# Mathematics and numerics for data assimilation and state estimation – Lecture 7

Summer semester 2020

# Summary of lecture 6

- Recurrence and construction of invariant distributions for finite discrete-time Markov chains.

- Prediction, filtering and smoothing of Markov Chains

# Overview

1. Random variables on $\mathbb{R}^d$
2. Conditional probability density functions
3. Mixed rv

# Overview

# Extending definitions from discrete to continuous state-space rv

- Given $(\Omega, \mathcal{F}, \mathbb{P})$, recall that a discrete rv $X : \Omega \to A$ was defined by measurability constraint

$$X^{-1}(a) \in \mathcal{F} \quad \forall a \in A.$$

- And $\mathbb{P}_X(a) := \mathbb{P}(X = a)$ is a probability measure on the measurable space $(A, \mathcal{A})$ where

$$\mathcal{A} = \sigma(\{a_k\}) := \text{smallest } \sigma\text{-algebra containing the sets } \{a_1\}, \{a_2\}, \dots$$

- Example elements of $C \in \mathcal{A}$ : $\{a_1, a_2\}, \dots$
- An equivalent definition of discrete rv that extends to the continuous state-space setting: $X$ is a **measurable** mapping between measurable spaces $X : (\Omega, \mathcal{F}) \to (A, \mathcal{A})$, meaning that

$$X^{-1}(C) \in \mathcal{F} \quad \forall C \in \mathcal{A}.$$

# Random values/vectors (rv) on $\mathbb{R}^d$

- For a mapping $X : \Omega \to \mathbb{R}^d$ what sets $C \subset \mathbb{R}^d$ are relevant, in the sense that we seek the probability of events

$$\{X \in C\} = \{\omega \in \Omega \mid X(\omega) \in C\}?$$

- If all open sets in $\mathbb{R}^d$ are relevant, then we should be able to evaluate

$$\mathbb{P}_X(C) = \mathbb{P}(X \in C) \quad \text{for all open } C \subset \mathbb{R}^d,$$

- and $\mathbb{P}_X$ should be a probability measure on $(\mathbb{R}^d, \sigma(\text{all open sets in } \mathbb{R}^d))$.
- the above is called the Borel $\sigma$-algebra:

$$\mathcal{B}^d := \text{smallest } \sigma\text{-algebra containing all open sets in } \mathbb{R}^d.$$

# Random variables/vectors on $\mathbb{R}^d$

**Definition 1**

An rv on $\mathbb{R}^d$ is a measurable mapping $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}^d)$ satisfying,

$$\{X \in C\} = X^{-1}(C) \in \mathcal{F} \quad \forall C \in \mathcal{B}^d.$$

**Comments:**

- The definition extends to measurable mappings $X : (\Omega, \mathcal{F}) \to (\mathbb{S}, \mathcal{S})$ for any measurable space $(\mathbb{S}, \mathcal{S})$.

- We will only consider rv on $(\mathbb{R}^d, \mathcal{B}^d)$, and often just write

$$X : \Omega \to \mathbb{R}^d.$$

## Example 2 (Uniform distribution)

- Let $X \sim U[0,1]^d$ denote the rv on $\mathbb{R}^d$ with

$$\mathbb{P}(X \in C) = \mathrm{Leb}(C \cap [0,1]^d) = \int_C \mathbb{1}_{[0,1]^d}(x)\, \mathrm{Leb}(dx) \tag{1}$$

  for any $C \in \mathcal{B}^d$ and with $\mathrm{Leb}(\cdot)$ being the Lebesgue measure on $\mathbb{R}^d$.

- This measure associates to volumes of sets: for instance, for any $C = (a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_d, b_d)$,

$$\mathrm{Leb}(C) = \prod_{k=1}^{d} (b_k - a_k).$$

- From now on $dx := \mathrm{Leb}(dx)$, and we rewrite (1) with a density:

$$\mathbb{P}(X \in C) = \int_C \mathbb{1}_{[0,1]^d}(x)\, dx$$

# Probability density function (pdf)

## Definition 3

- Consider an rv $X \sim \mathbb{P}_X$ on $\mathbb{R}^d$.
- If there exists a mapping $\pi : \mathbb{R}^d \to [0, \infty)$ such that

$$\mathbb{P}_X(C) = \int_C \pi(x)\, dx \quad \forall C \in \mathcal{B}^d,$$

  then $\pi$ is called the **pdf** of $X$.
- $X$ has a pdf whenever $\mathbb{P}_X$ is absolutely continuous wrt the Lebesgue measure, meaning that for all $C \in \mathcal{B}^d$,

$$\text{if} \quad dC = \text{Leb}(C) = 0, \quad \text{then also} \quad \mathbb{P}_x(C) = 0.$$

- And then

$$\pi(x) = \frac{\mathbb{P}_X(dx)}{dx}, \quad dx\text{''} = \text{''tiny set surrounding } x \in \mathbb{R}^d.$$

- An rv with a pdf is called a **continuous** rv.

## Example 4 (Uniform rv)

For $X \sim U[0,1]^d$, we have

$$\mathbb{P}_X(C) = \int_C \mathbb{1}_{[0,1]^d}(x)\, dx$$

pdf derived from the Radon-Nikodym derivative:

$$\pi(x) = \frac{\mathbb{P}_X(dx)}{dx} = \frac{\mathbb{1}_{[0,1]^d}(x)dx}{dx} = \begin{cases} 0 & x \in (-\infty, 0) \cup (1, \infty) \\ ? & x \in \{0,1\} \\ 1 & \in (0,1) \end{cases}$$

Value of $\pi|_{\{0,1\}}$ does not matter, whatever value we assign on this measure 0 set, $\pi$ will be the very same pdf:

$$\mathbb{P}(X \in C) = \int_C \pi(y)\, dy \quad \forall C \in \mathcal{B}^d.$$

So let us write $\pi(x) = \mathbb{1}_{[0,1]^d}(x)$.

## Example 5 (Real-valued normal distribution)

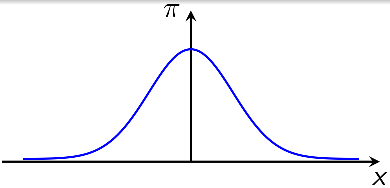
- Let $X \sim N(\mu, \sigma^2)$ denote the normal disribution on $\mathbb{R}$ with

$$\mathbb{P}(X \in C) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_C \exp\Big( -\frac{(x-\mu)^2}{2\sigma^2}\Big)\, dx$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$.

pdf:

$$\pi(x) = \frac{\exp\Big( -\frac{(x-\mu)^2}{2\sigma}\Big)}{\sqrt{2\pi\sigma^2}}$$



## Example 6 (Multivariate normal distribution)

- Let $X \sim N(\mu, \Sigma)$ denote the rv on $\mathbb{R}^d$ with

$$\mathbb{P}(X \in C) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \int_C \exp\Big( - \frac{(x-\mu)\cdot \Sigma^{-1}(x-\mu)}{2}\Big)\, dx$$

- With the pdf:

$$\pi(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} e^{-|x-\mu|_{\Sigma}^{2}/2}$$

Consider $X \sim N((0,0),\Sigma)$ on $\mathbb{R}^2$ with

$$\Sigma = [u_1\ u_2] \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} [u_1\ u_2]^T$$

$u_1, u_2$ orthonormal basis.

**Contour plot:**

# Cumulative distribution function (cdf)

## Definition 7

The cdf of a $d$-dimensional rv $X = (X_1, X_2, \dots, X_d)$ is defined by

$$F_X(x) = F_X(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_d \leq x_d) \quad \text{for } x \in \mathbb{R}^d.$$

Whenever $X$ is continuous $F_X$ is a primitive of $\pi$:

$$F_X(x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} \pi(y_1, \dots, y_d)\, dy_1 \, \dots dy_d$$

And in general, the cdf has the following properties:

- $F_X$ is non-decreasing and right continuous in each of its variables
- $$\lim_{x_1, x_2, \dots, x_d \to -\infty} F_X(x) = 0 \quad \text{and} \quad \lim_{x_1, x_2, \dots, x_d \to \infty} F_X(x) = 1.$$

## Example 8 (Discrete rv)

$X \sim \textit{Bernoulli}(1/2)$ yields

$$F(x) = \begin{cases} 0, & x < 0 \\ 1/2, & x \in [0, 1) \\ 1, & x > 1 \end{cases}$$

and formally

$$\pi(x) = F'(x) = 0.5\delta_0(x) + 0.5\delta_1(x).$$

## Example 9 (Sum of discrete and continuous rv that becomes ...)

Let $X = Y + 2Z$ where $Y \sim U[0,1]$ and $Z \sim \mathit{Bernoulli}(1/2)$ are independent. Then

$$F(x) = \mathbb{P}(Y + 2Z \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ x/2 & \text{if } x \in [0,1] \\ 1/2 & \text{if } x \in [1,2) \\ (x-1)/2 & \text{if } x \in [2,3) \\ 1 & \text{if } x \geq 3 \end{cases}$$

and

$$\pi(x) = F'(x) = \frac{1}{2}\mathbb{1}_{[0,1]\cup[2,3]}(x).$$

(Formally,

$$\pi(x) = \pi_Y * \pi_{2Z}(x)$$

where $\pi_{2Z} = 0.5\delta_0(x) + 0.5\delta_2(x)$.)

# Joint pdfs and cdfs

For rv $X : \Omega \to \mathbb{R}^d$ and $Y : \Omega \to \mathbb{R}^k$,

- the mapping $(X, Y) : (\Omega, \mathcal{F}) \to (\mathbb{R}^{d+k}, \mathcal{B}^{d+k})$ is also an rv
- with joint cdf
$$F_{XY}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$
where $X \leq x$ etc. should be read component-wise for vectors
- and, whenever $(X, Y)$ is continuous, the joint pdf
$$\pi_{XY}(x, y) = \frac{\mathbb{P}(X \in dx, Y \in dy)}{dx\, dy}.$$
- To avoid clutter, one often suppresses $XY$ subscripts.

# Joint pdfs and cdfs

The notation extends naturally to the joint distribution of a sequence of rv $\{X_k\}$,

$$\begin{aligned} F(x_1, \ldots, x_n) &:= \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) \\ &= \int_{y_1 \leq x_1, \ldots, y_n \leq x_n} \pi(y_1, \ldots, y_n) dy_1 \ldots dy_n, \end{aligned}$$

where the last equality with the pdf $\pi$ is valid when $(X_1, \ldots, X_n)$ is continuous.

# Independence of rv

A finite sequence of rv $X_k : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}^d)$ for $k = 1, \ldots, n$ is independent if

- $$\mathbb{P}(X_1 \in C_1, \ldots, X_n \in C_n) = \prod_{k=1}^{n} \mathbb{P}(X_k \in C_k)$$

  for all $C_1, \ldots, C_n \in \mathcal{B}^d$.
- or equivalently, if

  $$F(x_1, x_2, \ldots, x_n) = \prod_{k=1}^{n} F_{X_k}(x_k) \quad \forall x_1, x_2, \ldots, x_n \in \mathbb{R}^d,$$

- or, if all $X_k$ are continuous, equivalently if

  $$\pi(x_1, x_2, \ldots, x_n) = \prod_{k=1}^{n} \pi_{X_k}(x_k) \quad \text{for almost all} \quad x_1, x_2, \ldots, x_n \in \mathbb{R}^d$$

- A countable sequence of rv $\{X_k\}$ is independent if any of the above conditions hold for any finite subsequence.

# Expectations of rv

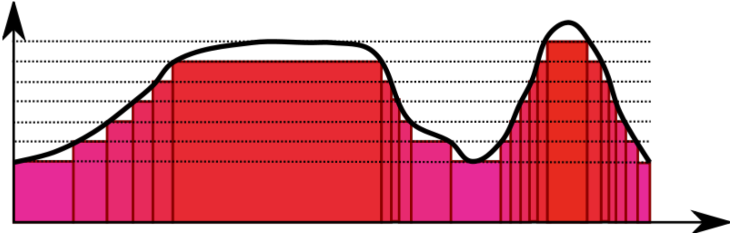
The expectation of $X : \Omega \to \mathbb{R}^d$ is defined by

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, \mathbb{P}(d\omega)$$

where for non-negative $X = (X_1, \ldots, X_d)$, each component of the right side is defined by

$$\int_\Omega X_j \, d\mathbb{P} := \sup_{Y \leq X_j, Y \text{ simple}} \int_\Omega Y \, d\mathbb{P}$$

where "simple" $= \{ Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}) \mid Y \text{ is simple} \}$.



# Expectations of rv

The expectation of $X : \Omega \to \mathbb{R}^d$ is defined by

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, \mathbb{P}(d\omega)$$

where for non-negative $X = (X_1, \ldots, X_d)$, each component of the right side is defined by

$$\int_\Omega X_j \, d\mathbb{P} := \sup_{Y \le X_j, Y \text{ simple}} \int_\Omega Y \, d\mathbb{P}$$

where "simple" $= \{ Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}) \mid Y \text{ is simple } \}$.
And in general,

$$\int_\Omega X_j \, d\mathbb{P} := \underbrace{\int_\Omega X_j^+ \, d\mathbb{P}}_{I_j^+} - \underbrace{\int_\Omega X_j^- \, d\mathbb{P}}_{I_j^-}$$

where

$$X_j^+ = \max\{X_j, 0\} \quad \text{and} \quad X_j^- = \max\{-X_j, 0\}.$$

**Observation:** $\mathbb{E}[X_j]$ exists whenever at least one of $I_j^+$ and $I_j^-$ are finite.

# Covariance function

- The definition extends straightforwardly to functions of rv:

$$\mathbb{E}\left[g(X)\right] = \int_{\Omega} g(X(\omega))\,\mathbb{P}(d\omega)$$

- Whenever $\mathbb{E}\left[g(X)\right]$ exists, the change of variables formula yields the equivalent representations (Durrett, Theorem 1.6.9)

$$\mathbb{E}\left[g(X)\right] = \int_{\mathbb{R}^d} g(x)dF_x(x) = \int_{\mathbb{R}^d} g(x)\pi_X(x)\,dx$$

(last equality valid when $\pi_X$ exists).

**Main idea of proof:**

1. In the $1D$ setting with $g(X) = \sum_{a\in A} g(a)\mathbb{1}_{X=a}$, then

   $\mathbb{E}\left[g(X)\right] =$

2. For non-discrete rv, approximate by sequence of simple functions.

## Covariance function

- In what remains, we will assume that the rvs are continuous so that we can employ pdfs in the expectations of rv.
- The covariance of the $d$-dimensional rv $X$ with mean $\mu$ is the $d \times d$ matrix defined by

$$\mathrm{Cov}(X) = \mathbb{E}\left[(X-\mu)(X-\mu)^T\right] = \int_{\mathbb{R}^d} (x-\mu)(x-\mu)^T \pi_X(x) dx$$

- In the special case of 1-dimensional rv, $\mathrm{Cov}(X) = \mathrm{Var}(X)$

**Example 10**

$X \sim U[0,1]$, yields

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \mathbb{1}_{[0,1]}(x) dx = 1/2$$

and

$$\mathrm{Var}(X) = \int_{\mathbb{R}} (x-1/2)^2 \mathbb{1}_{[0,1]}(x) dx = 1/12.$$

## Example 11 (Multivariate normal distribution)

For $X \sim N(\mu, \Sigma)$ with

$$\pi_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} e^{-(x-\mu)\cdot\Sigma^{-1}(x-\mu)/2},$$

one can show that

$$\mathbb{E}[X] = \mu$$

and

$$\mathrm{Cov}(X) = \Sigma.$$

# Overview

# Marginal and conditional densities

For a continuous rv $(X, Y) : \Omega \to \mathbb{R}^d \times \mathbb{R}^k$, we define

- the marginal pdf for $X$ by

$$\pi_X(x) := \int_{\mathbb{R}^k} \pi_{XY}(x, y)dy$$

- and the conditional density of $X$ given $Y = y$ by

$$\pi_{X|Y}(x|y) := \frac{\pi_{XY}(x, y)}{\pi_Y(y)} \quad \text{(using division-by-zero convention).}$$

**Properties:**

- $\pi_{X|Y}(\cdot|y)$ is a density whenever $\pi_Y(y) > 0$
- the disintegration property holds

$$\pi_{XY}(x, y) = \pi_{X|Y}(x|y)\pi_Y(y)$$

- and it extends to multiple rv $(X, Y, Z)$:

$$\pi(x, y, z) = \pi(x|y, z)\pi(y|z)\pi(z) \quad \text{etc}$$

**Formal motivation for scalar-valued $Y$ :** For any $C \in \mathcal{B}^d$,

$$\mathbb{P}(X \in C \mid Y \in [y, y + \Delta y]) = \frac{\int_y^{y+\Delta y} \int_C \pi_{XY}(x, y)\, dx\, dy}{\int_{\Delta y} \pi_Y(y)\, dy}$$

$$\approx$$

So when $\pi_{XY}$ is continuous, we have the relation

$$\lim_{\Delta y \to 0} \mathbb{P}(X \in C \mid Y \in \Delta y) = \int_C \pi_{X|Y}(x \mid y) dx,$$

for neighborhoods $\Delta y$ of $y$.

# Expectation of X given Y

**Definition 12 (Conditional expectation)**

For continuous rv $(X, Y) : \Omega \to \mathbb{R}^d \times \mathbb{R}^k$ and mapping $g(X)$ such that $\mathbb{E}[|g(X)|] < \infty$, we define

$$\mathbb{E}[g(X) \mid Y = y] := \int_{\mathbb{R}^d} g(x)\, \pi_{X|Y}(x|y)\, dx,$$

and the related continuous rv

$$\mathbb{E}[g(X) \mid Y](\omega) := \mathbb{E}[g(X) \mid Y = Y(\omega)].$$

# Properties

For integrable $g(X, Y)$, the **tower property** holds:

$$\mathbb{E}\left[\mathbb{E}\left[g(X, Y) \mid Y\right]\right] := \mathbb{E}\left[g(X, Y)\right]$$

and if $g(X, Y) = f(X)h(Y)$, then

$$\mathbb{E}\left[\mathbb{E}\left[g(X, Y) \mid Y\right]\right] := \mathbb{E}\left[h(Y)\mathbb{E}\left[f(X) \mid Y\right]\right].$$

**Verification:**

## Example 13

Let $Y, Z \sim U[0,1]$ and independent, and $X = Y + Z$.

Then for $y \in [0,1]$,

$$\text{shortcut: } X|\{Y = y\} = Z + y \sim U[y, 1+y]$$

giving

$$\pi_{X|Y}(x|y) = \mathbb{1}_{[y,y+1]}(x)$$

and

$$\mathbb{E}\left[X \mid Y = y\right] = \int_{\mathbb{R}} x \mathbb{1}_{[y,y+1]}\, dx = \frac{y+1}{2}.$$

Proper argument:

$$\pi_{XY}(x,y) = \pi_{ZY}(x-y, y) = \mathbb{1}_{[0,1]}(x-y)\mathbb{1}_{[0,1]}(y) \quad \text{etc.}$$

# Overview

1. Random variables on $\mathbb{R}^d$
2. Conditional probability density functions
3. Mixed rv

# General definition for conditional expectation

- Random variables may also be mixed, meaning neither discrete nor continuous.

**Example 14**

$X = YZ$ where $Y \sim \mathit{Bernoulli}(1/2)$ and $Z \sim U[0,1]$ with $Y \perp Z$.
Then formally,

$$\pi_X(x) = \frac{\delta_0(x) + \mathbb{1}_{[0,1]}(x)}{2}$$

- Mixed rv do not have a pdf, so Definition 12 does not apply to conditional expectations of mixed rv.

- **Objective for next lecture:** obtain a unifying definition for conditional probability.

# What describes an rv fully?

- An rv can be discrete, mixed or continuous.

- Regardless of that, $X$ is uniquely described by its distribution $\mathbb{P}_X$, and also by its cdf

$$F_X(x) = \mathbb{P}(X \leq x)$$

and, if it exists, also by its pdf

$$\pi_X(x) = \frac{\mathbb{P}(X \in dx)}{dx}.$$

# Next time

- Conditional expectations in general (also for mixed rv)
- Orthogonal projections on closed subspaces of $L^2(\Omega)$
- Bayesian inverse problems and well-posedness.