Mathematics and numerics for data assimilation and state estimation – Lecture 2

Summer semester 2020

Overview

1. Summary of lecture 1
2. Discrete random variables
   - Independence of random variables and events
   - Expected value and moments
3. Conditional probability and expectation

On ubungs, presentation and lectures

- 10:30-12:00 on most Fridays.
- Structure: 5-10 questions, which I will put up in pdf form on course webpage and on Moodle. Roughly 30 minutes work in groups or alone, where I will be present for discussions, thereafter solutions in plenary by me and/or you.
- No hand-ins, unless you want to (i.e., only for feedbac, kdoes not affect grade).
- The only "graded" part of the course, in the form of bonus points, is the presentation early July, and, of course, the final exam.
- Presentations can be done alone or in groups of maximum 2 people.
- Lectures after July 17th moved to first week of June.

Overview

Measurable spaces and probability measures

- introduced a probabilty space $(\Omega, \mathcal{F}, \mathbb{P})$
- discrete random variable $X : \Omega \to A = \{a_1, a_2, \dots, \}$ satisfies the event constraints

$$X^{-1}(a) = \{\omega \in \Omega \mid X(\omega) = a\} \in \mathcal{F} \quad \text{for all} \quad a \in A.$$

- X can be represented by a simple function

$$X(\omega) = \sum_{a \in A} a \mathbb{1}_{X=a}(\omega). \quad \text{where } \mathbb{1}_{X=a}(\omega) := \begin{cases} 1 & \text{if } X(\omega) = a \\ 0 & \text{otherwise} \end{cases}$$

Overview

1. Summary of lecture 1
2. Discrete random variables
   - Independence of random variables and events
   - Expected value and moments
3. Conditional probability and expectation

Discrete random variables 2

Example 1 (Coin toss, $X \sim$ Bernoulli(p))

- image-space outcomes $A = \{0, 1\}$,
- $$\Omega = \{Heads, Tails\}, \qquad \mathcal{F} = \{\emptyset, \{Heads\}, \{Tails\}, \Omega\}$$
- $X(Heads) = 1$ and $X(Tails) = 0$ and

$$\mathbb{P}(X = 1) = \mathbb{P}(X^{-1}(1)) = \mathbb{P}(Heads) = p, \quad \mathbb{P}(X = 0) = \mathbb{P}(Tails) = 1 - p.$$

Comment from last lecture: image-outcomes $\{a_1, a_2, \ldots, \}$ may not be associated uniquely to (probability-space) outcomes in Ω.

Larger set of outcomes in Ω than in A

Alternative, and admittedly confusing, probability space for the same rv as in the preceding example:

Example 2 (Coin toss, $X \sim \text{Bernoulli}(p)$)

- image-space outcomes $A = \{0, 1\} \subset \mathbb{R}$,
- $\Omega = \{Heads, Tails, Nose\}$ and

$$\mathcal{F} = \{\emptyset, \{Nose\}, \{Heads\}, \{Tails\}, \{Nose, Heads\},$$
$$\{Nose, Tails\}, \{Heads, Tails\}, \Omega\}$$

- $X^{-1}(1) = \{Heads, Nose\}$ and $X^{-1}(0) = \{Tails\}$ and

$$\mathbb{P}(X = 1) = \mathbb{P}(X^{-1}(1)) = \mathbb{P}(\{Heads, Nose\}) = p,$$
$$\mathbb{P}(X = 0) = \mathbb{P}(Tails) = 1 - p.$$

Motivation: if, for instance, you want to represent both a coin toss and a three-sided-die toss in the same probability space.

Joint rv

If $X : \Omega \to A$ and $Y : \Omega \to B = \{b_1, b_2, \ldots\}$ are two discrete rv on the same probability space, then

- $(X, Y) : \Omega \to A \times B$ is also a discrete rv with countable set of outcomes
$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$
- with joint distribution:
$$\mathbb{P}_{(X,Y)}((a, b)) = \mathbb{P}(X = a, Y = b).$$
- Question: why is $\mathbb{P}(X = a, Y = b)$ defined? Answer: when we say X and Y are defined on the same probability space, this entails that
$$\{X = a\}, \{Y = b\} \in \mathcal{F} \quad \underbrace{\Longrightarrow}_{\text{since } \mathcal{F} \text{ is } \sigma-\text{algebra}} \quad \{X = a\} \cap \{Y = b\} \in \mathcal{F},$$
and
$$\mathbb{P}(X = a, Y = b) = \mathbb{P}(\{X = a\} \cap \{Y = b\}).$$

Definition 3 (Independence of two rv)

If $X : \Omega \to A$ and $Y : \Omega \to B = \{b_1, b_2, \ldots\}$ are two discrete rv on the same probability space[a] are said to be independent random variables if

$$\mathbb{P}(X = a, Y = b) = \mathbb{P}(X = a)\mathbb{P}(Y = b), \qquad \forall a \in A \quad b \in B.$$

Notation: $X \perp Y$.

[a] From now on, it will be implicitly assumed that all rv are defined on the same probability space, unless otherwise stated.

Example 4

Given independent coin tosses $X_k \sim \mathit{Bernoulli}(1/2)$ for $k = 1, 2$, describe the smallest possible σ-algebra on which the rv (X_1, X_2) is defined.

Solution:

Example 5 (one coin toss and one three-sided-die toss)

- Consider $X : \Omega \to \{0, 1\}$ and and $Y : \Omega \to \{1, 2, 3\}$ both defined on the probability space from Example 2.
- Recall that $X^{-1}(1) = \{Heads, Nose\}$ and $X^{-1}(0) = \{Tails\}$ and let us assume that

$$\mathbb{P}(X = 1) = 1/2, \quad \mathbb{P}(X = 0) = 1/2$$

  and that $Y^{-1}(1) = \{Heads\}$, $Y^{-1}(2) = \{Nose\}$ and $Y^{-1}(3) = \{Tails\}$.
- Quation: For $p = 1/2$, what is

$$\mathbb{P}(X = 0, Y \in \{1, 2\}) =?$$

- Question: Are X and Y independent?

Independence of multiple rv

Definition 6

Let $X_k : \Omega \to A_k$ for $k = 1, 2, \ldots, N$, be a finite sequence of discrete rv. Then $X_1, X_2, \ldots, X_N$ are independent provided

$$\mathbb{P}(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N) = \prod_{k=1}^{N} \mathbb{P}(X_k = a_k) \tag{1}$$

for all $a_1 \in A_1$, $a_2 \in A_2$, $\ldots$, $a_n \in A_N$.
Extension: A **countable** sequence of discrete rv $X_1, X_2, \ldots$ are independent provided every finite subsequence $\{X_{k_j}\}_j$ satisfies (1).

Example 7

Let $X_i \sim \textit{Bernoulli}(p)$ for $i = 1, \ldots, N$ with joint distribution

$$\mathbb{P}\left(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N\right) = p^{\sum_{k=1}^{N} a_k}(1-p)^{N - \sum_{k=1}^{N} a_k}$$

for any $a_1, \ldots, a_N \in \{0, 1\}$. Then $X_1, X_2, \ldots$ are independent and identically distributed (iid).

Example 8 (Functions of joint discrete rv are also discrete rv)

Let $X_i \sim \mathit{Bernoulli}(p)$ be independent for $i = 1, 2, \ldots, N$ and

$$S_N = f(X_1, \ldots, X_N) := \sum_{i=1}^{N} X_i.$$

Then

$$\mathbb{P}(S_N = k) = \binom{N}{k}(1-p)^{N-k}p^k$$

S_N is called the **Binomial distribution** with degrees of freedom N and p, and we write $S_N \sim B(N, p)$.

Comment: the number of different ways the event $\{S_N = k\}$ when flipping N independent coins once equals factor in the $k+1$-th summand of

$$((1-p)+p)^N = (1-p)^N + \binom{N}{1}p(1-p)^{N-1} + \ldots + \binom{N}{k}p^k(1-p)^{N-k} + \ldots$$

Independence of events

Equation (1) is on the form:

$$\mathbb{P}\left(\bigcap_{k=1}^{N}\{X_k = a_k\}\right) = \mathbb{P}(\text{intersection of events}) = \text{Product of}\left[\mathbb{P}(\text{each event})\right]$$

Definition 9

A finite sequence of events $H_1, H_2, \ldots, H_N$ that belongs to $\mathcal{F}$ are independent provided

$$\mathbb{P}\left(\bigcap_{k=1}^{N} H_k\right) = \prod_{k=1}^{N} \mathbb{P}(H_k) \tag{2}$$

A **countable** sequence of events A_1, A_2, belonging to $\mathcal{F}$ are independent provided finite subsequence $\{A_{k_j}\}_j$ satisfies (2).

Connection between independence of rv and independence of events

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we can assign an rv to each event $H \in \mathcal{F}$ as follows

$$\mathbb{1}_H(\omega) := \begin{cases} 1 & \omega \in H \\ 0 & \text{otherwise} \end{cases}.$$

Easy consequence of preceding definition: $\mathbb{1}_{H_1}$ and $\mathbb{1}_{H_2}$ are independent if and only if

$$\mathbb{P}(H_1 \cap H_2) = \mathbb{P}(H_1)\mathbb{P}(H_2).$$

Expectation of rv

Definition 10

For a discrete rv $X : \Omega \to A \subset \mathbb{R}^d$, the expectation X is defined as

$$\mathbb{E}[X] := \int_\Omega X(\omega)\mathbb{P}(d\omega) = \sum_{a \in A} a\mathbb{P}\left(X = a\right)$$

Motivation of the above integral:

$$\int_\Omega X(\omega)\mathbb{P}(d\omega) =$$

- The condition

$$\mathbb{E}[|X|] = \sum_{a \in A} |a|\mathbb{P}\left(X = a\right) < \infty$$

  is a sufficent condition for $\mathbb{E}[X]$ being defined and bounded.
- Example for $X \sim \mathit{Beronoulli}(p)$

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

Expectation of rv

Definition 11

For a discrete rv $X : \Omega \to A \subset \mathbb{R}^d$, the expectation X is defined as

$$\mathbb{E}[X] := \int_\Omega X(\omega)\mathbb{P}(d\omega) = \sum_{a\in A} a\mathbb{P}\left(X = a\right)$$

- The condition
$$\mathbb{E}[|X|] = \sum_{a\in A} |a|\mathbb{P}\left(X = a\right) < \infty$$
is a sufficent condition for $\mathbb{E}[X]$ being defined and bounded.
- For mappings $f : \mathbb{R}^d \to \mathbb{R}^k$ and rv $f(X)$ the above definition readily extends:
$$\mathbb{E}[f(X)] = \sum_{a\in A} f(a)\mathbb{P}\left(X = a\right).$$
- Example for $X \sim \mathit{Beronoulli}(p)$
$$\mathbb{E}[X] =$$

Properties of the expectation

- For mappings $f : \mathbb{R}^d \to \mathbb{R}^k$ and rv $f(X)$, the expectation becomes

$$\mathbb{E}[f(X)] = \sum_{a \in A} f(a)\mathbb{P}(X = a).$$

- For a pair of rv $X : \Omega \to A \subset \mathbb{R}^d$ and $Y : \Omega \to B \subset \mathbb{R}^d$, it holds for any $c \in \mathbb{R}$, that

$$\mathbb{E}[X + cY] = \mathbb{E}[X] + c\,\mathbb{E}[Y]$$

provided $\mathbb{E}[|X|] + \mathbb{E}[|Y|] < \infty$ (sufficient condition).

Motivation:

Properties of the expectation 2

- Probability of events can be expressed through expectations:

$$\mathbb{P}(H) = \qquad\qquad = \mathbb{E}[\mathbb{1}_H]$$

  for any $H \in \mathcal{F}$.

- Expectation of discrete rv of the form $f(X, Y)$ where $X : \Omega \to A$ and $Y : \Omega \to B$:

$$\mathbb{E}[f(X, Y)] =$$

Variance of an rv

- For $X : \Omega \to A \subset \mathbb{R}$
$$F(k) = \mathbb{E}[(X-k)^2]$$
is the squared deviation of X from k in expectation.
- For $\mu := \mathbb{E}[X]$, and provided $\mathbb{E}[X^2] < \infty$, it can be shown that
$$F(\mu) \leq F(k) \quad \text{for all} \quad k \in \mathbb{R},$$
- Which motivates the variance of X:
$$\mathrm{Var}(X) := \mathbb{E}[(X-\mu)^2]$$
- For $X \sim \mathit{Bernolli}(p)$, $\mu = p$ and
$$\mathrm{Var}(X) =$$

Notation with same meaning

For events $H_1, H_2, \ldots \in \mathcal{F}$, the following notation is used interchangeably in the literature

$$\mathbb{P}\left(H_1 H_2 \ldots H_n\right) = \mathbb{P}\left(H_1, H_2, \ldots, H_n\right) = \mathbb{P}\left(\bigcap_{j=1}^{n} H_j\right).$$

And since

$$\mathbb{1}_{\bigcap_{j=1}^{n} H_j} = \prod_{i=1}^{n} \mathbb{1}_{H_j}.$$

we have that

$$\mathbb{P}\left(\bigcap_{j=1}^{n} H_j\right) = \mathbb{E}[\mathbb{1}_{\bigcap_{j=1}^{n} H_j}] = \mathbb{E}[\prod_{i=1}^{n} \mathbb{1}_{H_j}].$$

Overview

Conditional probability

Definition 12

For two events $G, H \in \mathcal{F}$ where $\mathbb{P}(H) > 0$, the conditional probability of G given H is given by

$$\mathbb{P}(G \mid H) = \frac{\mathbb{P}(G \cap H)}{\mathbb{P}(H)}$$

Whenever $\mathbb{P}(H) > 0$, the mapping $\mathbb{P}(\cdot \mid H) : \mathcal{F} \to [0, 1]$ is a probability measure.[1]

Verification:

[1] And it remains to define $\mathbb{P}(\cdot \mid H)$ for zero-probability events H.

Simplification in some settings (direct use of conditioning):
For X, Y and $f(X, Y)$ discrete rv,

$$\mathbb{P}\left(f(X,Y) = c \mid Y = b\right) = \mathbb{P}\left(f(X,b) = c\right), \qquad \text{if } \mathbb{P}\left(Y = b\right) > 0. \quad (3)$$

Example 13

Let $X_1, X_2, X_3 \sim \mathit{Bernoulli}(p)$ and independent rv. Let $Z = X_1 + X_2 + X_3$. Compute

$$\mathbb{P}\left(Z \geq 1 \mid X_1 = 0\right)$$

Solution:

Example 14 (Example where conditioning information is used "implicitly")

Let $X_1, X_2, X_3 \sim \mathit{Bernoulli}(p)$ and independent rv. Let $Z = X_1 + X_2 + X_3$. Compute

$$\mathbb{P}(X_1 = 1 \mid Z = 2)$$

Solution:

Definition 15 (Conditional expectation)

For a discrete rv $X : \Omega \to A$ and an event $H \in \mathcal{F}$ with $\mathbb{P}(H) > 0$, we define the conditional expectation of X given H as

$$\mathbb{E}[X \mid H] := \int_\Omega X(\omega)\mathbb{P}(d\omega \mid H) = \sum_{a \in A} a\mathbb{P}(X = a \mid H)$$

- Property: $\mathbb{E}[X \mid H] = \mathbb{E}[X \mathbb{1}_H]/\mathbb{P}(H)$.
  Verfication:

- Implication: $\mathbb{E}[|X| \mid H] \leq \mathbb{E}[|X|]/\mathbb{P}(H)$.

Example 16

Let X be a three-sided fair die, meaning

$$\mathbb{P}(X = k) = \frac{1}{3} \qquad \text{for } k = 1, 2, 3.$$

Compute $\mathbb{E}[X \mid X \geq 2]$.

Solution:

Conditioning on zero-probability events

For events $G, H \in \mathcal{F}$, it is not clear how interpret the definition

$$\mathbb{P}(G \mid H) := \frac{\mathbb{P}(G \cap H)}{\mathbb{P}(H)}$$

when $\mathbb{P}(H) = 0$.

Is an extension of the definition needed? May not seem needed as zero-probability events "never" happen anyway, but often it is convenient to use the same co-domain for all rv studied, say for example

$$X_k : \Omega \to \mathbb{N}$$

with $X_k(\Omega) = \mathbb{N} \setminus \{k\}$ for $k = 1, 2, \ldots$.

Also any event $\{Y = y\}$ is a zero-probability event for a continuous rv!

Conditioning on zero-probability events 2

Definition 17 (Division-by-zero convention)

For any $c \in \mathbb{R}$ we will, in all of this course, make use of the following convention

$$\frac{c}{0} := 0.$$

Motivation: Then $\frac{a}{b}$ is defined for any $a, b \in \mathbb{R}$, but it gives algebra a quirk

$$b(a/b) = \begin{cases} a & \text{if } b \neq 0 \text{ or } a = 0 \\ 0 & \text{if } b = 0. \end{cases}$$

Definition 18 (Generalization of Definition 12)

For **any** pair of events $G, H \in \mathcal{F}$, we define

$$\mathbb{P}(G \mid H) := \frac{\mathbb{P}(G \cap H)}{\mathbb{P}(H)}$$

where we note that by the division-by-zero convention

$$\mathbb{P}(G \mid H) = 0 \qquad \text{if } \mathbb{P}(H) = 0.$$

Implications:

- The definition of conditional expectation "naturally" extends to any zero-probability events $H \in \mathcal{F}$:

$$\mathbb{E}[X|H] := \sum_{a \in A} a\mathbb{P}(X = a \mid H) = 0.$$

- Direct use of conditioning, cf. equation (3), does **not** extend. Meaning, it is not generally true that

$$\mathbb{P}(f(X,Y) = c \mid Y = b) = \mathbb{P}(f(X,b) = c), \qquad \text{if } \mathbb{P}(Y = b) = 0.$$

(See exercises.)

Conditioning on random variables

- We have defined the conditional probability $\mathbb{P}(G \mid H)$ for any pair events G, H.
- So for rv $X : \Omega \to A$ and $Y\Omega \to B$, the following quantities are all defined
$$\mathbb{P}(X = a \mid Y = b) \qquad \text{for any } a \in A,\ b \in B.$$
- Fixing the event $\{X = a\}$, we may introduce the function $\psi : B \to [0, 1]$
$$\psi(b) = \mathbb{P}(G \mid \{Y = b\})$$
- and the function $\phi : \Omega \to [0, 1]$ by
$$\phi(\omega) := \mathbb{P}(X = a \mid \{Y = Y(\omega)\})$$
(curly brackets in the $\{Y = Y(\omega)\}$ notation here is only used to emphasize that we have events and is not really needed).
- Claim: ϕ is discrete rv. Verification: $\phi(\Omega) = \psi(B) = \{\psi(b_1), \psi(b_2), \ldots\}$, and for any element $\psi(b)$ in the image, $\phi^{-1}(\psi(b)) = \{Y = b\} \in \mathcal{F}$.

Conditioning on random variables 2

- The mapping ϕ above was just introduced for clarification. The customary notation for "the probability of $X = a$ given Y" is

$$\mathbb{P}(X = a \mid Y)(\omega) := \mathbb{P}(X = a \mid \{Y = Y(\omega)\}) \qquad \omega \in \Omega$$

- For each $a \in A$, the mapping $\mathbb{P}(X = a \mid Y) : \Omega \to [0, 1]$ is thus an rv.

Example 19

Consider the setting of Example 5: a coin toss $X : \Omega \to \{0, 1\}$ and a die roll $Y : \Omega \to \{1, 2, 3\}$, $\Omega = \{Heads, Nose, Tails\}$,

$$X^{-1}(1) = \{Heads, Nose\} \quad \text{and} \quad X^{-1}(0) = \{Tails\}$$

$$Y^{-1}(1) = \{Heads\}, \quad Y^{-1}(2) = \{Nose\} \quad \text{and} \quad Y^{-1}(3) = \{Tails\}.$$

and

$$\mathbb{P}(Heads) = \mathbb{P}(Nose) = 1/4, \quad \text{and} \quad \mathbb{P}(Tails) = 1/2.$$

Then

$\mathbb{P}(X = 0 \mid Y)(Heads) =$

$\mathbb{P}(Y = 1 \mid X)(Nose) =$

Definition 20 (Expectation of X given Y)

For discrete rv $X : \Omega \to A \subset \mathbb{R}^d$ and $Y : \Omega \to B \subset \mathbb{R}^k$ with $|\mathbb{E}[X]| < \infty$, the mapping $\mathbb{E}[X \mid Y] : \Omega \to \mathbb{R}^d$ is defined by

$$\mathbb{E}[X \mid Y](\omega) := \sum_{a \in A} a\, \mathbb{P}\left(X = a \mid Y\right)(\omega) = \sum_{a \in A} a\, \mathbb{P}\left(X = a \mid \{Y = Y(\omega)\}\right).$$

Example 21

Consider the setting of Example 19.

$$\begin{aligned} \mathbb{E}[Y \mid X](Nose) &= \sum_{k=1}^{3} k\, \mathbb{P}\left(Y = k \mid X\right)(Nose) \\ &= \sum_{k=1}^{3} k\, \mathbb{P}\left(Y = k \mid X = X(Nose)\right) \\ &= \dots \end{aligned}$$

Motivation for $\mathbb{E}[X \mid Y]$

Why is $\mathbb{E}[X \mid Y]$ relevant?

If you have an observation $Y(\omega)$ (i.e., you know $Y(\omega)$ but not ω), but seek $X(\omega)$, then what is the best function $g(Y(\omega))$ to approximate $X(\omega)$?

Answer: $\mathbb{E}[X \mid Y]$ is the best approximation of X in mean-square sense, meaning

$$\mathbb{E}[|X - \mathbb{E}[X \mid Y]|^2] \leq \mathbb{E}[|X - g(Y)|^2]$$

for all mappings $g : \mathbb{R}^k \to \mathbb{R}^n$ (assuming $X(\Omega) \subset \mathbb{R}^n$ and $Y(\Omega) \subset \mathbb{R}^k$).

Properties of $\mathbb{E}[X \mid Y]$ left to prove as ubung exercises:

- Verify that $\mathbb{E}[X \mid Y]$ is a discrete rv.

- If $X \perp Y$, then

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X] \qquad \mathbb{P} - \text{almost surely}$$

- The **tower property**

$$\mathbb{E}[\,\mathbb{E}[X \mid Y]\,] = \mathbb{E}[X].$$

Next time

- Conditional expectations
- Convergence of random variables
- Random walks and discrete time Markov Chains