LECTURE 4

The Bayesian estimation framework can be outlined as follows:

x: hidden "model" variable (e.g., kinematics)

y: observed variable (e.g., firing rate of neurons)

$$p_x(x|y) = \frac{p_y(y|x)\,p_x(x)}{p_y(y)} \qquad (1)$$

Bayes' Rule

$p_x(x) \triangleq$ prior of x (i.e., independent of any observation)

$p_y(y) \triangleq$ marginal probability of y (i.e., probability of making observation y)

$p_x(x|y) \triangleq$ posterior distribution of x

$p_y(y|x) \triangleq$ likelihood of an observation given the underlying variable x

First, we assume that – at generic step i – the prior of x is normally distributed with an estimated mean and variance (or covariance, in case x is a vector):

$$X_i^- \sim N(\hat{X}_i^-, \hat{P}_i^-)$$

Second, we assume that observations and model variables are linearly related according to the formula:

$$Y_i = A X_i + \eta \qquad \eta \sim N(0, R)$$

Under these two assumptions, we showed that:

- $$Y_i \sim N(A\hat{X}_i^-, A\hat{P}_i^- A^T + R)$$

- The random variables x and y, at step i, are jointly Gaussian, i.e., we can write:

$$p(X_i^-, Y_i) = N\left(\begin{bmatrix} \hat{X}_i^- \\ A\hat{X}_i^- \end{bmatrix}, \begin{bmatrix} \hat{P}_i^- & \hat{P}_i^- A^T \\ A\hat{P}_i^- & A\hat{P}_i^- A^T + R \end{bmatrix}\right) \quad (2)$$

Now, because $p(x,y) = p_y(y|x)\, p_x(x)$ in (1), we want to decompose the joint probability function (2) such that the marginal probability $p(y)$ is in evidence and can cancel out the denominator in (1) $\Rightarrow$ We recall:

$$p(x,y) = N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

$\Downarrow$ with x, y marginally Gaussian

$$p(x,y) = N\left(\mu_x + \Sigma_{12}\Sigma_{22}^{-1}(y-\mu_y),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right) N(\mu_y, \Sigma_{22})$$

$$\Downarrow$$

$$p_x(x|y) = N\left(\mu_x + \Sigma_{12}\Sigma_{22}^{-1}(y-\mu_y),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right) \quad (3)$$

In our case, we have:

$\mu_x = \hat{X}_i^-$; $\mu_y = A\hat{X}_i^-$; $\Sigma_{11} = \hat{P}_i^-$; $\Sigma_{21} = A\hat{P}_i^-$; $\Sigma_{12} = \hat{P}_i^- A^T$

$\Sigma_{22} = A\hat{P}_i^- A^T + R$ — By replacing in (3) we have:

$$\underset{x}{E}(x|y) = \hat{X}_i^+ = \hat{X}_i^- + \hat{P}_i^- A^T \left(A\hat{P}_i^- A^T + R\right)^{-1}\left(Y_i - A\hat{X}_i^-\right)$$

$$\underset{x}{E}\left(\left(x - \underset{x}{E}(x|y)\right)^2 \middle| y\right) = \hat{P}_i^+ = \hat{P}_i^- - \hat{P}_i^- A^T \left(A\hat{P}_i^- A^T + R\right)^{-1} A\hat{P}_i^-$$

By using fundamental algebric relationships, we can show:

$$\hat{P}_i^- - \hat{P}_i^- A^T\left(A\hat{P}_i^- A^T + R\right)^{-1} A\hat{P}_i^- = \left[(\hat{P}_i^-)^{-1} + A^T R^{-1} A\right]^{-1}$$

$$\hat{P}_i^- A^T \left(A \hat{P}_i^- A^T + R \right)^{-1} = \hat{P}_i^+ A^T R^{-1}$$

thus showing that the correction part of the KF returns the mean and covariance of the posterior distribution of x given y

In this formulation, though, we require a prior X_i^- that is Gaussian
$\Rightarrow$ We need to determine the conditions under which this is possible:

- One option consists of assigning a generative model:

(4) $$x_i = F x_{i-1} + \varepsilon_{i-1} \qquad \varepsilon_{i-1} \sim N(0, Q) \qquad x_i \triangleq \text{random variable}$$

In this way, the problem becomes recursive, i.e.:

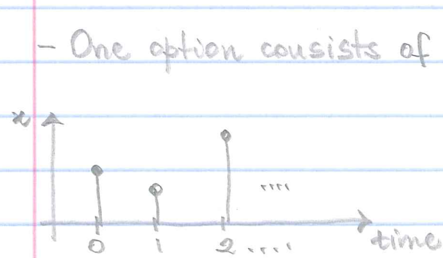
$$x_{i-1} \sim N(\hat{X}_{i-1}, \hat{P}_{i-1}) \Rightarrow x_i \sim N(F \hat{X}_{i-1}, F \hat{P}_{i-1} F^T + Q)$$

i.e, x_i is Gaussian if x_{i-1} is Gaussian $\Rightarrow$ This condition is suitable for the formulation of the predictive part of the KF, i.e.:

$$\hat{X}_{i-1}^+ \text{ - Gaussian} \Rightarrow \hat{X}_i^- \sim N(F \hat{X}_{i-1}^+, F \hat{P}_{i-1}^+ F^T + Q)$$

and the condition can be propagated as long as the initial estimate $\hat{X}_0^+$ is Gaussian (typically: $\hat{X}_0^+ \sim N(0, Q)$)

- One option consists of assigning a model for the pdf. Note this:



$$p_x(x_0, x_1, x_2, \ldots, x_i) =$$

$$p_x(x_i | x_{i-1}, x_{i-2}, \ldots, x_1, x_0) \cdot$$
$$p_x(x_{i-1} | x_{i-2}, \ldots, x_1, x_0) \cdot \ldots \cdot p_x(x_1 | x_0) \, p_x(x_0)$$

and

$$p_x(x_i) = \int dx_{i-1} \int dx_{i-2} \cdots \int p_x(x_i, x_{i-1}, \ldots, x_1, x_0)\, dx_0$$ (definition of marginal pdf)

We can relax the assumption of having a generative model like (4) but preserve the two fundamental properties of that model, i.e.:

A1) $p_x(x_i \mid x_{i-1}, x_{i-2}, \ldots, x_1, x_0) = p_x(x_i \mid x_{i-1})$

A2) $p_y(y_i \mid x_i, y_{i-1}) = p_y(y_i \mid x_i)$

In this way, we can still write:

$$p_x(x_i) = \int p_x(x_i \mid x_{i-1})\, p_x(x_{i-1})\, dx_{i-1}$$

$$\cong \int p_x(x_i \mid x_{i-1})\, p_x(x_{i-1} \mid y_{i-1})\, dx_{i-1}$$

where: $p_x(x_{i-1} \mid y_{i-1})$ is the posterior pdf at step $i-1$
$p_x(x_i)$ is the prior of x at step i
$p_x(x_i \mid x_{i-1})$ must be estimated

and we can still obtain a recursive formula:

$$p_x(x_i \mid y_i) = \frac{1}{p_y(y_i)}\, p_y(y_i \mid x_i) \int p_x(x_i \mid x_{i-1})\, p_x(x_{i-1} \mid y_{i-1})\, dx_{i-1} \qquad (*)$$

to determine the posterior distribution of $x_i \Rightarrow$ [The issue now is how to implement this in a way that is compatible with real-time est.]

Furthermore, the use of a Bayesian framework allows to determine a method to estimate the model parameters. In fact, let us suppose that:

$y_i = A x_i + \eta_i \qquad \eta_i \sim N(0, R)$ with A and R to be determined

$x_i = F x_{i-1} + \varepsilon_{i-1} \qquad \varepsilon_i \sim N(0, Q)$ with F and Q to be determined

Let us assume that training data (X_i, Y_i) is available for $i=1,2,\dots,M$
In this case, we can estimate (A, F, Q, R) such that the joint probability over the training data is maximized, i.e.:

$$(A^*, F^*, Q^*, R^*) = \arg\max \; p_x(X_1, X_2, \dots, X_M, Y_1, Y_2, \dots, Y_M)$$

$$= \arg\max \prod_{i=2}^{M} p_x(X_i \mid X_{i-1}) \, p_x(X_1) \prod_{i=1}^{M} p_y(Y_i \mid X_i)$$

It can be shown that the solution (under the normality assumptions) is:

$$A^* = \left(\sum_{i=1}^{M} Y_i X_i^T\right)\left(\sum_{i=1}^{M} X_i X_i^T\right)^{-1}$$

$$F^* = \left(\sum_{i=2}^{M} X_i X_{i-1}^T\right)\left(\sum_{i=2}^{M} X_{i-1} X_{i-1}^T\right)^{-1}$$

$$R^* = \frac{1}{M}\left(\sum_{i=1}^{M} Y_i Y_i^T - A^* \sum_{i=1}^{M} X_i Y_i^T\right)$$

$$Q^* = \frac{1}{M-1}\left(\sum_{i=2}^{M} X_i X_i^T - F^* \sum_{i=2}^{M} X_{i-1} X_i^T\right)$$

$\square$

Ex.: Let us consider the formula:

$$p_x(x_i \mid y_i) = \frac{1}{p_y(y_i)} p_y(y_i \mid x_i) \int p_x(x_i \mid x_{i-1}) \, p_x(x_{i-1} \mid y_{i-1}) \, dx_{i-1}$$

for $i=1$ and let us assume: $p_x(x_0) = N(\hat{X}_0, \hat{P}_0)$ (there is no observation at time 0); $p_x(x_1 \mid x_0) = N(F\hat{X}_0, F\hat{P}_0F^T + Q)$

In this case the integral can be computed, given that the function to be integrated is the product of two normal pdf $\Rightarrow$ Let us call:

$$G(x_1) = \int_{-\infty}^{+\infty} p(x_1, \mu_1, \sigma_1)\, p(x_0, \mu_0, \sigma_0)\, dx_0$$

where: $p(x, \mu, \sigma) = \dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ and:

$\mu_1 = F\hat{X}_0$; $\sigma_1 = F\hat{P}_0 F^T + Q$; $\mu_0 = \hat{X}_0$; $\sigma_0 = \hat{P}_0$

Let us now assume that, differently from the past, we have a log-normal pdf for the measurements:

$$p_y(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}}$$ and

$$p_y(y|x) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\ln y - \mu_0 - \mu_1 x)^2}{2\sigma^2}}$$

i.e., x affects the distribution by changing the parameter μ

In this way, we can compute:

$$p_x(x_1|y_1) = \frac{p_y(y_1|x_1)}{p_y(y_1)} G(x_1) = G(x_1) \exp\left(\frac{(\ln y_1 - \mu)^2}{2\sigma^2} - \frac{(\ln y_1 - \mu_0 - \mu_1 x_1)^2}{2\sigma^2}\right)$$

From this formula, we can estimate the most likely value $\hat{X}_1$ as the solution of:

$$\frac{dp_x}{dx_1}(x_1|y_1) = 0$$

i.e., we can determine an estimation that (i) is optimal in a precise mathematical sense, (ii) is obtained for non-Gaussian distributions of the measurements y, and (iii) can be estimated with statistical accuracy i.e, confidence bounds can be estimated. □

In summary, we have seen two approaches to the decoding of kinematic variables from neural recordings:

- A least-square approach: firing rates and kinematic variables are related via a regressive (linear) model. Model parameters are estimated offline by using LS and kinematic variables are obtained by inverting the model

The first case we encountered (i.e., the population vector) can be considered a special case

The result is a linear filter that processes the current and past firing rate values

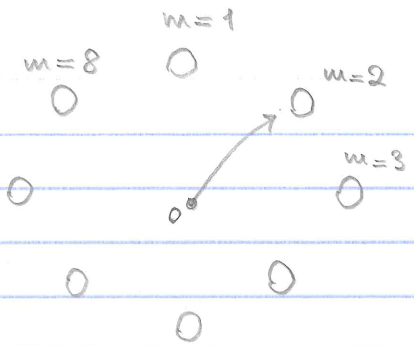
- A Bayesian approach: we study the conditional pdf of the kinematic variables given the neural measurements and and we determine the estimate as the value that maximizes the posterior pdf

The approach boils down to KF if all the pdf are Gaussian and linear (regressive) models describe the evolution of the kinematic variables and the relationship with the neural data

The approach is more general, as different classes of pdf can be explored, several tools are available to solve the problems, and solutions come with conf. bounds

Furthermore, the Bayesian framework allows a modular expansion of the problems that can be solved. For instance:

8



In the classic reach-out experiment, let us assume that the movement has different regimes (e.g., reach curvature and speed) depending on the goal, i.e.:

$$p_x(x_i \mid x_{i-1}, m) = N(A_m x_{i-1}, Q_m)$$

now depending on the goal

$p_y(y_i \mid x_i)$ - known

In this case, we can follow the Bayesian paradigm and write:

$$p_x(x_i \mid y_i) = \sum_{m=1}^{8} p_x(x_i \mid y_i, m)\, p_m(m \mid y_i)$$

Now: $p_x(x_i \mid y_i, m)$ is the posterior of x_i for the specific goal $m \Rightarrow$ We can apply formula $(*)$ for each goal

$p_m(m \mid y_i)$ is the probability that the actual regime is $m \Rightarrow$ It can be computed by using the Bayes' rule:

$$p_m(m \mid y_i) = \frac{p_y(y_i \mid m)\, p_m(m)}{p_y(y_i)} \quad \text{where:}$$

$p_y(y_i) \rightarrow$ It is the same as in $(*)$

$p_y(y_i \mid m) \rightarrow$ It is the posterior of y_i given the movement regime

$p_m(m) \rightarrow$ It is the prior belief about the regime m, i.e., it allows to incorporate prior knowledge about the identity of the regime to improve the estimation of the probabilities □

One of the bottlenecks of the Bayesian approach (when it does not boil down to the KF) is the necessary computational power:

- How do we estimate a kinematic variable? $\rightarrow$ We must update the parameters of pdf by using lenghty maximization procedures

- How do we solve the integral in (*)? $\rightarrow$ We must compute samples of the distributions and sum over them

- How do we determine the priors? $\rightarrow$ We need particle filters and MC simulations to determine these pdf offline

These issues suggest to keep looking into the KF approach $\Rightarrow$ Two alternatives can be explored:

- UKF: Performance enhancement is sought by using a nonlinear regression model and the unscented Kalman filter:

original model (linear) $\longleftrightarrow$ alternative model (linear)

$$Y_i = A X_i + \eta_i \qquad\qquad X_i = Y_i^T A + \eta_i$$

Y_i: $n \times 1$ firing rates (one per neuron); X_i: $p \times 1$ (kinematic variables)

X_i: $p \times 1$; Y_i^T: $(n \cdot m) \times 1$ firing rates (m per neuron, spanning m windows)

These models are now replaced by a quadratic model:

$y_i = b^T h(X_i) + \eta_i$ where $b = [b_1 \; b_2 \; \ldots \; b_{6n}]$ is a vector of coefficients and $h(X_i)$ is the quadratic tuning:

y_i: firing rate of a single neuron

$$\underbrace{X_i \triangleq \begin{bmatrix} pos_x(i) \\ pos_y(i) \\ vel_x(i) \\ vel_y(i) \end{bmatrix}}_{4\times 1} \Rightarrow h(X_i) = \left.\begin{bmatrix} pos_x(i) \\ pos_y(i) \\ vel_x(i) \\ vel_y(i) \\ \sqrt{pos_x^2 + pos_y^2} \\ \sqrt{vel_x^2 + vel_y^2} \end{bmatrix}\right\} 6\times 1$$

And the model can be expanded by adding previous hystory:

$$y_i = \begin{bmatrix} B_0^T & B_1^T & \dots & B_{m-1}^T \end{bmatrix} \begin{bmatrix} h(X_i) \\ h(X_{i-1}) \\ \vdots \\ h(X_{i-m+1}) \end{bmatrix} + \eta_i$$

NOTE: $m > 0$ here but anticipation is possible and does not change the matricial notation

By repeating the same procedure for every neuron we can write:

(a) $Y_i = \bar{B}\, W_i + \underline{\eta}_i$ where: $Y_i \triangleq n\times 1$ vector of firing rates

$$W_i \triangleq \begin{bmatrix} h(X_i) \\ h(X_{i-1}) \\ \vdots \\ h(X_{i-m+1}) \end{bmatrix}$$ $6m \times 1$ vector of kinematic features

$\bar{B} \triangleq n \times (6m)$ matrix of coefficients

Similarly, one can replace the evolution model of the state variables:

original model (1st order)		new model (m-th order)
$X_i = F X_{i-1} + \underline{\varepsilon}_{i-1}$	$\rightarrow$	$X_i = F_1 X_{i-1} + F_2 X_{i-2} + \dots + F_m X_{i-m} + \underline{\varepsilon}_{i-1}$

The new model can be rearranged as:

$$\underbrace{\begin{bmatrix} X_i \\ X_{i-1} \\ \vdots \\ X_{i-m+1} \end{bmatrix}}_{T_i} = \underbrace{\begin{bmatrix} F_1 & F_2 & \dots & F_m \\ I_4 & & \dots & \\ & I_4 & \ddots & \\ \dots & \underline{0}_4 & I_4 & \underline{0}_4 \end{bmatrix}}_{\bar{F}} \underbrace{\begin{bmatrix} X_{i-1} \\ X_{i-2} \\ \vdots \\ X_{i-m} \end{bmatrix}}_{T_{i-1}} + \underline{\varepsilon}_{i-1}$$

and it becomes: $T_i = \bar{F} T_{i-1} + \underline{\varepsilon}_{i-1}$ (b)

With the combination of the extended models (a) and (b), we can now implement the UKF:

1) Prediction Step

$$\hat{T}_i^- = \bar{F}\hat{T}_{i-1}^+$$

$$\hat{P}_i^- = \bar{F}\hat{P}_{i-1}^+\bar{F}^T + Q$$

with $Q \triangleq$ covariance matrix of the noise ε_i (it is $4m \times 4m$)

2) Sampling Step

$$\sigma_0 = \hat{T}_i^- \; ; \; \sigma_j = \hat{T}_i^- + \sqrt{(4m+1)}\,[\hat{P}_i^-]_j \; ; \; \sigma_{4m+j} = \hat{T}_i^- - \sqrt{(4m+1)}\,[\hat{P}_i^-]_j$$

for $j = 1, 2, \dots, 4m$ with $[\hat{P}_i^-]_j \triangleq$ j-th column of $\hat{P}_i^-$

$z_j = \bar{B}\hat{W}_j$ $\quad j = 0, 1, \dots, 8m$, where $\hat{W}_j = h(\sigma_j)$

It's the $6m \times 1$ vector of kinematic features evaluated at sigma point σ_j

3) Correction Step

$$\tilde{z}_i = \sum_{j=0}^{8m} w_j z_j$$

$w_j \triangleq$ weight to be assigned (e.g., $1/_{8m+1}$)

$$P_{zz} = \sum_{j=0}^{8m} w_j (z_j - \tilde{z}_i)(z_j - \tilde{z}_i)^T + R$$

$R \triangleq$ covariance matrix of noise η_i

$$P_{Tz} = \sum_{j=0}^{8m} (\sigma_j - \hat{T}_i^-)(z_j - \tilde{z}_i)^T$$

$$\hat{P}_i^+ = \hat{P}_i^- - P_{Tz} P_{zz}^{-1} P_{Tz}^T$$

$$\hat{T}_i^+ = \hat{T}_i^- + P_{Tz} P_{zz}^{-1} (y_i - \tilde{z}_i)$$

12

- Subject-in-the-loop: Performance enhancement is sought by keeping the (linear) KF but modifying the approach to parameter estimation and propagation of the uncertainty
(a.k.a. ReFIT-KF)

First, a two-tiered filtting procedure is implemented offline to adjust the model parameters to the intention

Step 1: offline estimation of F, A:

$$X_i = F X_{i-1} + \varepsilon_{i-1}$$

$$Y_i = A X_i + \eta_i$$

Step 2: online update of F, A based on the intended direction of movement

Second, as the KF runs online, the uncertainty on position (i.e., rows of $\hat{P}_i^+$ that refer to position) is set to 0 (i.e., the sight of the neuro-controlled cursor on the screen removes uncertainty)

In this way, intention of the subject and visual feedback are encompassed in the filter $\Rightarrow$ Performance may significantly enhance. See slides.

$\blacksquare$