Second dose of probability

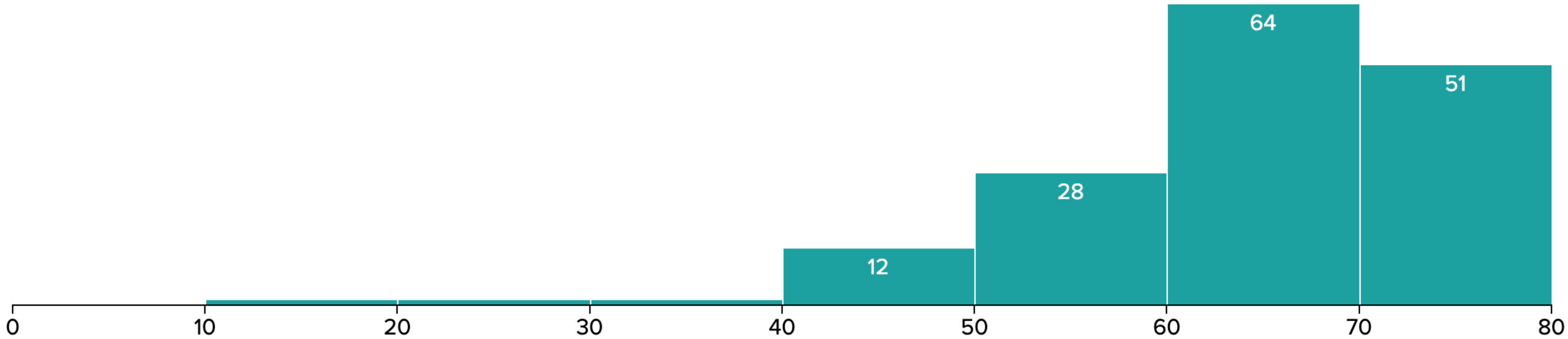
ROB 501

Necmiye Ozay

- **Slightly in-depth look at probability**
  - **Kalman filter peek**
- **(if time) Gaussian random variables**
  - **MVE another look (interpretation based on conditional probability using Schur complements)**

Announcements

- Exams are graded.
- Statistics:



MINIMUM	MEDIAN	MAXIMUM	MEAN	STD DEV
19.0	67.0	80.0	64.3	10.13

2a and 5d were the “hardest” (least # of correct answers) questions.

Announcements

- Extra credit for the exam. There will be a problem in the next problem set that allows you:
  - To pick up to 4 problems among the T/F part (first 5*4 questions) that you *missed* in the exam and submit the solution of them **including reasons**. Each solution with correct reasons will add +1 points to your exam score.
    - If you have less than 4 mistakes in the exam, mention that in the solutions too and we will take it into account

Announcements

- Wednesday lecture will be rescheduled for Thu or Fri (modulo availability of a room) due to additional travel…

Not Ex: $\mathcal{F} = \{\{2\}, \{1\}\}$

Let $\Omega = \{1,2\}$

Not Ex: Let $\mathcal{F} = \{\emptyset, \{1\}, \{1,2\}\}$ → not closed wrt. set complement

$\{1\}^c = \Omega \setminus \{1\} = \{2\} \notin \mathcal{F}$

Probability: A Second Dose in ROB 501

1 Probability Spaces

Ex: $\mathcal{F} = \{\emptyset, \underbrace{\{1,2\}}_{\Omega}\}$ is a sigma algebra

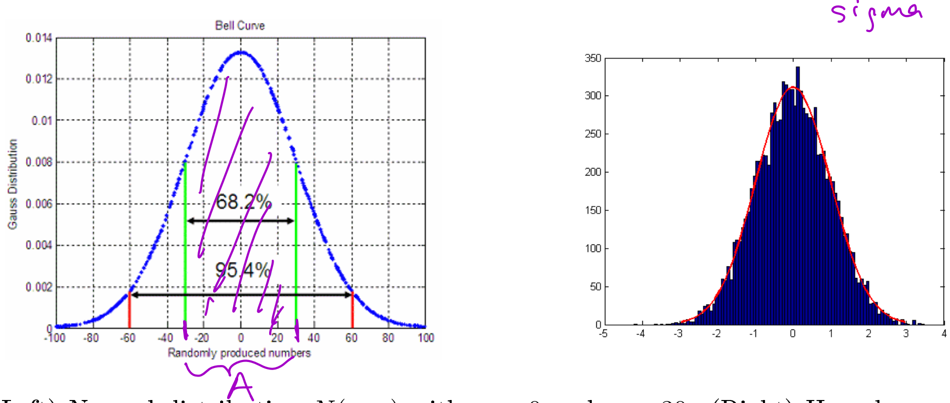


Figure 1: (Left) Normal distribution $N(\mu, \sigma)$ with $\mu = 0$ and $\sigma = 30$. (Right) How do you determine the density? You have to collect data! The figure shows a "fit" of a normal distribution to data.

Def. $(\Omega, \mathscr{F}, P)$ is called a probability space.

- Ω is the sample space. Think of it as the domain of a random variable $X : \Omega \to \mathbb{R}$ or random vector $X : \Omega \to \mathbb{R}^m$.

  $\mathcal{F} \subset 2^{\Omega}$ = power set of Ω ≈ the set of all subsets of Ω

- $A \subset \Omega$ is an event.

  → sigma algebra

- $\mathscr{F}$ is the collection of allowed events[1]. It must at least contain $\emptyset$ and Ω. It is closed w.r.t. countable unions and intersections, and set complement.

- $P : \mathscr{F} \to [0,1]$ is a probability measure. It has to satisfy a few basic operations

  $P(\Omega) = P(c \in \Omega)$

  - $P(\emptyset) = 0$ and $P(\Omega) = 1$.
  - For each $A \in \mathscr{F}$, $0 \leq P(A) \leq 1$
  - If the sets $A_1, A_2, \ldots$ are disjoint (i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

for a Gaussian random variable $\Omega = \mathbb{R}$

- Example event $A := \{\omega \in \Omega \mid |\omega - \mu| \leq \sigma\} \Rightarrow P(A) = 0.682$

[1]Though it is too deep for ROB 501, there are subsets of the reals, for example, that are so complicated one cannot define a reasonable notion of probability that agrees with how we would want to define the probability of an interval, such as $[a, b]$.

$\{\omega \in \Omega \mid X(\omega) \leq x\} \subset \Omega$

$\in \mathscr{F}$

2 Random Variables

X is a measurable function from Ω to $\mathbb{R}$

Def. $X : \Omega \to \mathbb{R}$ is a random variable if $\forall\ x \in \mathbb{R}$, the set $\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathscr{F}$. This just means that such sets can be assigned probabilities.

Remarks:

- Shorthand notation $\{X \leq x\} := \{\omega \in \Omega \mid X(\omega) \leq x\}$
- Because $\mathscr{F}$ is closed under set complements, (countable) unions, and (countable) intersections, we can also assign probabilities to

set complement

a) $\{X > x\} = \sim \{X \leq x\} = \{X \leq x\}^C$
b) $\{x < X \leq y\} = \{X \leq y\} \bigcap \{X > x\}$

Def. $X : \Omega \to \mathbb{R}$ is a continuous random variable if there exists a density $f : \mathbb{R}^p \to [0, \infty)$ such that,

$$\forall\ x \in \mathbb{R}, \quad P\big(\{X \leq x\}\big) = \int_{-\infty}^{x} f(\bar{x})d\bar{x} \quad (\bar{x} \text{ dummy variable in the integral})$$

Remarks:

- $\int_a^b f(x)dx = P\big(a < X \leq b\big) = P\big(a \leq X \leq b\big)$
  $= P\big(\{\omega \in \Omega \mid X(\omega) \in [a, b]\}\big)$
- mean: $\mu := \mathcal{E}\{X\} := \int_{-\infty}^{\infty} xf(x)dx$
- Variance: $\sigma^2 := \mathcal{E}\{(X - \mu)^2)\} := \int_{-\infty}^{\infty}(x - \mu)^2 f(x)dx$
- Standard Deviation $\sigma := \sqrt{\sigma^2}$ (Std. Dev.)

3 Random Vectors

Def. Let $(\Omega, \mathscr{F}, P)$ be a probability space. A function $X : \Omega \to \mathbb{R}^p$ is called a <u>random vector</u> if each component of $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$ is a random variable, that is, $\forall\ 1 \leq i \leq p$, $X_i : \Omega \to \mathbb{R}$ is a random variable.

Consequence: $\forall x \in \mathbb{R}^p$, the set $\{\omega \in \Omega \mid X(\omega) \leq \overset{\in \mathbb{R}^p}{x}\} \in \mathscr{F}$ (i.e., it is an allowed event), where the inequality is understood pointwise, that is,

$$\{\omega \in \Omega \mid X(\omega) \leq x\} := \{\omega \in \Omega \mid \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_p(\omega) \end{bmatrix} \leq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}\} = \bigcap_{i=1}^{p} \{\omega \in \Omega \mid X_i(\omega) \leq x_i\}$$

shorthand $\{X \leq x\}$

Def. $X : \Omega \to \mathbb{R}^p$ is a <u>continuous random vector</u> if there exists a <u>density</u> $f : \mathbb{R}^p \to [0, \infty)$ such that,

$$\forall\ x \in \mathbb{R}^P, \quad P\big(\{X \leq x\}\big) = \int_{-\infty}^{x_p} \dots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_X(\bar{x}_1, \bar{x}_2 \dots \bar{x}_p) d\bar{x}_1 d\bar{x}_2 \dots d\bar{x}_p$$

Ex: Multivariate Gaussian vector is a continuous random vector

4 Moments

expectations of $g(X)$ where g is a monomial

Def. Suppose $g : \mathbb{R}^p \to \mathbb{R}$

$$E\{g(X)\} = \int_{\mathbb{R}^p} g(x) f_X(x) dx = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, ..., x_p) f_X(x_1, ..., x_p) dx_1 ... dx_p$$

Mean or Expected Value 1^{st} order moment

$$\mu = E\{X\} = E\{\begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix}\} = \begin{bmatrix} \mathcal{E}\{X_1\} \\ \vdots \\ \mathcal{E}\{X_p\} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}$$

$E\{XX^T\}$

Covariance Matrices 2^{nd} order moment (sometimes this is called central moment)

$$\Sigma := \text{cov}(X) = \text{cov}(X, X) = E\{(X - \mu)(X - \mu)^T\}$$

where

$$(X - \mu) \text{ is } p \times 1, \ (X - \mu)^T \text{ is } 1 \times p, \ (X - \mu)(X - \mu)^T \text{ is } p \times p$$

Exercise cov(X) is positive semidefinite

take $v \in \mathbb{R}^p$

$$v^T \Sigma v = v^T E\{(X-\mu)(X-\mu)^T\} v$$
$$= E\{v^T (X-\mu)(X-\mu)^T v\}$$
$$= E\{[(X-\mu)^T v]^T [(X-\mu)^T v]\}$$
$$= E\{\underbrace{\|(X-\mu)^T v\|_2^2}_{g}\}$$

$g: \mathbb{R}^p \to \mathbb{R}$

$g(X) \doteq \|(X-\mu)^T v\|_2^2$

Σ is psd.

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \underbrace{\|(x-\mu)^T v\|_2^2}_{\geq 0} \underbrace{f_X(x_1, x_2, \dots x_p)}_{\geq 0} dx_1 \dots dx_p \geq 0$$

5 Marginal Densities, Independence, and Correlatation

Suppose the random vector $X : \Omega \to \mathbb{R}^p$ is partitioned into two components $X_1 : \Omega \to R^n$ and $X_2 : \Omega \to R^m$, with $p = n + m$, so that,

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

Notation: We denote the density of X by

$$f_X(x) = f_{\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}}(x_1, x_2) = f_{X_1 X_2}(x_1, x_2)$$

and it is called the joint density of X_1 and X_2. As before, we can define the mean and covariance.

- Mean is $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mathcal{E}\{X\} = E\{\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\} = \begin{bmatrix} \mathcal{E}\{X_1\} \\ \mathcal{E}\{X_2\} \end{bmatrix}$
- Covariance is

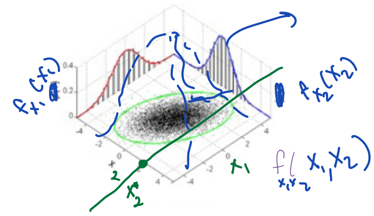
$$\begin{aligned}
\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} &= \mathcal{E}\{\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix}^\top\} \\
&= \mathcal{E}\{\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix} [(X_1 - \mu_1)^\top \quad (X_2 - \mu_2)^\top]\} \\
&= \mathcal{E}\{\begin{bmatrix} (X_1 - \mu_1)(X_1 - \mu_1)^\top & (X_1 - \mu_1)(X_2 - \mu_2)^\top \\ (X_2 - \mu_2)(X_1 - \mu_1)^\top & (X_2 - \mu_2)(X_2 - \mu_2)^\top \end{bmatrix}
\end{aligned}$$

where $\Sigma_{12} = \Sigma_{21}^\top = cov(X_1, X_2) = \mathcal{E}\{(X_1 - \mu_1)(X_2 - \mu_2)^\top\}$ is also called the correlation of X_1 and X_2.

Def: X_1 and X_2 are uncorrelated if $cov(X_1, X_2) =$

$f_{x_1}(x_1) = \int_{-\infty}^{\infty} f_{x_1 x_2}(x_1, x_2)\, d x_2$

if $x_1 \in \mathbb{R}$
$x_2 \in \mathbb{R}$



If $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} : \Omega \to \mathbb{R}^{n+p}$ is a continuous random vector, then its components

$$X_1 : \Omega \to \mathbb{R}^n \quad \text{and} \quad X_2 : \Omega \to \mathbb{R}^m$$

are also continuous random vectors and have densities, $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$. These densities are given a special name.

Def. $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ are called the marginal densities of X_1 and X_2.

Fact: In general the marginal densities are a nightmare to compute.

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{X_1 X_2}(x_1, x_2) dx_2 \\ &:= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 X_2}\big(\underbrace{\bar{x}_1, \ldots, \bar{x}_n}_{x_1}, \underbrace{\bar{x}_{n+1}, \cdots, \bar{x}_{n+m}}_{x_2}\big) \underbrace{d\bar{x}_{n+1} \cdots d\bar{x}_{n+m}}_{dx_2} \end{aligned}$$

$$\begin{aligned} f_{X_2}(x_2) &= \int_{-\infty}^{\infty} f_{X_1 X_2}(x_1, x_2) dx_1 \\ &:= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 X_2}\big(\underbrace{\bar{x}_1, \ldots, \bar{x}_n}_{x_1}, \underbrace{\bar{x}_{n+1}, \cdots, \bar{x}_{n+m}}_{x_2}\big) \underbrace{d\bar{x}_1 \cdots d\bar{x}_n}_{dx_1} \end{aligned}$$

For Normal Random Vectors, however, we can read them directly from the joint density! We will not be doing any iterated integrals.

Def. X_1 and X_2 are independent random vectors if their joint density factors

$$f_X(x) = f_{X_1 X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2).$$

Def. X_1 and X_2 are uncorrelated if their "cross covariance" or "correlation " is zero

$$cov(X_1, X_2) := \mathcal{E}\{(X_1 - \mu_1)(X_2 - \mu_2)^\top\} = 0_{n \times m}$$

$\underbrace{\int_{-\infty}^{\infty} x_1 f_{x_1}(x_1)\, dx_1}_{\mu_1} - \mu_1 \underbrace{\int_{-\infty}^{\infty} f_{x_1}(x_1) dx_1}_{1}$
$\mu_1 - \mu_1 \cdot 1 = 0$

Proof of fact for the scalar case: (where x_1, x_2 are scalar random variables)

$E\{(x_1-\mu_1)(x_2-\mu_2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1-\mu_1)(x_2-\mu_2) \underbrace{f_{x_1 x_2}(x_1,x_2)}_{\text{by ind.} \approx f_{x_1}(x_1) f_{x_2}(x_2)} dx_1\, dx_2 = \underbrace{\int_{-\infty}^{\infty} (x_1-\mu_1) f_{x_1}(x_1) dx_1}_{0} \underbrace{\int_{-\infty}^{\infty} (x_2-\mu_2) f_{x_2}(x_2) dx_2}_{0}$

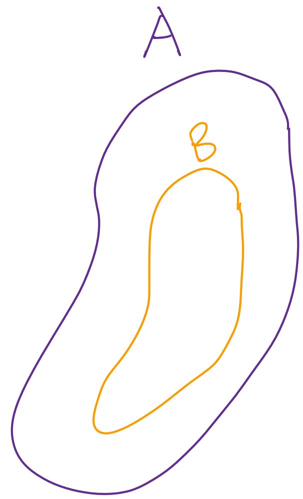
$= 0$

independent $\Longrightarrow$ uncorrelated
$\not\Longleftarrow$

Fact: If X_1 and X_2 are independent, then they are also uncorrelated. **The converse is in general false.**

* For Gaussian random vectors, there is an exception in that they uncorrelated iff they are independent.

$P(A|B) = 1$



6 Conditioning

Def. For two events $A, B \in \mathscr{F}$, $P(B) > 0$

$$P(A \mid B) := \frac{P(A \bigcap B)}{P(B)}$$

is the conditional probability of A given B.

Remarks:

- Suppose $P(A)$ is our current estimate of the probability that our robot is near a certain location and B is a measurement of the robot's location, with confidence in the measurement being $P(B)$. The conditional probability of A given B occurred is how we "fuse" the two pieces of information

$$P(A \mid B) := \frac{P(A \bigcap B)}{P(B)}$$

- $$B \subset A,\ P(A \mid B) = \frac{P(A \bigcap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

- $$A \subset B,\ P(A \mid B) = \frac{P(A \bigcap B)}{P(B)} = \frac{P(A)}{P(B)} \geq P(A)$$

$\in (0, 1]$

Consider again our partitioned random vector $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$

Def. The conditional density of X_1 given $X_2 = x_2$ is

$$f_{X_1|X_2}(x_1 \mid x_2) = \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)}.$$

Sometimes we simply write $f(x_1 \mid x_2)$

Remarks on Conditional Random Vectors:

- **Very important:** X_1 given $X_2 = x_2$ is (still) a random vector. It's density is $f_{X_1|X_2}(x_1 \mid x_2)$
- **Conditional Mean:**

$$\mu_{X_1|X_2=x_2} := \mathcal{E}\{X_1 \mid X_2 = x_2\}$$

$$:= \int_{-\infty}^{\infty} \boxed{x_1} f_{X_1|X_2}(x_1 \mid x_2) dx_1$$

$\mu_{X_1|X_2=x_2}$ is a function of x_2. Think of it as a function of the value read by your sensor!

- **Conditional Covariance:**

$$\Sigma_{X_1|X_2=x_2} := \mathcal{E}\{(X_1 - \mu_{X_1|X_2=x_2})(X_1 - \mu_{X_1|X_2=x_2})^\top \mid X_2 = x_2\}$$

$$:= \int_{-\infty}^{\infty} (X_1 - \mu_{X_1|X_2=x_2})(X_1 - \mu_{X_1|X_2=x_2})^\top f_{X_1|X_2}(x_1 \mid x_2) dx_1$$

$\Sigma_{X_1|X_2=x_2}$ is a function of x_2. Think of it as a function of the value read by your sensor!

Peek at the KF (Kalman Filter)

Model

$$\begin{cases} x_{k+1} = A_k x_k + G_k w_k, & x_0 \text{ initial condition} \\ y_k = C_k x_k + v_k \end{cases}$$

$x \in \mathbb{R}^n$, $w \in \mathbb{R}^p$, $y \in \mathbb{R}^m$, $v \in \mathbb{R}^m$. Moreover, the random vectors x_0, and, for $k \geq 0$, w_k, v_k are all independent Gaussian (normal) random vectors.

Definition of Terms:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \begin{matrix} \sim x_k \\ \sim y_k \\ \sim y_0, \cdots, y_{k-1} \end{matrix}$$

$\widehat{x}_{k|k} := \mathcal{E}\{x_k | y_0, \cdots, y_k\}$ ← conditional mean at time k

$P_{k|k} := \mathcal{E}\{(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^\top | y_0, \cdots, y_k\}$ ← conditional cov at time k

$\widehat{x}_{k+1|k} := \mathcal{E}\{x_{k+1} | y_0, \cdots, y_k\}$

$P_{k+1|k} := \mathcal{E}\{(x_{k+1} - \widehat{x}_{k+1|k})(x_{k+1} - \widehat{x}_{k+1|k})^\top | y_0, \cdots, y_k\}$

Initial Conditions: $\widehat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}$, and $P_{0|-1} := P_0 = \text{cov}(x_0)$

For $k \geq 0$

Measurement Update Step:

$$K_k = P_{k|k-1} C_k^\top \left(C_k P_{k|k-1} C_k^\top + Q_k \right)^{-1}$$

(Kalman Gain)

$$\widehat{x}_{k|k} = \widehat{x}_{k|k-1} + K_k \left(y_k - C_k \widehat{x}_{k|k-1} \right)$$
$$P_{k|k} = P_{k|k-1} - K_k C_k P_{k|k-1}$$

Time Update or Prediction Step:

$$\widehat{x}_{k+1|k} = A_k \widehat{x}_{k|k}$$
$$P_{k+1|k} = A_k P_{k|k} A_k^\top + G_k R_k G_k^\top$$

End of For Loop (Just stated this way to emphasize the recursive nature of the filter)

Recursion in Kalman Filter

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \begin{matrix} \sim x_k \\ \sim y_k \\ \sim y_0, \dots, y_{k-1} \end{matrix}$$

We want to compute

$$\hat{x}_{k|k} = E\{\hat{x}_k \mid y_0, y_1, \dots, y_k\}$$

$X \mid \begin{bmatrix} Y \\ Z \end{bmatrix}$ (in Kalman filter to compute $\hat{x}_{k|k}$, we need to understand the distribution of random vector $X \mid \begin{bmatrix} Y \\ Z \end{bmatrix}$

Moreover,

→ for recursive computation, we will try to express

$\underbrace{X \mid \begin{bmatrix} Y \\ Z \end{bmatrix}}$ using $\underline{X \mid Z}$ (which is $\hat{x}_{k|k-1}$)

and Y