EE 126 — Probability and Random Processes

April 17, 2018 — Joint Gaussians

1 Lecture

Joint Gaussian Random Variables

From the MMSE/LLSE picture, it is easy to see that MMSE $\neq$ LLSE in general. When are they equal? *When X and Y are jointly Gaussian.*

(Accordingly, joint Gaussian random variables are nice because the MMSE is hard to compute in many cases – you have to integrate over the whole joint distribution – whereas computing the LLSE is far easier.)

Theorem. *If X and Y are jointly Gaussian, then $L[X \mid Y] = \mathbb{E}[X \mid Y]$.*

Definition. If (X_1, X_2) are **jointly Gaussian**, then $\bar{X} = (X_1, X_2)$ has a **multivariate Gaussian distribution** where $\bar{X}$ is a **random vector**.

Alternatively, $f_{X_1,X_2}(x_1, x_2)$ is such that

$$\alpha_1 x_1 + \alpha_2 x_2 \sim \mathcal{N}(\cdot, \cdot) \quad \forall \quad \alpha_1, \alpha_2 \in \mathbb{R}$$

(For *every* weighted combination of x_1 and x_2, we have a normal distribution.)

In this case, if X_1 and X_2 are **uncorrelated** ($cov(X_1, X_2) = 0$), then they are also independent. (In general, "independent" $\implies$ "uncorrelated," but "uncorrelated" $\not\Longrightarrow$ "independent," so this is powerful.)

Multivariate Gaussian (Generalization of Univariate)

$$f_{X_1,\dots,X_k}(x_1, \dots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(X - \mu_X)^T \Sigma^{-1} (X - \mu_X)\right)$$

where $\underset{k\times 1}{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$, $\underset{k\times 1}{\mu_X} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix} = \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_k] \end{bmatrix}$, and

$$\begin{aligned}
\underset{k\times k}{\Sigma} &= \text{covariance matrix} \\
&= \begin{bmatrix} cov(x_i, x_j) \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \cdots \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \ddots \\ \vdots & \ddots & \sigma_k^2 \end{bmatrix} \\
&= \begin{bmatrix} \rho\sigma_i\sigma_j \end{bmatrix} \quad \text{(symmetric)}
\end{aligned}$$

Recall: $cov(x, y) = \rho\sigma_x\sigma_y$, where ρ is the correlation coefficient.

In the case where $k = 7$ and $\mu_X = 0, \mu_Y = 0$,

$$f_{X,Y} = \frac{1}{\sqrt{(2\pi)^2 \begin{vmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{vmatrix}}} \exp\left(-\frac{1}{2} \begin{bmatrix} x & y \end{bmatrix} \Sigma^{-1} \begin{bmatrix} x \\ y \end{bmatrix}\right)$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x\sigma_y}\right)\right)$$

If $\rho = 0$,

$$\begin{aligned} f_{X,Y} &= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}\sigma_y} e^{-y^2/2} \\ &= f_X(x) \cdot f_Y(y) \\ &\implies X, Y \text{ independent} \end{aligned}$$

(This generalizes for $k > 2$.)

Thus if X and Y are uncorrelated, they are independent. *Intuition: if we have pairwise uncorrelation, Σ becomes a diagonal matrix and we have no cross terms.*

Theorem 6.4. *Jointly Gaussian random variables are independent iff they are uncorrelated.*

Theorem 6.5. *Linear combinations of jointly Gaussian random variables are jointly Gaussian.*

Example. If X, Y are i.i.d. $\mathcal{N}(0, 1)$, are $(X + Y)$ and $(X - Y)$ independent? (Note that any collection of i.i.d. Gaussians are jointly Gaussian ($\rho = 0$).

We only need to check correlation because $(X + Y)$ and $(X - Y)$ are jointly Gaussian by Theorem 6.5.

$$\begin{aligned} cov((X+Y),(X-Y)) &= \mathbb{E}[(X+Y)(X-Y)] - \mathbb{E}[(X+Y)]\mathbb{E}[(X-Y)] \\ &= \mathbb{E}[X^2 - Y^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[Y^2] \\ &= \sigma_x^2 - \sigma_y^2 \\ &= 0 \end{aligned}$$

Therefore $(X + Y)$ and $(X - Y)$ are uncorrelated and independent.

Example. Let $X \sim \mathcal{N}(0, 1)$, $\mathbb{E}[X] = 0$, and $W = \begin{cases} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}$, where W and X are independent. Let $Y = W \cdot X$. What is the distribution of Y?

Intuition: first we get a standard normal $\mathcal{N}(0, 1)$ from X. Then we either multiply by $+1$ or -1, but the standard normal is symmetric so this doesn't change the distribution.

$$cov(X, Y) = \mathbb{E}[XY] = \mathbb{E}[WX^2] = \mathbb{E}[W]\mathbb{E}[X^2] = 0$$

Hence X and Y are uncorrelated but not independent; they are *not* jointly Gaussian (even though they are Gaussian independently). Note, for example, that $X + Y \not\sim \mathcal{N}$. $X + Y$ is either $2X$ or 0, and half of the mass is at 0 so it is not Gaussian.

Kalman Filter

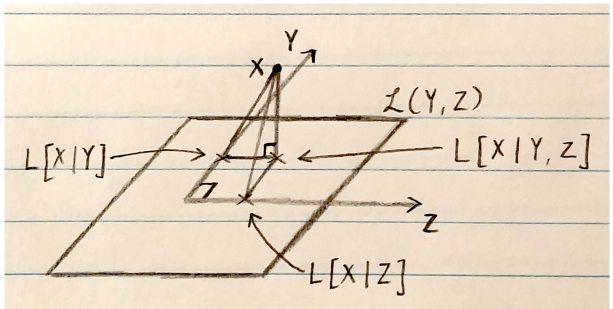
The **Kalman filter** is an algorithm for estimating a sequence of states of a process, given some noisy observations related to the states. The process will describe a sequence of random variables evolving over time. For example, we might want to track how a robot moves over time, and the states could be the positions of the robot.

Note: these are called *state-space dynamics*; the sequence $X_1, ..., X_n$ represents the state of the system. However, we don't get access to the states. We only get access to some noisy observations (themselves random variables) that are related to the states, and we want to *infer* the states. Furthermore, we would like to track states in an online fashion (in real time).

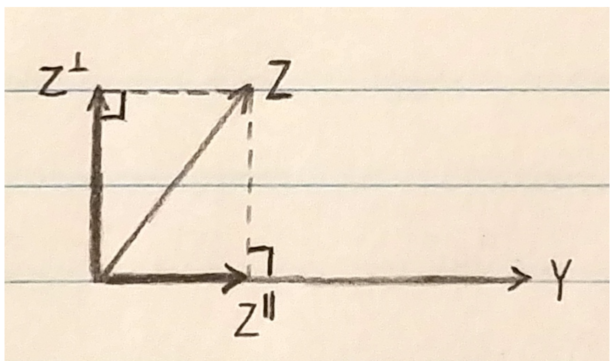
The good news is that we will stick to linear estimation!

`tl;dr:` Given some noisy observations that are related to a sequence of states, we want to infer the sequence of states. (Note that both the observations and states are random variables, and we can't observe the states.) We want to do this *online*.

Theorem. $L[X \mid Y, Z] = L[X \mid Y] + L[X \mid Z]$ *if* Y *and* Z *have nothing to do with each other.* In other words, Y and Z must be orthogonal. If $Y \perp Z$, then we establish orthogonal "axes" such that $L[X \mid Y, Z] = L[X \mid Y] + L[X \mid Z]$.



Suppose Y and Z are not orthogonal (i.e. correlated). We know the projection of X onto Y and the projection of X onto Z, but Y and Z are no longer at a 90 degree angle to each other. What do we do? *Make* Y and Z orthogonal; take only the orthogonal component of Z!



If Y and Z are not uncorrelated, then $L[X \mid Y, Z] = L[X \mid Y] + L[X \mid Z^\perp]$. The parallel component of Z brings no new information, so we only want the portion of Z that is perpendicular (i.e. the part we haven't seen before).

$$\underbrace{Z^\perp = Z - Z^\parallel = Z - \text{Proj}_Y Z = Z - L[Z \mid Y]}_{\text{``innovation,'' ``new stuff''}}$$

The Kalman filter is essentially Gram-Schmidt. It is an iterative way of doing Gram-Schmidt on the fly.