, is due to the addition of all the errors of the measured speed, since we started the prediction process.

How do we know the PDF of $\xi_x(t)$?

How do we know the PDF of $\xi_x(t)$?

$$p_{\xi_x(t)}\left(\xi_x(t)\right)$$

It can be calculated.

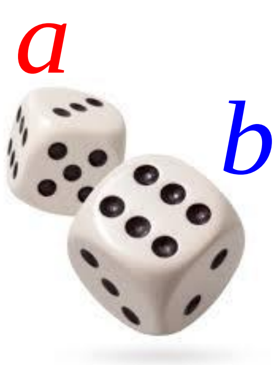
First, we need to understand what happens when we add two RVs

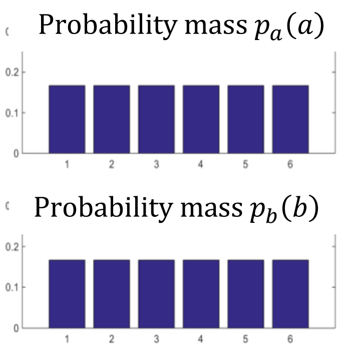
We assume that the sequence of errors in the speed measurements is "white", i.e. that

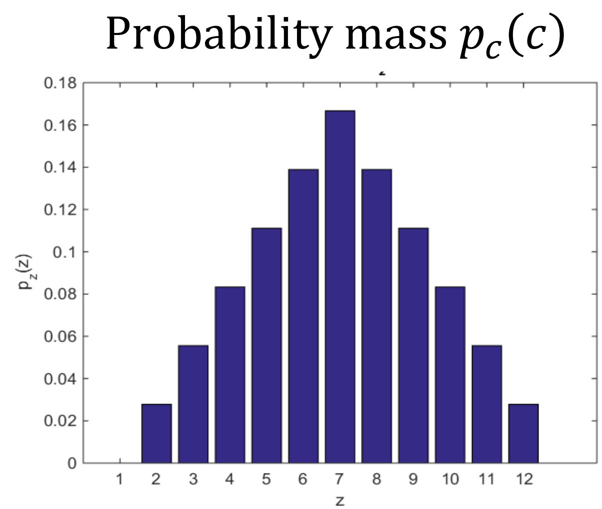
Any couple of RVs $\xi_v(t_i), \xi_v(t_k)$ are statistically independent (for $t_i \neq t_k$)

Now, do you remember the “2-steps” dice game?

$$\begin{cases} p_a(a) \\ p_b(b) \\ c = a + b \end{cases}$$







$p_c(c)$

Any connection with this case? (yes, we are adding RVs)

We have the statistical process of adding independent RVs,

What happen when we add two independent analog RVs?

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

It can be demonstrated that

$$p_z(z) = \int_{\substack{\text{domain of} \\ x}} p_x(x) \cdot p_y(z-x) \cdot dx$$

$$= \int_{\substack{\text{domain of} \\ y}} p_y(y) \cdot p_x(z-y) \cdot dy$$

The relevant conclusion from this is that

$$\left.\begin{cases} a \sim p_a(a) \\ b \sim p_b(b) \\ c = a + b \end{cases}\right\} \Rightarrow \text{we can obtain } p_c(c)$$

In addition (we will describe it later)

$$\left.\begin{cases} a \sim p_a(a) \\ c = G(a) \end{cases}\right\} \Rightarrow \text{we can obtain } p_c(c)$$

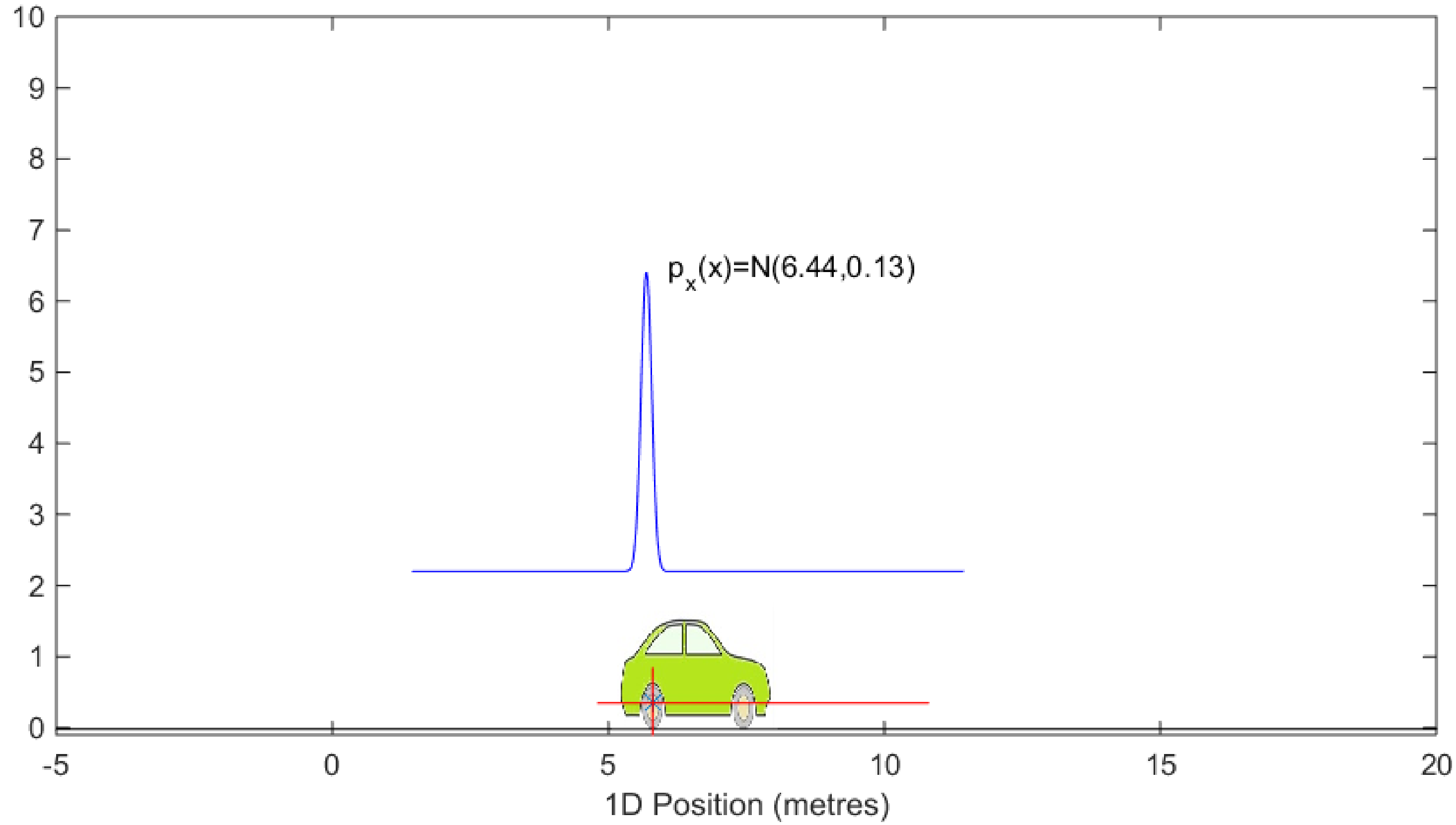
as well.

Comment1: We usually do not perform those operations by following the mentioned convolution operation)
Comment 2: When the involved PDFs are Gaussian, the actual implementations of these operations are cheap and easy (a matter to be seen later)

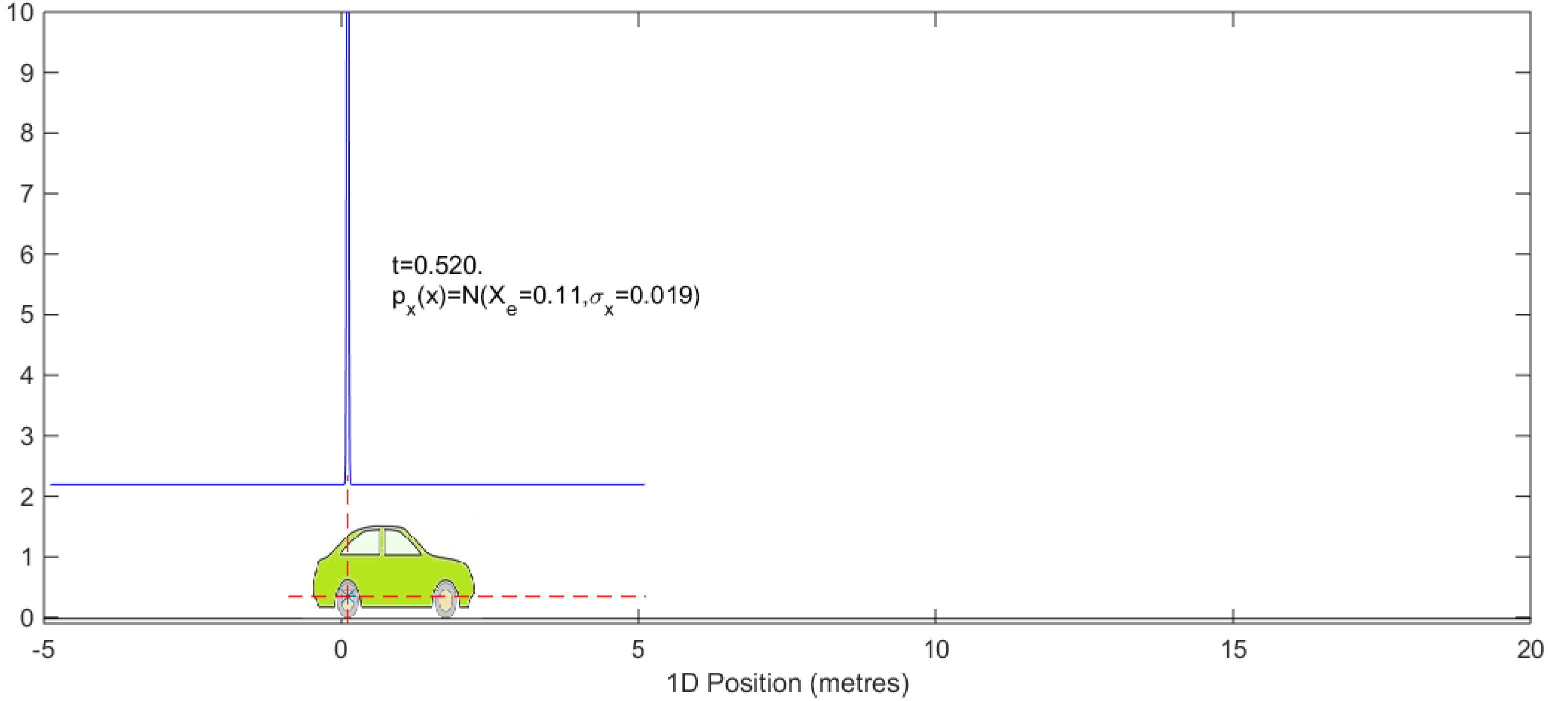
Let’s see the usual effect of the noise, on the quality of the predictions:

Case: Accurate initial position; but speed sensor introducing some error (GWN, $\sigma_v = 0.2 metre/s$).

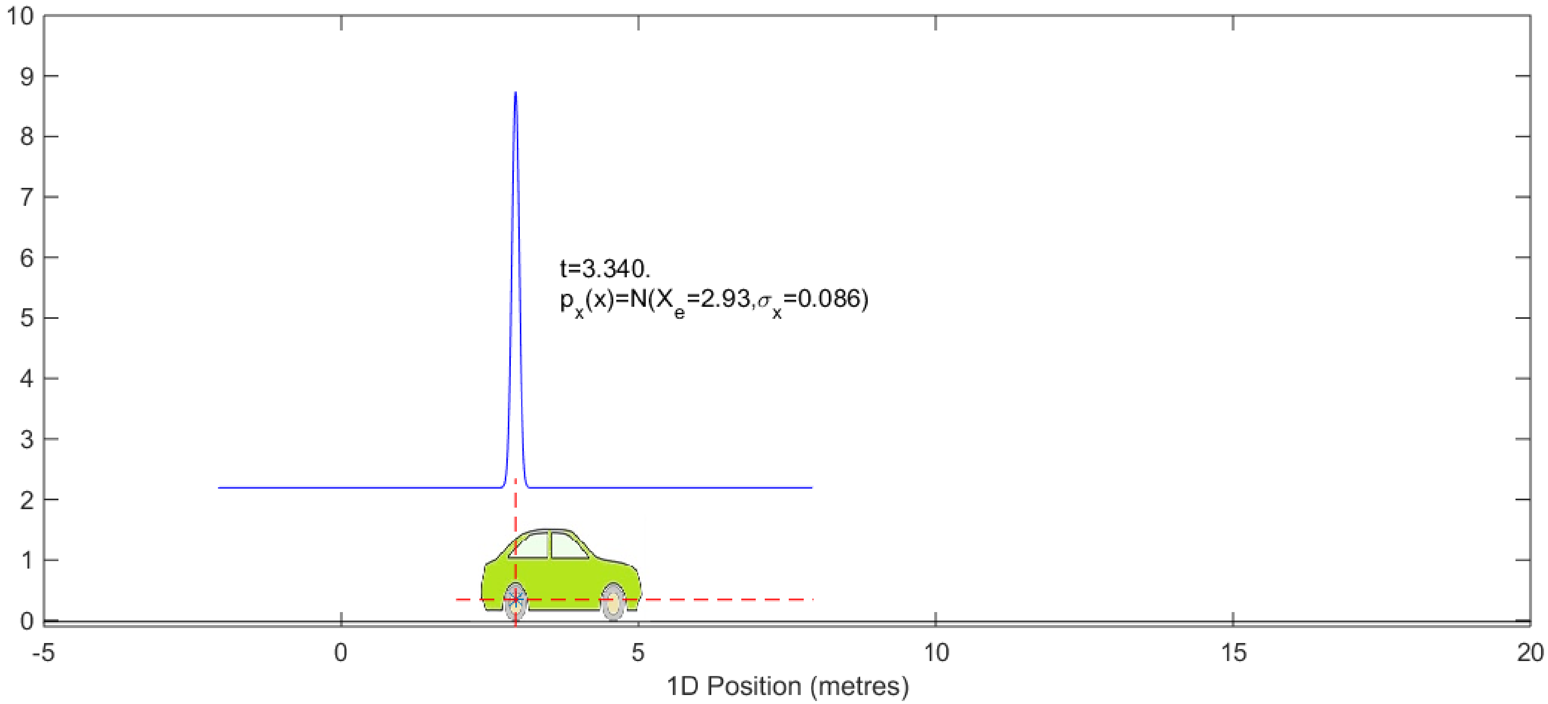
What would happen if we kept predicting position?
(we try in a brief simulation / animation ,now)



Very first prediction is of good quality, because we knew the initial condition shortly before.

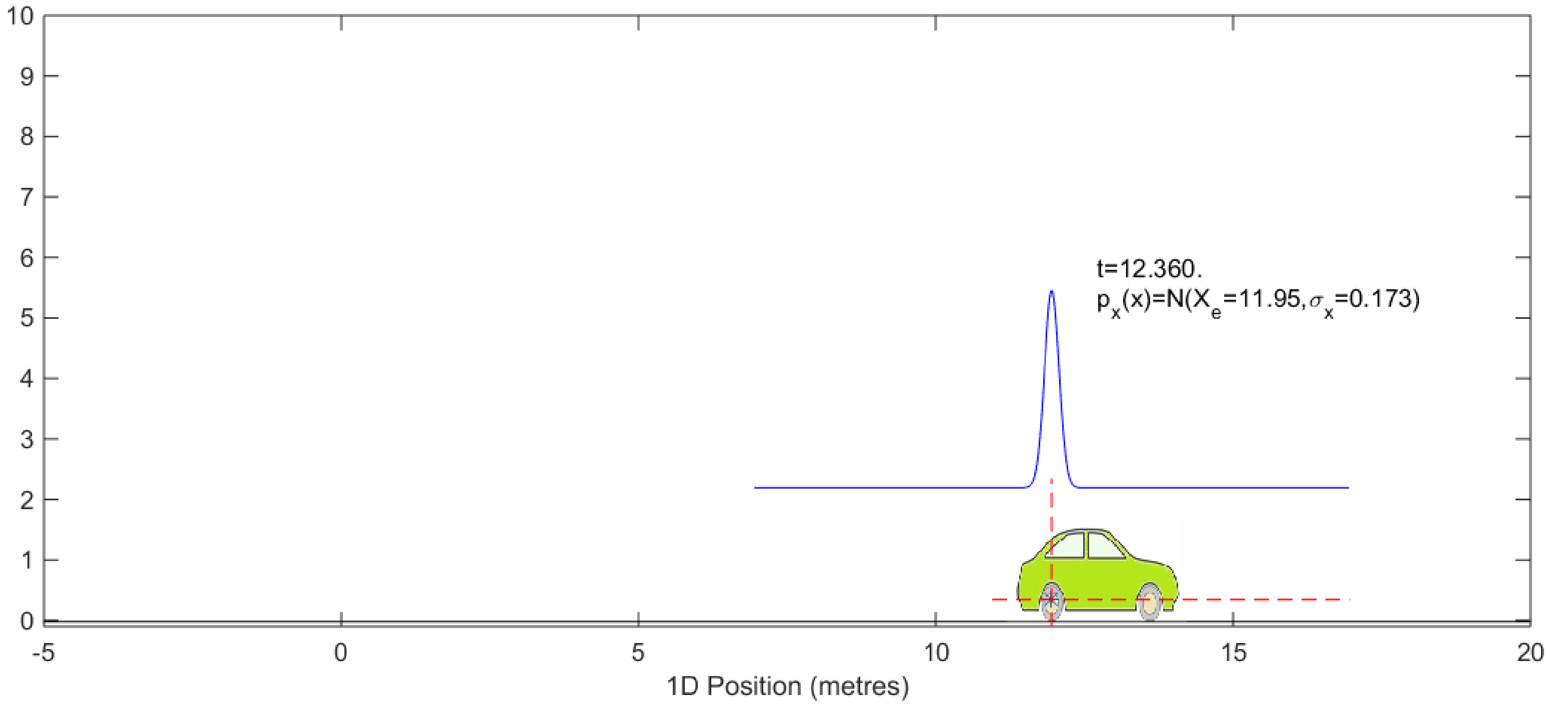


“confident PDF” : narrow, concentrating almost all probability in a small area of the domain of x, As more predictions are performed: the uncertainty grows (i.e. our confidence decreases)



We are considering the effect of the adding noise. Our PDF about the car position at time t, has its variance increased.

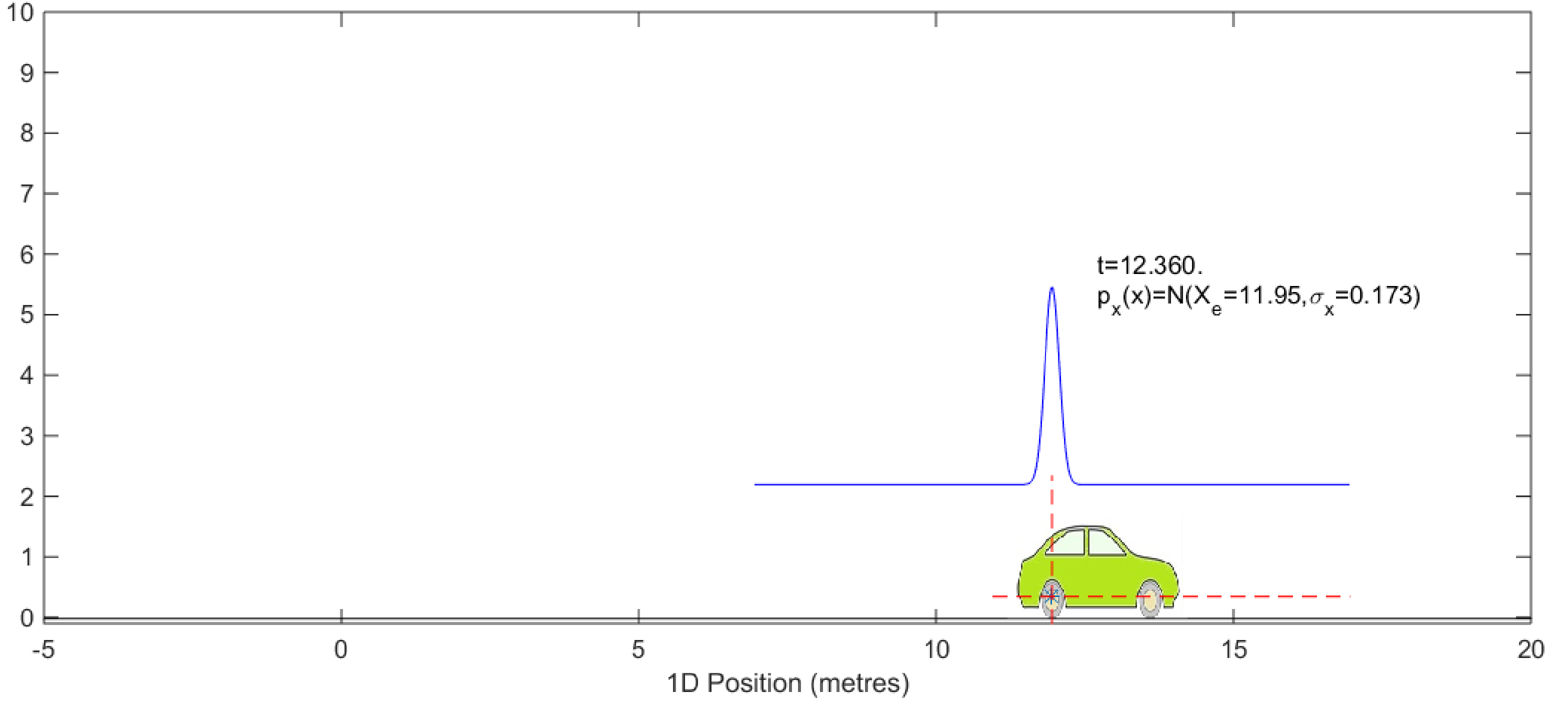
And the degradation of accuracy would continue if we simply keep predicting, due to the uncertainty which we (involuntarily) add each time we apply our imperfect process model.



If our belief about the RV x is represented by Gaussian PDF (as shown), we will appreciate how the variance does increase as our confidence decreases.

(question: what happens with the shape of the PDF?)

How to stop this degradation process?



→We will need to exploit other sources of information,
to “update” our estimates.

Until now we have briefly described the degradation of our estimate, due to applying our approximate process model (e.g. kinematic model), in this case due to using polluted measurements of speed as model's input.

We try to maintain a PDF for describing our current state x(t+T) based on the PDF of x(t), on our model and on the known values of inputs at time t.

For the moment we stop talking about predictions.

Now we will see how to use information of other sources, for performing "updates" (or "corrections")

How other sources of information do help to refine our PDF about the variables we want to estimate.

BAYES' RULE

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

This equation can be obtained by manipulating equations of conditional probabilities

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

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

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

BAYE's RULE

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

It may be a simple relationship, but its implication is relevant, because it is the support for all the Bayesian estimation theory .

But, first, we describe our whole estimation problem, to see how to make this equation to work for implementing an estimation process.

Consider a system whose state , X(t), we need to know at all times. The system follows certain dynamics, whose model we may know approximately, e.g. in this case, being a continuous time process.

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{f}\left(\mathbf{X}(t), \mathbf{u}(t)\right)$$

$$\mathbf{X}(t) \in R^n$$

We may also know its initial condition, at time t0,

$$\mathbf{X}(t_0) = \mathbf{X}_0$$

And we are usually able to measure (*) certain system's outputs, at certain times.

$$\mathbf{y}(t) = \mathbf{h}\big(\mathbf{X}(t)\big), \quad t \in \{t_1, t_2, ..\}$$

(*) Usually using sensors.

Bayesian Estimation

It seems that based on those "sources of information" we should be able to generate perfect estimates of **X**(t).

E.g. "If I know the state of the system at time t0, and I if know the process model and the inputs **u**(t) ➔ I can know **X**(t)

However, that assumption is WRONG.

Why?

Because: We do not have perfect knowledge about all or some of the assumed sources of information (initial condition, process model)

If we assume that hose sources of information are perfect when the reality is not that, then we will obtain wrong estimates of **X**(t).

We model our UNCERTAINTY using some RVs

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{f}\left(\mathbf{X}(t), \mathbf{u}(t), \boldsymbol{\xi}(t)\right)$$

$$\widehat{\mathbf{X}}_0 = \mathbf{X}_0 + \boldsymbol{\mu}$$

In which we include some sources of uncertainty to “blame”

One is to represent the error of our nominal (assumed) process model.

We call it, in our notation, $\boldsymbol{\xi}(t)$. We also represent our uncertainty about the initial condition, by the variable $\boldsymbol{\mu}$.

Of course, we do not know the values of those components, so that for the moment they are not useful for the purpose of keeping generating X(t).

Could we get some extra sources of information?

➔ YES

Usually, we have access to measurements of certain system's outputs., at certain times.

$$\mathbf{y}(t_i) = \mathbf{h}(\mathbf{X}(t_i))$$

These outputs are variables which are function of the state of the system.

BUT: we are able to know the values of those outputs, incurring in certain errors ,

$$\mathbf{y}(t_i) = \mathbf{h}(\mathbf{X}(t_i), \boldsymbol{\eta}(t_i))$$

because we do not perfectly know the function *h*(.), or because our measurements of those outputs are polluted with noise (sensors are noisy!)

Usually, those uncertainties can be assumed additive

$$\mathbf{y}(t_i) = \mathbf{h}\left(\mathbf{X}(t_i), \boldsymbol{\eta}(t_i)\right)$$

$$\Downarrow$$

$$\mathbf{y}(t_i) = \mathbf{h}\left(\mathbf{X}(t_i)\right) + \boldsymbol{\eta}(t_i)$$

Bayesian Estimation

IDEA:

We do not know the values of $\left(\boldsymbol{\xi}(t), \boldsymbol{\eta}(t), \mu\right)$

But we may know their statistical properties, e.g. we may have PDFs describing them.

$$p_{\boldsymbol{\xi}(t)}\left(\boldsymbol{\xi}(t)\right), p_{\boldsymbol{\eta}(t)}\left(\boldsymbol{\eta}(t)\right), p_{\mu}(\mu)$$

If we knew the statistical properties (i.e. the PDFs) of those RVs, then we would be able to obtain estimates of X(t); and those estimates would be expressed in a statistical way as well (i.e., through PDFs).

$$p_{\mathbf{X}(t)}\left(\mathbf{X}(t)\right)$$

How confident is that resulting PDF depends on the information we use for generating that PDF.

Bayesian Estimation

We also assume that those sources of "pollution" are independent,

$$\left(\boldsymbol{\xi}(t), \boldsymbol{\eta}(t), \mu\right)$$

$$\boldsymbol{\xi}(t) \text{ independent of } \left(\mathbf{x}(\tau), \boldsymbol{\eta}(\tau), \mu\right) \text{ for any } \tau \le t$$

$$\boldsymbol{\eta}(t) \text{ independent of } \left(\mathbf{x}(\tau), \boldsymbol{\xi}(\tau), \mu\right) \text{ for any } \tau \le t$$

We may say that $\boldsymbol{\xi}(t), \boldsymbol{\eta}(t)$ both behave as white noises

Bayesian Estimation

So, we fuse all our sources of information for obtaining a PDF about the sate of the system at current time,

$$\left.\begin{cases} p_{\mathbf{X}(t)}\left(\mathbf{X}(t)\right), p_{\boldsymbol{\xi}(t)}\left(\boldsymbol{\xi}(t)\right), p_{\boldsymbol{\eta}(t)}\left(\boldsymbol{\eta}(t)\right), p_{\mu}(\mu), \\ \text{model } \mathrm{X}(t+\tau) = \mathbf{F}\left(\mathrm{X}(t), \mathrm{u}(t), \boldsymbol{\xi}(t)\right) \\ \text{model } \mathbf{y}(t+\tau) = \mathbf{h}\left(\mathrm{X}(t+\tau), \boldsymbol{\eta}(t+\tau)\right) \\ \text{measurements of } \mathbf{y}(t+\tau) \end{cases}\right\} \Rightarrow p_{\mathbf{X}(t+\tau)}\left(\mathbf{X}(t+\tau)\right)$$

Note: In this last description, we used a discrete time version process model

$$\mathrm{X}(t+\tau) = \mathbf{F}\left(\mathrm{X}(t), \mathrm{u}(t), \boldsymbol{\xi}(t)\right)$$

Bayesian Estimation

$$\left\{\begin{array}{l} p_{\mathbf{X}(t)}\left(\mathbf{X}(t)\right) \;:\; \text{our current PDF about } \mathbf{X}(t) \\ p_{\boldsymbol{\xi}(t)}\left(\boldsymbol{\xi}(t)\right) \;:\; \text{PDF about noise } \boldsymbol{\xi}(t) \\ p_{\boldsymbol{\eta}(t)}\left(\boldsymbol{\eta}(t)\right) \;:\; \text{PDF about noise } \boldsymbol{\eta}(t) \\ \text{model } \mathrm{X}(t+\tau) = \mathbf{F}\left(\mathrm{X}(t), \mathrm{u}(t), \boldsymbol{\xi}(t)\right) \\ \text{model } \mathbf{y}(t+\tau) = \mathbf{h}\left(\mathrm{X}(t+\tau), \boldsymbol{\eta}(t+\tau)\right) \\ \text{measurements of } \mathbf{y}(t+\tau) \end{array}\right\}$$

$$\Downarrow$$

$$p_{\mathbf{X}(t+\tau)}\left(\mathbf{X}(t+\tau)\right) \;:\; \text{resulting PDF about } \mathbf{X}(t+\tau)$$

Output of the system?

It is a variable that is function of the state of the system, and which can also be measured by us.

$$\mathbf{y} = \mathbf{h}\left(\mathbf{X}\right)$$

Example: For the localization problem, in which we want to know the pose of a platform/robot; the distance of the robot to a certain known landmark (whose position is known to us), is an output of the system.

Why? Because that distance is function of the state!

Particular case, for estimating vehicle's pose.

$$\mathbf{X} = \begin{bmatrix} x \\ y \\ \phi \end{bmatrix}$$

$$\mathbf{r} = h(\mathbf{X}) = \sqrt{(x-a)^2 + (y-b)^2}$$

In which the variable **r** is an "output of the system"

(because it is a variable which is visible to us, and it is function of **X**)

(**a** and **b**, are simply constants, i.e. the position of the related landmark in certain CF (in which we express the estimated pose as well)

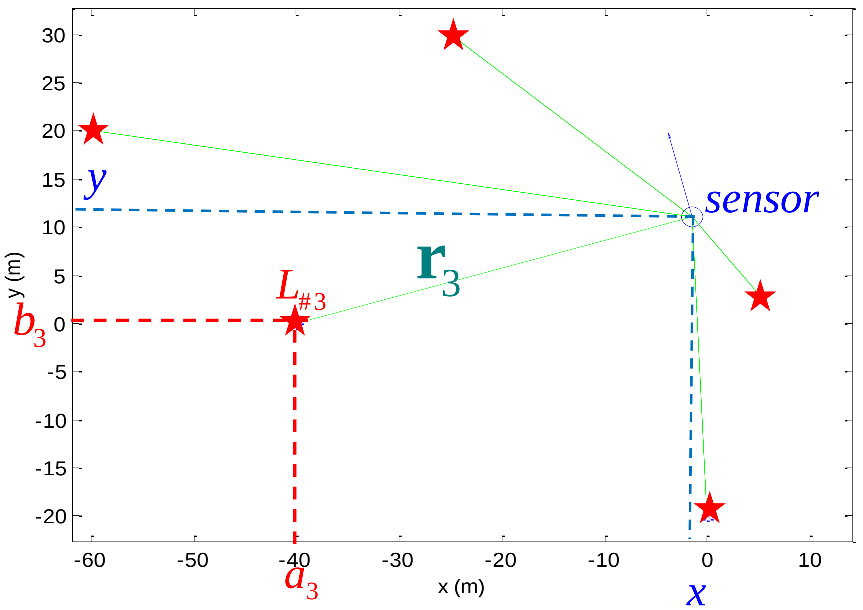
$$\mathbf{r}_1 = h\left(\mathbf{X}\right) = \sqrt{\left(x - a_1\right)^2 + \left(y - b_1\right)^2}$$

$$\mathbf{r}_2 = h\left(\mathbf{X}\right) = \sqrt{\left(x - a_2\right)^2 + \left(y - b_2\right)^2}$$

If we can measure more outputs, we can then improve our estimation of the state of the system.

If we can measure more outputs, we can then improve our estimation of the state of the system.

$$\left\{ \mathbf{r}_i = h(\mathbf{X}) = \sqrt{(x - a_i)^2 + (y - b_i)^2} \right\}_{i=1}^{N}$$



What happens if the equations are not perfectly defined?

$$\mathbf{y} = h(\mathbf{X}) + \boldsymbol{\eta}$$

$$\boldsymbol{\eta} \sim p_{\boldsymbol{\eta}}(\boldsymbol{\eta})$$

We cannot simply intersect constraints ! (because they are “lying a bit”)

We use, instead, likelihood functions

What is a likelihood function ? (we see some example , in a simulation)

Suppose we measure a distance to a landmark located at position (5,8), in the reference coordinate frame. We measure 10 metres. We know the sensor may introduce an error , which can be any value from -1 cm to 1cm. We assume a uniform PDF for that noise.

What points (x,y) would be solutions of that constraint?

Answer: any point inside the ring

$$9.99 \le \sqrt{(x-5)^2 + (y-8)^2} \le 10.01$$

So, we define a trivial "likelihood function", which is =1 at any point which belongs to the ring, and =0 at any point that does not belong to that region.

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

(comment: Likelihood functions do not usually have "plateau" shapes, because the uncertainties that pollute the measurement are not necessarily uniform.)

Why de we use the notation $p_{r|y,x}(r=10\,|\,x,y)$?

(conditional PDF of r given (x,y), evaluated at r=10m)

We need to see the Bayes' Rule to see that.

Bayes' RULE

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

How do we exploit it?

Suppose the variable X represents the state we want to estimate, and variable y a measurement of certain output of the system.

Both are RVs for us.

We assume there is a joint PDF which describes those variables.

(we know it does exist, but we do not need to know its expression)

We also have some prior idea about X (we have certain PDF about it)

$$p_{\mathbf{x}}^{(0)}\left(\mathbf{x}\right)$$

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

Now we want to combine that prior knowledge about X with the information provided by a measurement of y.

$p_{\mathbf{x},}^{(0)}(\mathbf{x})$: initial knowledge about **x**

$\mathbf{y} = \mathbf{y}_{measured}$ measurement of y (containing new information)

Suppose that we know the functional relationship between X and y

$$\mathbf{y} = \mathbf{h}(\mathbf{X})$$

and that a measurement of y is available. The measurement is polluted by noise, whose PDF is known by us,

$$\mathbf{y}_{measured} = \mathbf{h}(\mathbf{X}) + \boldsymbol{\eta}$$

$$\boldsymbol{\eta} \sim p_{\boldsymbol{\eta}}(\boldsymbol{\eta})$$

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

Based on that we define the conditional probability of the measurement of y given X, but in place of instancing X we instance y using its measured value.

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} = \mathbf{y}_{measured}\,|\,\mathbf{x})$$

This is a function about X, which is defined by { the measured value of y, the output equation and the PDF of the polluting noise } as follows:

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} = \mathbf{y}_{measured}\,|\,\mathbf{x}) = p_{\boldsymbol{\eta}}\left(\mathbf{y}_{measured} - \mathbf{h}(\mathbf{X})\right)$$

It is a function of X, and is called "**likelihood function of X**".

It is a likelihood function about X, associated to that particular measurement of y.

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

Then we can obtain what we actually want: the conditional PDF of X given that measurement of y:

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\,\mathbf{y} = \mathbf{y}_{measured}) = \frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} = \mathbf{y}_{measured}\,|\,\mathbf{x}) \cdot p_{\mathbf{x}}^{(0)}(\mathbf{x})}{p_{\mathbf{y}}(\mathbf{y}_{measured})}$$

In which:

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} = \mathbf{y}_{measured}\,|\,\mathbf{x}) = p_{\boldsymbol{\eta}}(\mathbf{y}_{measured} - \mathbf{h}(\mathbf{X}))$$

$p_{\mathbf{y}}(\mathbf{y}_{measured})$ is unknown, but it is a simple constant

$p_{\mathbf{x}}^{(0)}(\mathbf{x})$ is our prior PDF about $\mathbf{x}$

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\,\mathbf{y} = \mathbf{y}_{measured}) \propto p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} = \mathbf{y}_{measured}\,|\,\mathbf{x}) \cdot p_{\mathbf{x}}^{(0)}(\mathbf{x})$$

So we conclude :

$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\,|\,\mathbf{y} = \mathbf{y}_{measured})$ is called Posterior PDF

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

The missing factor is obtained by just normalizing

$$c=\int_{\text{domain of }\mathbf{X}} p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}=\mathbf{y}_{measured}\mid\mathbf{x}\right)\cdot p_{\mathbf{x}}^{(0)}\left(\mathbf{x}\right)\cdot d\mathbf{x}$$

$$p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x}\mid\mathbf{y}=\mathbf{y}_{measured}\right)=\frac{1}{c}\cdot p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}=\mathbf{y}_{measured}\mid\mathbf{x}\right)\cdot p_{\mathbf{x}}^{(0)}\left(\mathbf{x}\right)$$

$$btw: c=p_{\mathbf{y}}\left(\mathbf{y}_{measured}\right)$$

Can we process multiple measurements?

(so we can fuse multiple individual sources of information)

$$p_{x|y}(\mathbf{x}\,|\,\mathbf{y}) = \frac{p_{y|x}(\mathbf{y}\,|\,\mathbf{x}) \cdot p_x(\mathbf{x})}{p_y(\mathbf{y})}$$

→ YES.

If the instances of noises in those measurements are statistically independent we simply repeat the procedure, recursively.

If we have a set of measurements of variables which have functional relationship with the unknown variable of interest X

$\{\mathbf{y}_i\}_{i=1}^{M}$: set of measurements of output variables

$$p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x}\,|\,\mathbf{Y}=\{\mathbf{y}_1\}\right) \propto p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}=\mathbf{y}_1\,|\,\mathbf{x}\right) \cdot p_{\mathbf{x}}^{(0)}(\mathbf{x})$$

$$p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x}\,|\,\mathbf{Y}=\{\mathbf{y}_1,\mathbf{y}_2\}\right) \propto p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}=\mathbf{y}_2\,|\,\mathbf{x}\right) \cdot p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x}\,|\,\mathbf{Y}=\{\mathbf{y}_1\}\right)$$

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

in which

$p_{\mathbf{x}|\mathbf{Y}}\left(\mathbf{x}\,|\,\mathbf{Y}=\{\mathbf{y}_i\}_{i=1}^{K}\right)$: conditional PDF of X based on first k measurements of the set.

$p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}=\mathbf{y}_K\,|\,\mathbf{x}\right)$: likelihood function of X based on individual measurement #k

Note: the set of measurements can be measurements of diverse output variables, having those variables different output equations, and the polluting noises having different PDFs.

We are just multiplying all those available likelihood functions and the prior PDF, and finally normalizing the resulting function of X, which will result in a cumulative posterior PDF.

That resulting PDF does include all the information provided by the measurements, and by the initial PDF.

Example in our case of study (localization based on LiDAR)

Example in our case of study (localization based on LiDAR)

At time ***t1*** we have a prior PDF about **X(t1)**.

We process a LiDAR scan which is taken at time **t1**. We detect **n** landmarks. We have **n** range measurements, each of them having a different functional relationship with the variable **X(t1)**, and each of them independently affected by noise.

So, we can generate **n** likelihood functions about **X(t1)**, and obtain the posterior PDF about **X(t1)**

Usually, the more likelihood functions are available, the better estimates we obtain about X(t1).

We see a simulation/animation about such a “triangulation” process.

Note:

Example in our case of study (localization based on LiDAR)

This example is simplified to the case in which we estimate X for describing the 2D position (not including heading)

So, we can visualize its PDF via 2D surfaces or via images.

We see a simulation/animation about such a “triangulation” process.

First we see cases, having different number of map landmarks.

We test different deployments of landmarks.

We inspect

- how likelihood functions are, in 2D.
- How PDF evolves by applying likelihood functions.

In all those cases, we DO know the identity of the OOIs/Landmarks being used.

(we see it now)

Extra complication: cases in which we do not know the identity of the landmarks being detected.

So, multi hypotheses likelihood functions are generated and processed through the Bayes rule.

If enough information is available, the resulting PDF turns to be unimodal and narrow.

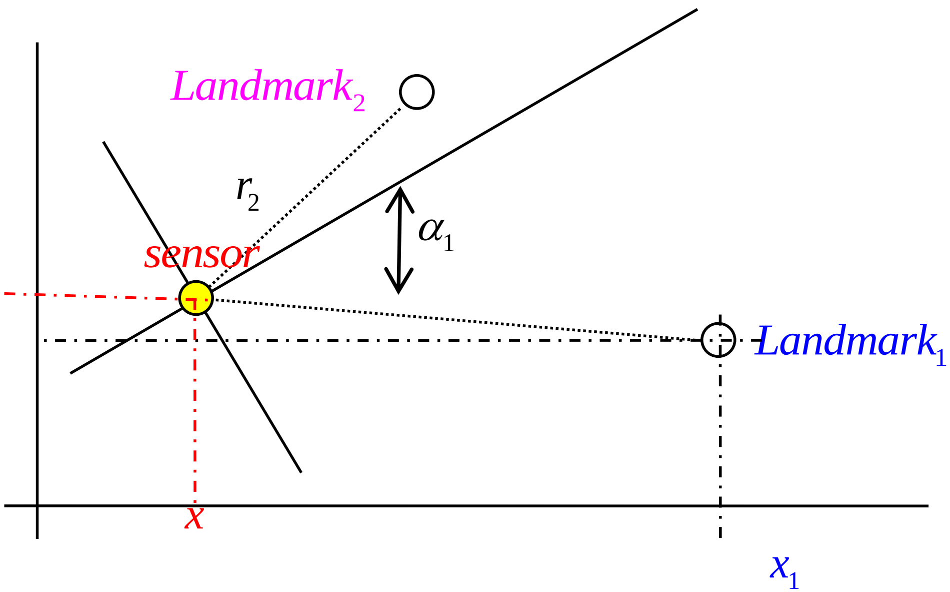
Let’s see those cases.

We end this part here.

We have seen how to process multiple sources of information for generating an estimate of an unknown, X, by exploiting the Bayes' Rule.

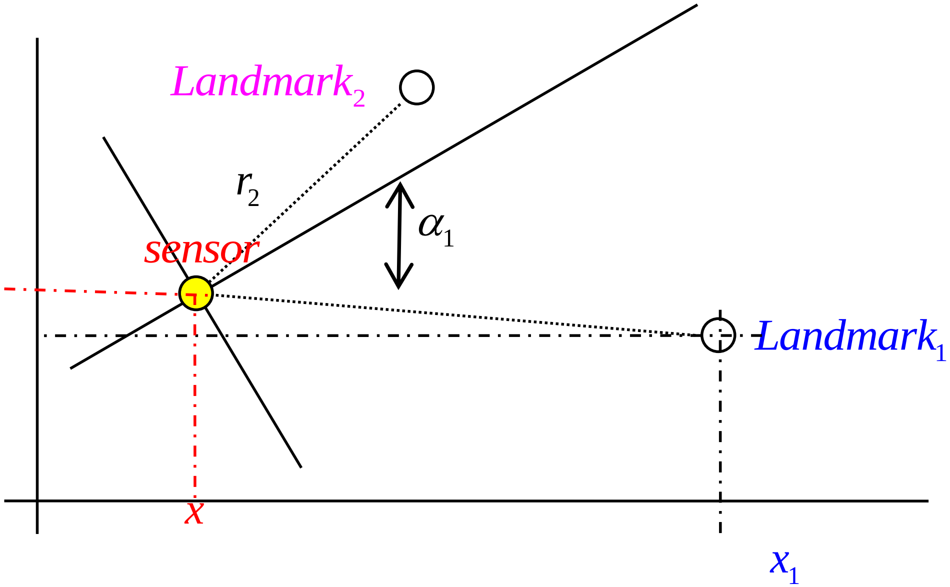
The estimate is produced in the form of a PDF about X.

Simple Localization based on Triangulation and Trilateration



LiDAR:	Range and Bearing
Camera:	Bearing
Sonar:	Range
Depth Camera:	Range and Bearing

2D Trilateration based on Range measurements.



Sonar: Range Only

$$\begin{Bmatrix} r_1 - \sqrt{(x-x_1)^2 + (y-y_1)^2} = 0 \\ r_2 - \sqrt{(x-x_2)^2 + (y-y_2)^2} = 0 \\ r_3 - \sqrt{(x-x_3)^2 + (y-y_3)^2} = 0 \end{Bmatrix}$$

It cannot solve heading.

2D Trilateration based on Range measurements.

$$\left\{ \begin{array}{l} r_1 - \sqrt{\left(x - x_1\right)^2 + \left(y - y_1\right)^2} = 0 \\ r_2 - \sqrt{\left(x - x_2\right)^2 + \left(y - y_2\right)^2} = 0 \\ r_3 - \sqrt{\left(x - x_3\right)^2 + \left(y - y_3\right)^2} = 0 \end{array} \right\}$$

$$\left\{ r_k - \sqrt{\left(x - x_k\right)^2 + \left(y - y_k\right)^2} = 0 \right\}_{k=1}^{N}$$

r_k : range to landmark #k (measured in local or in global CF)

$\left(x_k, y_k\right)$: position of landmark #k (expressed in global CF)

$\left(x, y\right)$: position of sensor (expressed in global CF) (unknown)

2D Triangulation based on Angular measurements.

$$\left\{\begin{array}{l}\alpha_1 = \operatorname{atan2}\left(y_1 - \mathrm{y},\ x_1 - x \right) - \phi \\ \alpha_2 = \operatorname{atan2}\left(y_2 - \mathrm{y}, x_2 - x \right) - \phi \\ \alpha_3 = \operatorname{atan2}\left(y_3 - \mathrm{y}, x_3 - x \right) - \phi \end{array}\right\}$$

$$\left\{\alpha_i = \operatorname{atan2}\left(y_i - y, x_i - x\right) - \phi\right\}_{i=1}^{N}$$

We can estimate full pose (if enough number of equations, i.e. 3)

The participant variables $\left\{(x_i, y_i)\right\}_{i=1}^{N}, \quad x, y, \phi$

Are expressed in the global coordinate frame.

2D Localization based on Range and Angular measurements.

$$\left\{ \begin{array}{l} r_i - \sqrt{\left(x_i - x\right)^2 + \left(y_i - y\right)^2} = 0 \\ \alpha_i = \operatorname{atan2}\left(y_i - y, x_i - x\right) - \phi \end{array} \right\}_{i=1}^{N}$$

Can estimate full pose (if enough number of equations)

$$\left\{ p_k^a = R(\phi) \cdot p_k^b + T \right\}_{k=1}^{N}$$

Alternative interpretation.

$$R(\phi) = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}$$

Can estimate full pose (if enough number of equations)

p_k^b : position of object in coordinate frame CF_b (e.g. OOI in local CF)

p_k^a : position of object in coordinate frame CF_a (e.g. Landkmark, associated to OOI, expressed in global CF)

$R(\phi)$: rotation matrix associated to heading ϕ of the sensor in (in CF_a, e.g., global CF)

T : translation = position of the sensor (in CF_a)

Note: Those mentioned equations will allow us to implement likelihood functions in our estimations processes (to be used next week), when we use EKF (approach based on Bayesian estimation), for solving localization.

The equations may also help you to solve part D of project 1.

Now we need to complete the picture of the estimation process.

Until know ,we have seen how to combine multiple likelihood functions , generated from measurements of the system outputs.

But we do not know how to fuse those with our process model.

From now, we will talk about discrete time process systems.

We use the discrete time index k, assuming some implicit “dt”

$$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right)$$

If we had a PDF about $\mathbf{x}(k)$

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

what would be $p_{\mathbf{x}(k+1)}\left(\mathbf{x}(k+1)\right)$

If we had a perfect model

$$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right)$$

?

It would be

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

$$\mathbf{x}(k+1)=F\left(\mathbf{x}(k),\mathbf{u}(k)\right)$$

$$\Downarrow$$

$$p_{\mathbf{x}(k+1)}\left(\mathbf{x}(k+1)\right)=\left(p_{\mathbf{x}(k)}(\mathbf{x})\cdot J(\mathbf{x},\mathbf{u})^{-1}\right)\Big|_{\mathbf{x}=F^{-1}\left(\mathbf{x}(k+1),\mathbf{u}(k)\right)}$$

$$J(\mathbf{x},\mathbf{u})=\left|\frac{\partial F}{\partial \mathbf{x}}(\mathbf{x},\mathbf{u})\right|$$

Do not panic! We do not perform that calculation in that way.

But we must simply know that YES, we can obtain that PDF.

We need a new PDF because , now, we are at time k+1, and we need to know about $\mathbf{x}(k+1)$; we do not have interest on $\mathbf{x}(k)$ anymore.

$$p_{\mathbf{x}(k+1)}\left(\mathbf{x}(k+1)\right)$$

But we still need an extra step, because our process model is not perfect,

$$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right) + \boldsymbol{\xi}(k)$$

(we assume here an additive error)

But we need an extra step, because our model is not perfect,

$$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right) + \boldsymbol{\xi}(k)$$

In which $\boldsymbol{\xi}(k)$ behaves as white noise (independent of the rest of RVs)

So, it is a case of adding two independent RVs $\mathbf{z} = \mathbf{w} + \boldsymbol{\xi}$

being $$\mathbf{w} = F\left(\mathbf{x}(k), \mathbf{u}(k)\right), \quad \boldsymbol{\xi} = \boldsymbol{\xi}(k)$$

We know how to obtain the PDF of z (convolution!) because we have the PDFs of $\mathbf{w}$ and $\boldsymbol{\xi}$ from previous intermediate steps

So that we are able to generate the PDF about X(k+1) considering the imperfect model,

$$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right) + \boldsymbol{\xi}(k)$$

Because we have

$$\left\{\begin{array}{l} p_{\mathbf{x}(k)}\left(\mathbf{x}(k)\right) \\ p_{\xi}\left(\boldsymbol{\xi}(k)\right) \\ \mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right) + \boldsymbol{\xi}(k) \\ \mathbf{u}(k) \end{array}\right\} \Rightarrow p_{\mathbf{x}(k+1)}\left(\mathbf{x}(k+1)\right)$$

mathematically possible !

We call this step: PREDICTION

What happens if a time k+1 we have measurements of certain system's outputs?

$$\mathbf{x}(k+1) = F\left(\mathbf{x}(k), \mathbf{u}(k)\right) + \boldsymbol{\xi}(k)$$

$$\mathbf{y}(k+1) = h\left(\mathbf{x}(k+1)\right) + \boldsymbol{\eta}(k+1)$$

We should use their likelihood functions for improving our PDF about **x**(k+1) (as we saw before, applying Bayes' rule)

We call this step "UPDATE"

(the update is based on measurements of system outputs. We refer to those as "observations"

So, we always perform predictions (to keep synchronized with the time, k)

and after the prediction, if there is any available likelihood function, we process it.

We express it as follows,

$$p_{\mathbf{x}(k)|\mathbf{Y}(k-1)}\left(\mathbf{x}(k) \mid \mathbf{Y}(k-1)\right) = \text{Prediction}\left(p_{\mathbf{x}(k-1)|\mathbf{Y}(k-1)}\left(\mathbf{x}(k-1) \mid \mathbf{Y}(k-1)\right)\right)$$

$$p_{\mathbf{x}(k)|\mathbf{Y}(k)}\left(\mathbf{x}(k) \mid \mathbf{Y}(k)\right) \quad \propto \quad p_{\mathbf{y}(k)|\mathbf{x}(k)}\left(\mathbf{y}(k) \mid \mathbf{x}(k)\right)\Big|_{\mathbf{y}(k)=\mathbf{y}_K} \cdot p_{\mathbf{x}(k)|\mathbf{Y}(k-1)}\left(\mathbf{x}(k) \mid \mathbf{Y}(k-1)\right)$$

The meaning of the notation is the following:

$\mathbf{Y}(i)$: all the observations, which have occurred till time ***i*** **(**including the observations at that time as well)

$\mathbf{y}(i)$: the observation (or the set of observations) that occurred at time ***i***,

$\mathbf{y}_i$: measured value of $\mathbf{y}(i)$.

$p_{\mathbf{x}(k-1)|\mathbf{Y}(k-1)}\left(\mathbf{x}(k-1)\,|\,\mathbf{Y}(k-1)\right)$: PDF describing the RV $\mathbf{x}(k-1)$, based on observations collected till time $k-1$.

$p_{\mathbf{x}(k)|\mathbf{Y}(k-1)}\left(\mathbf{x}(k)\,|\,\mathbf{Y}(k-1)\right)$: the PDF describing the RV $\mathbf{x}(k)$, based on observations collected till time $k-1$.

$p_{\mathbf{x}(k)|\mathbf{Y}(k)}\left(\mathbf{x}(k)\,|\,\mathbf{Y}(k)\right)$: the PDF for describing the RV $\mathbf{x}(k)$ based on observations collected till time k.

$p_{\mathbf{y}(k)|\mathbf{x}(k)}\left(\mathbf{y}(k)\,|\,\mathbf{x}(k)\right)\Big|_{\mathbf{y}(k)=\mathbf{y}_K}$: likelihood function based on the measurement $\mathbf{y}_k$, and the observation model of $\mathbf{y}(k)$.

We keep applying the sequence:

$$\xrightarrow[\textit{prediction}]{\substack{\mathbf{x}(k)= \\ F(\mathbf{x}(k-1),\mathbf{u}(k-1))}} p_{\mathbf{x}(k)|\mathbf{Y}(k-1)}\left(\mathbf{x}(k)\,|\,\mathbf{Y}(k-1)\right) \xrightarrow[\substack{\text{Apply} \\ \text{observations} \\ \text{performed at time k}}]{\text{update}} p_{\mathbf{x}(k)|\mathbf{Y}(k)}\left(\mathbf{x}(k)\,|\,\mathbf{Y}(k)\right)$$

$$\xrightarrow[\textit{prediction}]{\substack{\mathbf{x}(k+1)= \\ F(\mathbf{x}(k),\mathbf{u}(k))}} p_{\mathbf{x}(k+1)|\mathbf{Y}(k)}\left(\mathbf{x}(k+1)\,|\,\mathbf{Y}(k)\right) \xrightarrow[\substack{\text{Apply} \\ \text{observations} \\ \text{performed at time k+1}}]{\text{update}} p_{\mathbf{x}(k+1)|\mathbf{Y}(k+1)}\left(\mathbf{x}(k+1)\,|\,\mathbf{Y}(k+1)\right)$$

$$\xrightarrow[\textit{prediction}]{\substack{\mathbf{x}(k+2)= \\ F(\mathbf{x}(k+1),\mathbf{u}(k+1))}} p_{\mathbf{x}(k+2)|\mathbf{Y}(k+1)}\left(\mathbf{x}(k+2)\,|\,\mathbf{Y}(k+1)\right) \xrightarrow[\substack{\text{Apply} \\ \text{observations} \\ \text{performed at time k+2}}]{\text{update}} p_{\mathbf{x}(k+2)|\mathbf{Y}(k+2)}\left(\mathbf{x}(k+2)\,|\,\mathbf{Y}(k+2)\right)$$

This seems complicated.

But we have a card to play, to simplify matters.

→ **Gaussian PDFs**

→ In many cases (not in general) we can approximate the involved PDFs by using Gaussian PDFs.

→ Due to that, all the complicated operations turn to be easy to implement, and very cheap to process.

→ To be seen, next week.