Mathematics and numerics for data assimilation and state estimation – Lecture 4

UQ Mathematics for Uncertainty Quantification

RWTH AACHEN UNIVERSITY

Summer semester 2020

Summary of lecture 3

- Probability of G given H for events $G, H \in \mathcal{F}$:

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

  where we use the division-by-zero convention $c/0 := 0$ whenever $\mathbb{P}(H) = 0$

- Probability of $X = a$ given Y for rv X, Y:

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

$$\{Y = Y(\omega)\} = \{\beta \in \Omega \mid Y(\beta) = Y(\omega)\}$$
$$\in \mathcal{F}$$

Summary of lecture 3

- Expectation of discrete rv $X : \Omega \to A$ given $H \in \mathcal{F}$:

$$\mathbb{E}\left[X \mid H\right] = \sum_{a \in A} a\mathbb{P}(X = a \mid H) = \frac{\mathbb{E}\left[X \mathbb{1}_H\right]}{\mathbb{P}(H)}$$

- Expectation of X given the rv Y:

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

- Optimal approximation property: Interesting property

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

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

Last slides of lecture 3

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

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

$g : B \to \mathbb{R}$

satisfies

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

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

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

Definition 1 ($\mathbb{P}$-almost surely equal)

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

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

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 2

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

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

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

Theorem 3

Consider discrete rv $X : \Omega \to A \subset \mathbb{R}^d$ *and* $Y : \Omega \to B$. *If* $g : \mathbb{R}^k \to \mathbb{R}^d$ *is a mapping such that for every bounded mapping* $f : \mathbb{R}^k \to \mathbb{R}$,

$$\mathbb{E}\left[f(Y)g(Y)\right] = \mathbb{E}\left[f(Y)X\right] \tag{1}$$

then

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

Interpretation: $\mathbb{E}[X \mid Y]$ is an a.s. unique rv of form $g(Y)$ satisfying (1).

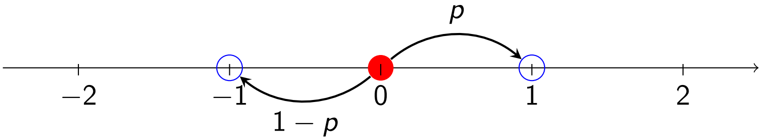
Usage: If a mapping $B \ni b \mapsto g(b) \in \mathbb{R}^d$ satisfies (1), i.e.,

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

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

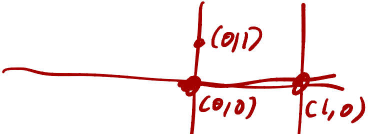
Plan for this lecture

- Properties of Random walks (steps, symmetry, recurrence)



- Convergence of random variables

Random walks



(i,j)

- Are sequences of rv $\{X_n\}$ taking values on the lattice $\mathbb{Z}^d$ for some $d \geq 1$.

- The subindex n can be associated to discrete time, and $\mathbb{Z}^d$ to discrete space (really discrete state-space).

Definition 4 (Random walk (RW))

$X_n : \Omega \to \mathbb{Z}^d$ for $n = 0, 1, \ldots$ is an RW if the sequence of steps $\Delta X_n := X_{n+1} - X_n$ is identically distributed and

$$X_0, \Delta X_1, \Delta X_2, \ldots \text{ are independent.}$$

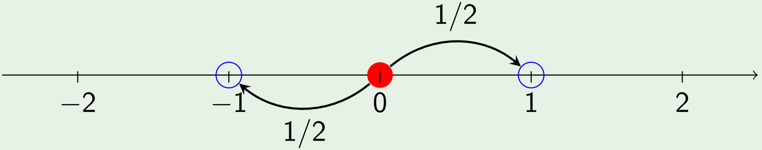
Random walk 2

Since $\{\Delta X_n\}$ are iid, an RW is defined by the two distributions:

- the initial state $\mathbb{P}_{X_0}(z) = \mathbb{P}(X_0 = z)$
- the step $\mathbb{P}_{\Delta X_0}(z) = \mathbb{P}(\Delta X_0 = z)$

Example 5 (Simple and symmetric RW on $\mathbb{Z}^1$)

Let $X_0 = 0$ and $\mathbb{P}(\Delta X_0 = \pm 1) = 1/2$, and **let us compute** $\mathbb{P}(X_n = k)$.



Solution:

Observe that the sequence $Y_k := \mathbb{1}_{\{\Delta X_k = 1\}} \sim \mathit{Bernoulli}(1/2)$ is iid and satisfies

$$\Delta X_k = 2Y_k - 1 = \begin{cases} 1 & \text{if } Y_k = 1 \\ -1 & \text{if } Y_k = 0 \end{cases}$$

Consequently,

$$X_n = \underbrace{X_0}_{=0} + \sum_{k=0}^{n-1} \Delta X_k = \sum_{k=0}^{n-1} (2Y_k - 1)$$

$$= 2\sum_{k=0}^{n-1} Y_k - n =: 2M_n - n$$

$$M_n = \sum_{k=0}^{n-1} Y_k \sim \mathrm{Binom}(n, \tfrac{1}{2})$$ [lecture 2]

$$\mathbb{P}(X_n = k) = \mathbb{P}(2M_n - n = k)$$

$$= \mathbb{P}\left(M_n = \frac{n+k}{2}\right) = \begin{cases} 2^{-n} \binom{n}{(n+k)/2} & \text{if } n+k \text{ is even and } |k| \le n \\ 0 & \end{cases}$$

What if X_0 is an rv with $\mathbb{P}(X_0 = k) > 0$ for multiple k?

Basically

$$\mathbb{P}(X_n = j) = \sum_{k \in \mathbb{Z}} \mathbb{P}(X_n = j \mid X_0 = k) \cdot \mathbb{P}(X_0 = k)$$

Symmetric random walks

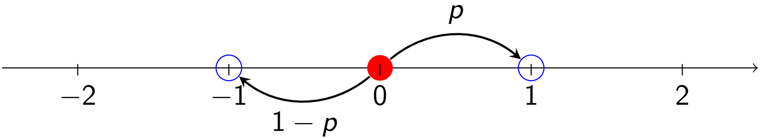
For rv X and Y, we introduce notation $X \stackrel{D}{=} Y$ to say that X and Y are identically distributed.

Definition 6 (Symmetric RW)

An RW on $\mathbb{Z}^d$ is called symmetric if the step and the "reverse step" are identically distributed, meaning

$$X_1 - X_0 \stackrel{D}{=} X_0 - X_1.$$

Intuition: Equally likely to step in opposite directions.



The above RW symmetric if and only if $p = 1/2$.

Simple RW

$$\pm\begin{bmatrix}1\\0\end{bmatrix} \pm\begin{bmatrix}0\\1\end{bmatrix}$$

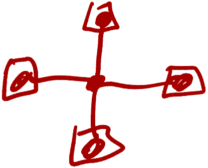
Definition 7 (Simple RW)

An RW on $\mathbb{Z}^d$ is called **simple** if the values of the step ΔX_0 belong to the set $\{e_k\}_{k=1}^d$ of canonical basis vectors in $\mathbb{R}^d$. In other words,

$$\{X_n\} \text{ is simple} \iff \mathbb{P}(|\Delta X_0| = 1) = 1.$$

Furthermore, an RW is called **simple symmetric** if

$$\mathbb{P}(\Delta X_0 = e_k) = \mathbb{P}(\Delta X_0 = -e_k) = \frac{1}{2d}, \quad k = 1, 2, \ldots, d.$$



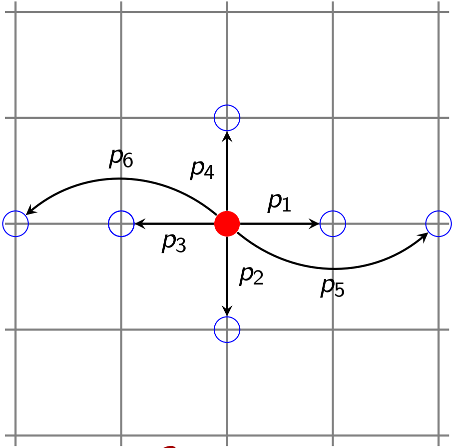
Example

Consider RW with steps satisfying

$$\sum_{i=1}^{6} p_i = 1.$$

Constraints for the RW being

- symmetric?
- simple?
- simple symmetric?



Symmetric:

$P_1 = P_3, \; P_2 = P_4, \; P_5 = P_6$

Simple: $P_5 = 0, \; P_6 = 0$

& $\sum_{i=1}^{4} P_i = 1$

simple symmetric: $P_5 = P_6 = 0$

$P_k = \frac{1}{4} \quad k = 1, 2, \dots, 4.$

Matlab implementation of simple symmetric RW on $\mathbb{Z}^2$

Core idea $X_{n+1} = X_n + \Delta X_n$ where

$$\mathbb{P}(\Delta X_n = \pm e_1) = \mathbb{P}(\Delta X_n = \pm e_2) = 1/4.$$

Use **randi(4)** in matlab to draw random integer in $[1, 4]$, all integers with same probability, and assign walk direction from drawn integer.

```
% Four step directions
Step = [-1 0;1 0;0 1;0 -1];
X = [0,0];
for n =1:200
    direction = randi(4);
    dX = Step(direction,:);
    plot([X(1) X(1)+dX(1)], [X(2), X(2)+dX(2)])
    X = X+dX;
end
```

See **randWalk2d.m** for more details.

Recurrence and transience

Definition 8

An RW on $\mathbb{Z}^d$ with is **recurrent** if it (over its whole path $\{X_n\}_{n\in\mathbb{N}}$) visits its initial state infinitely often $\mathbb{P}-$almost surely, and **transient** otherwise (i.e., if it visits its initial state only a finite number of times $\mathbb{P}-$almost surely).

- Description of a quasi-stable property: assume you are gambling, you win with probability $\mathbb{P}(\Delta X_n = 1) = p$ lose with $\mathbb{P}(\Delta X_n = -1) = 1 - p$. Unless $p = 1/2$, $\{X_n\}$ is transient.
- Recurrence is a form of quasi-periodic behavior. In some settings (but not for RW) it connects spatial distribution of limit processes and time-averages over path realizations

$$\mathbb{P}(X_\infty = y) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} \mathbb{1}_{X_n=y}.$$

Theorem 9

Consider an RW on $\mathbb{Z}^d$ *with* $X_0 = 0$ *and let*

$T : \Omega \to \mathbb{N} \cup \{\infty\}$

$$T := \inf\{n \geq 1 \mid X_n = 0\}$$

with the convention that $\inf \emptyset := \infty$ *and*

$$N := \sum_{n \in \mathbb{N}} \mathbb{1}_{X_n = 0} \quad \textit{(total visits of origin)}$$

Then $\{X_n\}$ *is recurrent if and only if* $\lambda := \mathbb{P}(T < \infty) = 1$ *and for* $j \in \mathbb{N} \cup \{\infty\}$,

$$\mathbb{P}(N = j) = \begin{cases} (1-\lambda)\lambda^{j-1} & \textit{if } \lambda < 1 \\ \mathbb{1}_{j=\infty} & \textit{if } \lambda = 1 \end{cases}$$

Note that $N : \Omega \to \mathbb{N} \cup \{\infty\}$.

Proof of Theorem 9

Define $\tau_0 = 0$, and

$$\tau_{k+1} = \inf\{n > \tau_k \mid X_n = 0\} \quad \text{for} \quad k = 0, 1, \ldots$$

Note that $\Delta\tau_k = \tau_{k+1} - \tau_k$ is a sequence of independent and T-distributed rv.

Introducing the rv

$$\bar{k} = \sup\{k \geq 0 \mid \tau_k < \infty\},$$

we can write

$$N = \sum_{n=0}^{\infty} \mathbb{1}_{X_n=0} = \sum_{k=0}^{\bar{k}} \mathbb{1}_{X_{\tau_k}=0} = \bar{k} + 1.$$

Observe that for $j \in \mathbb{N}$

$$\mathbb{P}(\bar{k} = j) = \mathbb{P}(\Delta\tau_0 < \infty, \ldots, \Delta\tau_{j-1} < \infty, \Delta\tau_j = \infty)$$

$$\overset{\perp}{=} \prod_{r=0}^{j-1} \mathbb{P}(\Delta\tau_r < \infty) \cdot \mathbb{P}(\Delta\tau_j = \infty)$$

$$= \lambda^j (1 - \lambda)$$

$\bar{k} = j \Rightarrow \tau_{j+1} = \infty$ & $\tau_r < \infty \ \forall r < j$.

and for $j = \infty$,

$$\mathbb{P}(\bar{K} = \infty) = \lambda^{\infty}$$

We obtain that

$$\mathbb{P}(N = j) = \mathbb{P}(\bar{K} + 1 = j)$$

$$= \mathbb{P}(\bar{K} = j - 1)$$

$$= \begin{cases} \lambda^{j-1}(1-\lambda) & \text{if } \lambda < 1 \\ \mathbb{1}_{j=\infty} & \text{if } \lambda = 1 \end{cases}$$

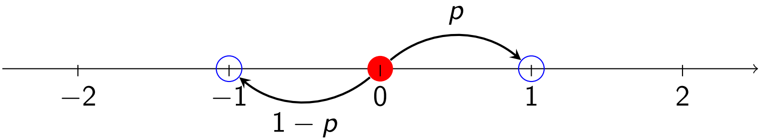
for $j \in \mathbb{N} \cup \{\infty\}$.

Which RW are recurrent?

- **(Related to FJK 2.1.13)** Symmetric and simple RW on $\mathbb{Z}^d$ are recurrent if $d \leq 2$ and transient otherwise.

> A drunk man will eventually find his way home, but a drunk bird may get lost forever
>
> *Shizuo Kakutani*

- **(Related to FJK 2.1.14)** Non-symmetric RW are always transient.



Always transient when $p \neq 1/2$.

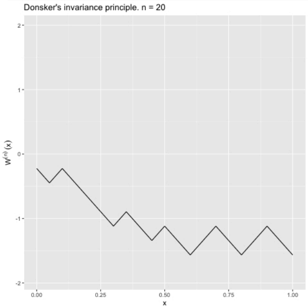
Scaling property of RW

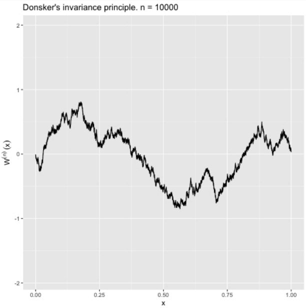
Theorem 10 (Random walk case of Donsker's theorem)

Let $\{X_n\}$ be a simple symmetric RW on $\mathbb{Z}$ with $X_0 = 0$ and consider

$$W^{(n)}(t) := \frac{X_{\lfloor nt \rfloor}}{\sqrt{n}} \quad t \in [0,1],$$

where $\lfloor x \rfloor := \max\{k \in \mathbb{Z} \mid k \leq x\}$. Then $\{W^{(n)}(t)\}_{t\in[0,1]}$ converges in distribution to a standard Brownian motion $\{W(t)\}_{t\in[0,1]}$.





Convergence of random variables

Assume you can draw iid samples $X_k \sim \mathbb{P}_X$ and that you approximate $\mu = \mathbb{E}[X]$ by the sample average

$$\bar{X}_M := \frac{1}{M} \sum_{k=1}^{M} X_k. \tag{2}$$

Questions:

- Will $\bar{X}_M \to \mu$ as $M \to \infty$, and, if so, in what sense?
- Is there a convergence rate of the form

  $$\|\bar{X}_M - \mu\| \leq \frac{C}{M^\beta}$$

  for some norm $\|\cdot\|$ and some rate $\beta > 0$?

Mean-square convergence

- For rv $Y, Z : \Omega \to \mathbb{R}^d$ we introduce the scalar product

$$\langle Y, Z \rangle_{L^2(\Omega)} := \mathbb{E}[\, Y \cdot Z]$$

- the function space

$$L^2(\Omega) := \{\mathcal{F} - \text{measurable mappings } Y : \Omega \to \mathbb{R}^d \mid \mathbb{E}\left[\,|Y|^2\right] < \infty\}$$

with norm

$$\|Y\|_{L^2(\Omega)} := \sqrt{\mathbb{E}[\,|Y|^2]},$$

is a Hilbert space.

- The notation is shorthand for $L^2(\Omega) = L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$.

Returning to the approximation

$$\bar{X}_M = \frac{1}{M} \sum_{k=1}^{M} X_k$$

- Since $\mathbb{E}[X_k] = \mu$, it holds that

$$\bar{X}_M - \mu = \sum_{k=1}^{M} \frac{X_k - \mu}{M}$$

- Since $\{X_k - \mu\}$ is a mean-zero and independent sequence of rv, it holds for $j \neq k$ that

$$\begin{aligned}
\langle X_k - \mu, X_j - \mu \rangle_{L^2(\Omega)} &= \mathbb{E}[(X_k - \mu) \cdot (X_j - \mu)] \\
&= \sum_{(x_k, x_j) \in A \times A} (x_k - \mu) \cdot (x_j - \mu) \underbrace{\mathbb{P}(X_k = x_k, X_j = x_j)}_{= \mathbb{P}(X_k = x_k)\mathbb{P}(X_j = x_j)} \\
&= \mathbb{E}[(X_k - \mu)] \cdot \mathbb{E}[(X_j - \mu)] = 0
\end{aligned}$$

(Here we assumed discrete rv $X_k : \Omega \to A$, but it also holds for continuous rv.)

This yields

$$\|\bar{X}_M - \mu\|^2_{L^2(\Omega)} = \left\langle \sum_{k=1}^{M} \frac{X_k - \mu}{M}, \sum_{k=1}^{M} \frac{X_k - \mu}{M} \right\rangle$$

$$= \sum_{j,k=1}^{M} \frac{1}{M^2} \langle \bar{X}_j - \mu, \bar{X}_k - \mu \rangle =$$

$$= \sum_{k=1}^{M} \frac{\langle \bar{X}_k - \mu, \bar{X}_k - \mu \rangle}{M^2}$$

$$= \frac{\|\bar{X} - \mu\|^2_{L^2(\Omega)}}{M}$$

Conclusion: For a sequence of d-dimensional discrete independent rv $X_i \sim \mathbb{P}_X$,

$$\|\bar{X}_M - \mu\|_{L^2(\Omega)} = \frac{\|X - \mu\|_{L^2(\Omega)}}{\sqrt{M}}, \tag{3}$$

i.e., the mean-square convergence rate is $1/2$.

Weaker form of convergence

Definition 11 (Convergence in probability)

A sequence of rv $\{\bar{Y}_k\}$ converges in probability towards the rv Y if for all $\epsilon > 0$,

$$\lim_{k\to\infty} \mathbb{P}(|Y_k - Y| > \epsilon) = 0.$$

Theorem 12 (Weak law of large numbers (Durrett 2.2.14))

For a sequence of d-dimensional independent rv $X_i \sim \mathbb{P}_X$ with $\mathbb{E}\left[|X_i|\right] < \infty$ it holds that

$$\bar{X}_M \to \mu \quad \text{in probability.}$$

Chebychev's inequality

To prove the theorem, we will apply **Chebychev's inequality**: for any rv Y with $\bar{\mu} = \mathbb{E}[Y]$

$$\mathbb{P}(|Y - \bar{\mu}| > \epsilon) \leq \mathbb{E}\left[\frac{|Y - \bar{\mu}|^2}{\epsilon^2}\right]$$

Verification:

$$\mathbb{P}(|Y - \bar{\mu}| > \epsilon) = \mathbb{E}\left[\mathbb{1}_{\{|Y - \bar{\mu}| > \epsilon\}}\right] \leq \mathbb{E}\left[\frac{|Y - \bar{\mu}|^2}{\epsilon^2}\right]$$

Proof of Theorem 12 (Assuming $\mathbb{E}[|X_i|^2] < \infty$.)

$$\mathbb{P}(|\bar{X}_M - \mu| > \epsilon) \leq \mathbb{E}\left[\frac{|\bar{X}_M - \mu|^2}{\epsilon^2}\right]$$

$$\leq \frac{\|X - \mu\|_{L^2}^2}{M\epsilon^2} \to 0 \text{ as } M \to \infty$$

Next time

Discrete time and space Markov Chains

Caption: Quantum Cloud, designed by Antony Gormley. Random walk algorithm starting from points on the surface of an enlarged figure based on Gormley's body.