Lecture 6: The Kalman filter

STAT 545: Introduction to computational statistics

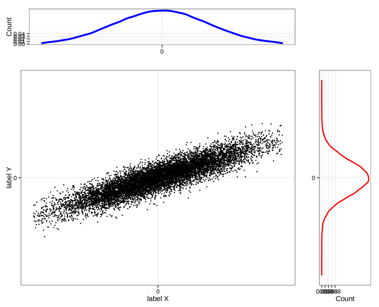
Vinayak Rao

Department of Statistics, Purdue University

September 6, 2018

Marginalization:

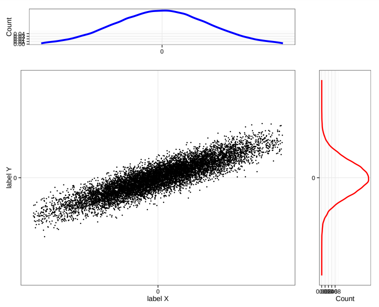
$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right), \quad X \sim ?$$



$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 10 & 4 \\ 4 & 2 \end{bmatrix}$$

Marginalization:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}\right), \quad X \sim ?$$

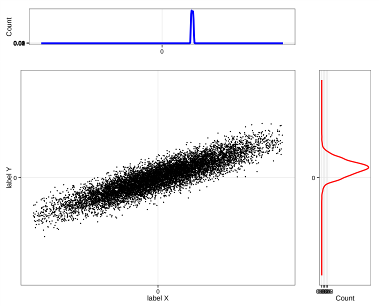


$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 10 & 4 \\ 4 & 2 \end{bmatrix}$$

$$X \sim \mathcal{N}(\mu_X, \Sigma_{XX})$$

Conditioning:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right), \quad Y|(X = a) \sim ?$$

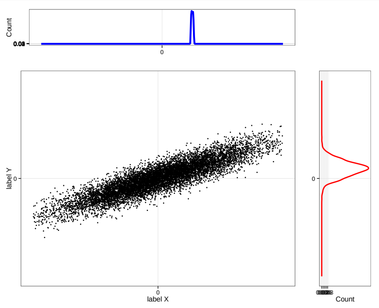


$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 10 & 4 \\ 4 & 2 \end{bmatrix}$$

Some properties of the Gaussian

Conditioning:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right), \quad Y|(X = a) \sim ?$$



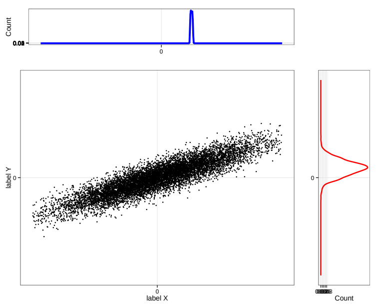
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 10 & 4 \\ 4 & 2 \end{bmatrix}$$

$$Y|X \sim \mathcal{N}\left(\mu_Y + \Sigma_{XY}\Sigma_{XX}^{-1}(a - \mu_X)\ ,\ \Sigma_{YY} - \Sigma_{XY}\Sigma_{XX}^{-1}\Sigma_{YX}\right)$$

Some properties of the Gaussian

Conditioning:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right), \quad Y|(X = a) \sim ?$$



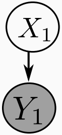
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 10 & 4 \\ 4 & 2 \end{bmatrix}$$

$$Y|X \sim \mathcal{N}\left(\mu_Y + \Sigma_{XY}\Sigma_{XX}^{-1}(a - \mu_X)\,,\; \Sigma_{YY} - \Sigma_{XY}\Sigma_{XX}^{-1}\Sigma_{YX} \right)$$

Cost?

The Gaussian distribution, conjugacy and Bayes' rule

$$X_1 \sim \mathcal{N}(\mu, \Sigma)$$

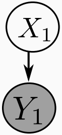


$$\epsilon \sim \mathcal{N}(0, \Sigma_E)$$

$$Y_1 = AX_1 + \epsilon$$

The Gaussian distribution, conjugacy and Bayes' rule

$$X_1 \sim \mathcal{N}(\mu, \Sigma)$$



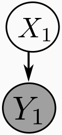
$$\epsilon \sim \mathcal{N}(0, \Sigma_E)$$

$$Y_1 = AX_1 + \epsilon$$

We have a Gaussian ‘prior’ on X_1.

We observe a noisy measurement $Y_1 | X_1 \sim \mathcal{N}(AX_1, \Sigma_E)$.

The Gaussian distribution, conjugacy and Bayes' rule



$$X_1 \sim \mathcal{N}(\mu, \Sigma)$$

$$\epsilon \sim \mathcal{N}(0, \Sigma_E)$$

$$Y_1 = AX_1 + \epsilon$$

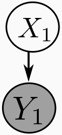
We have a Gaussian 'prior' on X_1.

We observe a noisy measurement $Y_1 | X_1 \sim \mathcal{N}(AX_1, \Sigma_E)$.

$$\begin{bmatrix} X \\ \epsilon \end{bmatrix} \to \begin{bmatrix} X \\ Y \end{bmatrix}$$

X and Y jointly Gaussian: what is its mean and covariance?

The Gaussian distribution, conjugacy and Bayes' rule



$$X_1 \sim \mathcal{N}(\mu, \Sigma)$$

$$\epsilon \sim \mathcal{N}(0, \Sigma_E)$$

$$Y_1 = AX_1 + \epsilon$$

We have a Gaussian 'prior' on X_1.

We observe a noisy measurement $Y_1|X_1 \sim \mathcal{N}(AX_1, \Sigma_E)$.

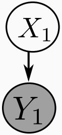
$$\begin{bmatrix} X \\ \epsilon \end{bmatrix} \to \begin{bmatrix} X \\ Y \end{bmatrix}$$

X and Y jointly Gaussian: what is its mean and covariance?

Y is marginally Gaussian: what is its mean and covariance?

The Gaussian distribution, conjugacy and Bayes' rule

$$X_1 \sim \mathcal{N}(\mu, \Sigma)$$



$$\epsilon \sim \mathcal{N}(0, \Sigma_E)$$

$$Y_1 = AX_1 + \epsilon$$

We have a Gaussian 'prior' on X_1.
We observe a noisy measurement $Y_1 | X_1 \sim \mathcal{N}(AX_1, \Sigma_E)$.

$$\begin{bmatrix} X \\ \epsilon \end{bmatrix} \to \begin{bmatrix} X \\ Y \end{bmatrix}$$

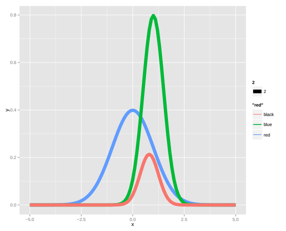
X and Y jointly Gaussian: what is its mean and covariance?

Y is marginally Gaussian: what is its mean and covariance?

$X|Y$ is Gaussian: what is its mean and covariance?

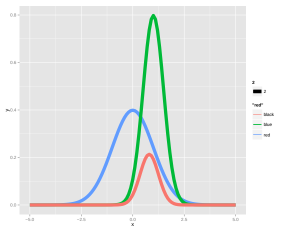
Product of Gaussian densities:

Product of Gaussian densities is Gaussian a Gaussian density (upto a multiplication constant)



Intuition: sum of two quadratic functions is a quadratic

Product of Gaussian densities is Gaussian a Gaussian density (upto a multiplication constant)



Intuition: sum of two quadratic functions is a quadratic

Aside: need only specify prob. distrib. to a constant

- $p(x)$ and $C \cdot p(x)$ represents the same, if C is independent of x
- Probabilities must integrate to 1

Markov chains

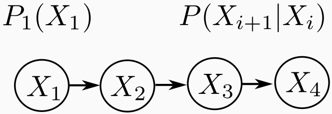
$$P_1(X_1) \qquad\qquad P_i(X_{i+1}|X_i)$$

$$X_1 \to X_2 \to X_3 \to X_4$$

A sequence of random variables such that

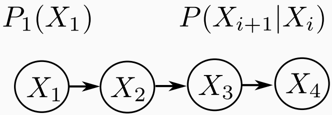
$$P(X_{i+1}|X_i, X_{i-1}, \cdots, X_1) = P(X_{i+1}|X_i)$$

We'll stick to homogeneous chains:



Markov chains

We'll stick to homogeneous chains:



In fact, with $X_i \in \Re^D$, we will consider:

$$X_1 \sim \mathcal{N}(\mu_0, \Sigma_0)$$

$$X_{i+1} = AX_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \Sigma_E)$$

We'll stick to homogeneous chains:

$$P_1(X_1) \qquad P(X_{i+1}|X_i)$$

$$X_1 \to X_2 \to X_3 \to X_4$$

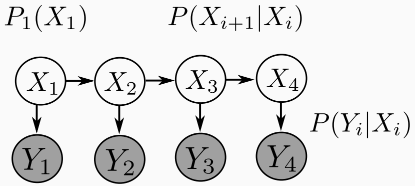
In fact, with $X_i \in \Re^D$, we will consider:

$$X_1 \sim \mathcal{N}(\mu_0, \Sigma_0)$$

$$X_{i+1} = AX_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \Sigma_E)$$

If our chain has T steps, a TD-dimensional Gaussian!
In the figure, $T = 4$. In practice: thousands to millions.

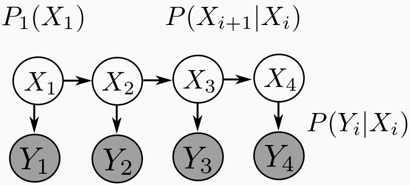
We don't observe the chain directly:



$$Y_i = BX_i + \zeta_i, \quad \zeta \sim \mathcal{N}(0, \Sigma_z), \quad Y_i \in \Re^d$$

A hidden Markov model

We don't observe the chain directly:

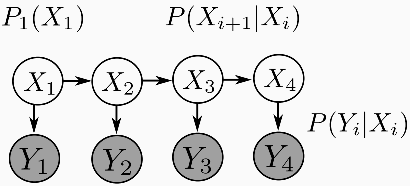


$$Y_i = BX_i + \zeta_i, \quad \zeta \sim \mathcal{N}(0, \Sigma_z), \quad Y_i \in \Re^d$$

We want to answer questions like: What is $p(X_i|Y_1, \cdots, Y_T)$?
$\{X_i, Y_i\}$ is a $(D+d)T$-dimensional Gaussian.

A hidden Markov model

We don't observe the chain directly:



$$Y_i = BX_i + \zeta_i, \quad \zeta \sim \mathcal{N}(0, \Sigma_z), \quad Y_i \in \Re^d$$

We want to answer questions like: What is $p(X_i|Y_1, \cdots, Y_T)$?
$\{X_i, Y_i\}$ is a $(D+d)T$-dimensional Gaussian.
We 'just' have to look at conditionals?

[board]