Mathematics and numerics for data assimilation and state estimation – Lecture 3

Mathematics for Uncertainty Quantification

RWTH AACHEN UNIVERSITY

Summer semester 2020

Summary of lecture 2

- Random vectors $(X, Y) : \Omega \to A \times B$ and joint distributions

$$\mathbb{P}_{(X,Y)}((a, b)) = \mathbb{P}(X = a, Y = b).$$

- Independence of rv

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

and of events

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

- Expectation of $X : \Omega \to A$,

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

Summary of lecture 2

- Variance of X. Defined for a scalar-valued rv (meaning $X : \Omega \to A \subset \mathbb{R}^d$ with $d = 1$),

$$\mathrm{Var}(X) := \mathbb{E}\left[\,(X-\mu)^2\right].$$

- Property: μ is the best constant-value approximation of X in the following sense

$$\mathbb{E}\left[\,(X-\mu)^2\right] \leq \mathbb{E}\left[\,(X-k)^2\right] \qquad \text{for all} \quad k \in \mathbb{R}.$$

Plan for this lecture

- Conditional probabilities and expectations
- Conditioning on events: "probability of X given H"

$$\mathbb{P}(X = a \mid H) \qquad H \in \mathcal{F},$$

- Conditioning on rv: "probability of X given rv Y":

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

Interesting property

$$\mathbb{E}\left[\,|X - \mathbb{E}[X \mid Y]\,|^2\right] = \mathbb{E}\left[\,|X - f(Y)|^2\right]$$

for any mapping $f(Y) \in \mathbb{R}^d$.

Conditional probability

Definition 1

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,

$$\mathbb{P}(f(X, Y) = c \mid Y = b) = \frac{\mathbb{P}(f(X, b) = c)}{\mathbb{P}(Y = b)}, \qquad \text{if } \mathbb{P}(Y = b) > 0. \quad (1)$$

Example 2

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

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

Solution:

Example 3 (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 4 (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

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

- Property:

$$\mathbb{E}\left[X \mid H\right] = \mathbb{E}\left[X \mathbb{1}_H\right] / \mathbb{P}(H) \tag{2}$$

  Verfication:

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

Example 5

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 6 (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 7 (Generalization of Definition 1)

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 (1), extends. Meaning,

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

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).

Conditioning on random variables 2

- The mapping $\phi(\omega)$ was introduced to clarify that $\mathbb{P}(X = a \mid \{Y = Y(\omega)\})$ is a function of ω.
- The customary notation for these conditional probabilities is as follows:

Definition 8 (Probability of X given Y)

Consider the discrete rv X and Y on the previous slide. Then for each $a \in A$, the mapping $\mathbb{P}(X = a \mid Y) : \Omega \to [0, 1]$ is the discrete rv defined by

$$\mathbb{P}(X = a \mid Y)(\omega) = \mathbb{P}(X = a \mid \{Y = Y(\omega)\}).$$

Verification that $\phi(\omega) = \mathbb{P}(X = a \mid Y)(\omega)$ is a discrete rv:

- The set of outcomes/ image space

$$\begin{aligned}\phi(\Omega) &= \cup_{\omega\in\Omega}\{\mathbb{P}(X = a \mid Y = Y(\omega))\} \\ &= \cup_{b\in B}\{\mathbb{P}(X = a \mid Y = b)\} =: C \subset [0,1]\end{aligned}$$

  is countable since B is countable.

- For each $c \in C$, there exists a $b(c) \in B$ such that

$$c = \mathbb{P}(X = a \mid Y = b(c))$$

  and

$$\phi^{-1}(c) = \{\omega \in \Omega \mid Y(\omega) = b(c)\} \in \mathcal{F}.$$

Example 9

Consider the a coin toss $X : \Omega \to \{0, 1\}$ and a die roll $Y : \Omega \to \{1, 2, 3\}$, on sample space $\Omega = \{Heads, Nose, Tails\}$ with

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

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

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 10 (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

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

Note, $\mathbb{E}[X \mid Y]$ is a discrete rv.

Example 11

Consider the Bernoulli rv X, Y with joint probabilities

$$\mathbb{P}(X = i, Y = j) = \begin{bmatrix} 1/8 & 1/4 \\ 1/2 & 1/8 \end{bmatrix} \qquad i, j \in \{0, 1\}$$

$\mathbb{E}[Y \mid X](Heads) =$

Definition 10 (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

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

Note, $\mathbb{E}[X \mid Y]$ is a discrete rv.

Note that,

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

and that it can be associated to a deterministic mapping $g : \mathbb{R}^k \to \mathbb{R}^d$ as follows

$$g(Y(\omega)) = \mathbb{E}[X \mid Y = Y(\omega)]. \tag{3}$$

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

Say you have an observation $Y(\omega)$ (i.e., you know $Y(\omega)$ but not ω), and that what you really seek is the value of $X(\omega)$. Then what is the best function $g(Y(\omega))$ to approximate $X(\omega)$?

Theorem 11 (Mean-square sense best approximation)

For discrete rv $X : \Omega \to A \subset \mathbb{R}^d$ and $Y : \Omega \to B \subset \mathbb{R}^k$ with $\mathbb{E}[X^2] < \infty$, it holds that

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

for all $f : \mathbb{R}^k \to \mathbb{R}^d$ such that $\mathbb{E}\left[|f(Y)|^2\right] < \infty$.

Interpretation Since the constant function $f(Y) = \mathbb{E}[X]$ is one possible mapping, we conclude that

$$\mathbb{E}\left[|X - \mathbb{E}[X \mid Y]|^2\right] \leq \mathbb{E}\left[|X - \mathbb{E}[X]|^2\right].$$

To prove Theorem 11, we will need a few intermediary results.

Lemma 12 (The **tower property**)

For discrete rv $X : \Omega \to A \subset \mathbb{R}^d$ *and* $Y : \Omega \to B \subset \mathbb{R}^k$ *with* $|\mathbb{E}[X]| < \infty$, *it holds that*

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

Proof:

Lemma 13 (The Direct conditioning of expectations)

For the setting in Lemma 12, it holds for any mapping $f : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}$ *such that* $|\mathbb{E}[f(X,Y)]| < \infty$ *that*

$$\mathbb{E}[f(X,Y) \mid Y = b] = \mathbb{E}[f(X,b) \mid Y = b] \quad \forall b \in B.$$

Special case: $f(x,y) = g(x)h(y)$ yields

$$\mathbb{E}[g(X)h(Y) \mid Y = b] = h(b)\mathbb{E}[g(X) \mid Y = b] \quad \forall b \in B$$

Since this holds for all b,

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

And tower property 2:

$$\mathbb{E}[h(Y)\mathbb{E}[g(X) \mid Y]] = \mathbb{E}[h(Y)g(X)] \tag{4}$$

Using (4), let us prove Theorem 11 in the 1D setting, i.e., that

$$\mathbb{E}\left[(X - \mathbb{E}[X \mid Y])^2\right] \le \mathbb{E}\left[(X - f(Y))^2\right]$$

for all $f : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}\left[(f(Y))^2\right] < \infty$.

Proof:

$$\mathbb{E}\left[(X - f(Y))^2\right] = \mathbb{E}\left[\Big((X - \mathbb{E}[X \mid Y]) + (\mathbb{E}[X \mid Y] - f(Y))\Big)^2\right]$$

$$=$$

For $X : \Omega \to A \subset \mathbb{R}^d$ and $Y : \Omega \to B$, the mapping

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

satisfies

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

Concludsion: $\mathbb{E}[X \mid Y]$ is an rv induced from the rv Y through the mapping g.

Question: Is $\mathbb{E}[X \mid Y]$ in some sense unique?

Question: Given a candidate mapping $g : B \to \mathbb{R}^d$, is there a way to verify whether $g(Y) = \mathbb{E}[Y \mid X]$?

Definition 14 ($\mathbb{P}$-almost surely equal)

Two rv X, Y are said to be $\mathbb{P}$-almost surely equal provided

$$\mathbb{P}(\{\omega \in \Omega \mid X(\omega) = Y(\omega)\}) = 1.$$

We write

$$X = Y \quad \mathbb{P} - a.s.$$

(or just "a.s." whenever it is clear which probability measure $\mathbb{P}$ is considered).

Motivation:

Example 15

$X : \Omega \to \{0, 1\}$ and $Y : \Omega \to \{0, 1, 2\}$ with

$$\mathbb{P}(X = Y) = 1 \quad \text{and} \quad \{Y = 2\} \neq \emptyset.$$

Then $X(\omega) \neq Y(\omega)$ for any $\omega \in \{Y = 2\}$, but $X = Y$ a.s.

Theorem 16

Consider the setting in Lemma 12.
If $g : \mathbb{R}^k \to \mathbb{R}^d$ *is a mapping such that for every bounded mapping* $f : \mathbb{R}^k \to \mathbb{R}$,

$$\mathbb{E}\left[f(Y)g(Y) \right] = \mathbb{E}\left[f(Y)X \right] \tag{5}$$

then

$$g(Y) = \mathbb{E}\left[X \mid Y \right] \quad a.s.$$

Interpretation: $\mathbb{E}\left[X \mid Y \right]$ is a a.s. unique rv of form $g(Y)$ satisfying (5).

Usage: If a mapping $B \ni b \mapsto g(b) \in \mathbb{R}^d$ satisfies (5), i.e.,

$$\sum_{b \in B} f(b)g(b)P(Y = b) = \sum_{a \in A, b \in B} f(b)aP(X = a, Y = b) \qquad \forall f : B \to \mathbb{R},$$

then $g(Y(\omega)) = \mathbb{E}\left[X|Y \right](\omega)$ for $\mathbb{P}$-almost all $\omega \in \Omega$.

Next time

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