Inference for dynamic systems: a toy example with Gaussian Linear Models

Suppose you are tracking a vehicle moving on a road, and coordinate \( x \) defines its position.

**Open-loop long term prediction**

The initial position of the vehicle is \( x_0 \). Time is discretized into intervals, and in the \( k \)-th interval the vehicle moves of \( \Delta_k \). So position at time \( t_k \), just after the end of the \( k \)-th interval, is \( x_k \), related to that at previous step to \( x_k = x_{k-1} + \Delta_k \).

\[
\begin{align*}
x_k &= x_0 + \sum_{i=1}^{k} \Delta_i \quad (1)
\end{align*}
\]

\( x_0 \) and all \( \Delta_s \) are random variables. From (1) we derive that the expected value (\( \mathbb{E} \)) of \( x_k \) is:

\[
\begin{align*}
\mu_{x,k} &= \mathbb{E}[x_k] = \mathbb{E}[x_0] + \sum_{i=1}^{k} \mathbb{E}[\Delta_i] = \mu_{x,0} + \sum_{i=1}^{k} \mu_{\Delta,i} \quad (2)
\end{align*}
\]

where \( \mu_{x,0} = \mathbb{E}[x_0] \), \( \mu_{\Delta,i} = \mathbb{E}[\Delta_i] \). Suppose these random variables are uncorrelated. Then the variance (\( \mathbb{V} \)) of \( x_k \) is:

\[
\begin{align*}
\sigma_{x,k}^2 &= \mathbb{V}[x_k] = \mathbb{V}[x_0] + \sum_{i=1}^{k} \mathbb{V}[\Delta_i] = \sigma_{x,0}^2 + \sum_{i=1}^{k} \sigma_{\Delta,i}^2 \quad (3)
\end{align*}
\]

To focus on a simple case, suppose the vehicle is initially in position zero for sure, the average step-size is zero for all intervals, and variance of the step-size is constant at value \( \sigma_\Delta^2 \) (i.e. \( \mu_{x,0} = 0, \sigma_{x,0}^2 = 0, \forall k \mu_{\Delta,k} = 0, \sigma_{\Delta,k}^2 = \sigma_\Delta^2 \)). Then we conclude that \( \mu_{x,k} = 0, \sigma_{x,k}^2 = k \sigma_\Delta^2 \): the expected value is the initial position (by symmetry) and uncertainty (variance) grows linearly with the number of intervals.

Let us assign a specific model: suppose that each variable is independently normally distributed. Mean and variance are as before, so \( p(x_k) = \mathcal{N}(x_k, \mu_{x,k}, \sigma_{x,k}^2) \), where \( \mathcal{N}(x, \mu, \sigma^2) \) is the normal distribution with mean \( \mu \), variance \( \sigma^2 \), computed at \( x \). For the case presented before, \( p(x_k) = \mathcal{N}(x_k, 0, k \sigma_\Delta^2) \), and therefore it is easy to compute the probability of the vehicle being over any assigned threshold \( L \) at time \( t_k \), \( \mathbb{P}[x_k > L] \), by using the corresponding normal CDF. On the other hand, note that it is not easy to get the probability that the random walk overpass level \( L \) at least once in any time between \( 0 \) and \( k \).

**Open-loop one-step ahead iterative prediction**

Note also the following: system dynamics is encoded in probabilities: \( p(x_0) \) and \( p(x_{k+1} | x_k) \). Latter one is the probability of going from \( x_k \) to \( x_{k+1} \) in one step, and can derived from \( p(\Delta_k): p(x_{k+1} | x_k) = \)
\( p(\Delta_k = x_{k+1} - x_k) \). Under the assumption of normality, this is: 
\[ p(x_{k+1} | x_k) = \mathcal{N}(x_{k+1}, x_k + \mu_{\Delta,k+1}, \sigma^2_{\Delta,k+1}). \]

At any step, suppose \( p(x_{k} | I_k) \) describes your belief on where the vehicle at time \( t_k \), given observations and background information, indicated with \( I_k \). Probability for position at time \( t_{k+1} \) is:

\[
p(x_{k+1} | I_k) = \int_{x_k} p(x_{k+1}, x_k | I_k) dx_k = \int_{x_k} p(x_{k+1} | x_k) p(x_k | I_k) dx_k
\]

(4)

Under normality, suppose \( p(x_{k} | I_k) = \mathcal{N}(x_k, \mu_{k|I_k}, \sigma^2_{k|I_k}) \), and \( x_k \) is independent of \( \Delta_k \). Then previous equation can be solved by summing normal random variables, to get
\[
p(x_{k+1} | I_k) = \mathcal{N}(x_{k+1}, \mu_{k+1|I_k}, \sigma^2_{k+1|I_k})
\]

with \( \mu_{k+1|I_k} = \mu_k + \mu_{\Delta,k} \) and \( \sigma^2_{k+1|I_k} = \sigma^2_{k|I_k} + \sigma^2_{\Delta,k} \).

Inference after observation

Suppose after the \( k + 1 \)-th interval the position of the vehicle is measured by \( y_{k+1} = x_{k+1} + \eta_{k+1} \), where noise \( \eta_{k+1} \) is independently distributed by \( p(\eta_{k+1}) \). Corresponding likelihood can be expressed as \( p(y_{k+1} | x_{k+1}) = p(\eta_{k+1} = y_{k+1} - x_{k+1}) \). Again, given \( p(x_{k+1} | I_k) \) defined as before, inference after getting measure \( y_{k+1} \) is computed by Bayes’ formula:

\[
p(x_{k+1} | y_{k+1}, I_k) \propto p(y_{k+1} | x_{k+1}) p(x_{k+1} | I_k)
\]

(5)

Suppose \( p(\eta_{k+1}) = \mathcal{N}(\eta_{k+1}, 0, \sigma^2_{\eta,k+1}) \), then
\[
p(y_{k+1} | x_{k+1}) = \mathcal{N}(y_{k+1}, x_{k+1}, \sigma^2_{\eta,k+1}) = \mathcal{N}(x_{k+1}, y_{k+1}, \sigma^2_{\eta,k+1}).
\]

Under normality of the prior, Eq.5 reads \( p(x_{k+1} | y_{k+1}, I_k) = \mathcal{N}(x_{k+1}, \mu_{k+1|I_k}, \sigma^2_{k+1|I_k}) \), with posterior parameters:

\[
\begin{align*}
\mu_{k+1|I_k} &= \left( \sigma^{-2}_{\eta,k+1} + \sigma^{-2}_{\eta,k+1} \right)^{-1} \\
\sigma^2_{k+1|I_k} &= \left( \sigma^{-2}_{\eta,k+1} \right)^{-1} + \left( \sigma^{-2}_{\eta,k+1} \right)^{-1} \sigma^{-2}_{\eta,k+1} y_{k+1}
\end{align*}
\]

(6)

where \( I_{k+1} = \{ l_k, y_{k+1} \} \). If another observation \( z_{k+1} \) is collected at the same time, defined by \( z_{k+1} = x_{k+1} + \varepsilon_{k+1} \) with \( \varepsilon_{k+1} \) independently distributed, one can apply Bayes formula again:
\[
p(x_{k}|y_k, z_{k}, l_k) \propto p(x_{k} | y_{k}, l_k) p(z_{k} | x_{k}).
\]

In that case, where \( I_{k+1} = \{ l_k, y_{k+1}, z_{k+1} \} \). Under normality, Eq.6 can be applied again. On the other hand, if no observation is collected, no inference is needed.

Iterative filtering

Suppose we track the vehicle processing data of position sensors. We can alternate prediction and inference. Consider the case of a single sensor, producing measurements \( \{ y_1, y_2, \ldots \} \) where \( y_k \) is collected after \( k \)-th interval. Let us define \( I_k = \{ y_1, y_2, \ldots, y_k \} \). We can combine prediction and inference to compute \( p(x_{k+1} | I_{k+1}) \) from \( p(x_{k} | I_k) \), processing measure \( y_k \). To do so, we can follow this scheme:
\[
p(x_k | I_k) \xrightarrow{\text{prediction}} p(x_{k+1} | I_k) \xrightarrow{\text{inference}} p(x_{k+1} | I_{k+1})
\]

where obviously \( I_{k+1} = \{ y_{k+1}, l_k \} \). Under normality, we can track vehicle position by parameters \( \mu_{k|I_k} \) and \( \sigma^2_{k|I_k} \).
Example of application

Suppose we assume normality. Specifically, system dynamic is defined by:

\[ x_0 \sim \mathcal{N}(\mu_{x,0}, \sigma_{x,0}^2) = \mathcal{N}(0,1) \]
\[ \forall k \Delta_k \sim \mathcal{N}(\mu_\Delta, \sigma_\Delta^2) = \mathcal{N}(-1,4^2) \]

Two sensors (a) and (b) are recording vehicle position:

\[
\begin{aligned}
Y_{a,k} &= x_k + \eta_{a,k} \\
Y_{b,k} &= x_k + \eta_{b,k}
\end{aligned}
\]

Sensor (a) is far more precise than (b), as defined by noise:

\[
\begin{aligned}
\forall k \eta_{a,k} &\sim \mathcal{N}(0, \sigma_{\eta,a}^2) = \mathcal{N}(0,1) \\
\forall k \eta_{b,k} &\sim \mathcal{N}(0, \sigma_{\eta,b}^2) = \mathcal{N}(0,12^2)
\end{aligned}
\]

Here is an example of simulation, showing vehicle actual position in black, sensor measures in magenta dots and red crosses, posterior mean of filtering in continuous blue, posterior 95% confidence interval in dashed blue. Note that observations are not always available. Measures from (a) are seldom available, from (b) more often, but no observation is available for many steps.
You can compare previous figure with following ones, referring to measures \((b)\) only, or none measures at all.
The toy example shows many features of dynamic systems, sensor fusion, missing data. At any time, you can predict future evolution. For example, at time 45, you can predict the position at time 100, but predicting 55 steps ahead: $p(x_{100} | l_{45})$. 
\( \mu_{x,0} = 0, \sigma_{x,0} = 1, \mu_\Delta = -1, \sigma_\Delta = 4, \sigma_{\eta} = 1, \sigma_{\eta b} = 12 \)