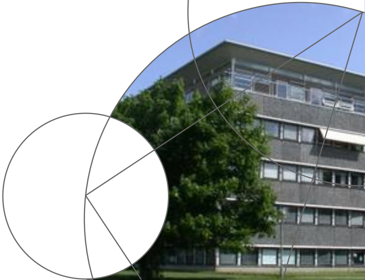
Teaching complicated mathematics at the university

Anders Rønn-Nielsen
Department of Mathematical Sciences, KU



Overview of the talk

1. Type and aim of courses
2. General approach
3. Activation of students
4. Conclusion

Type of courses

Theoretical courses in probability theory (VidSand2 and Beting):

- Decomposition of measures
- Conditional expectations and conditional distributions
- Martingale theory: Optional sampling, convergence, CLT
- Brownian motion
- Conditional independence
- Markov chains: Markov property, ergodic theorem

Aim of courses

From course descriptions:

- "describe and prove the results on decomposition of signed measures"
- "derive and describe the main results on martingales"
- "discuss the relation between decomposition of measures and conditional expectations"
- "Give an oral presentation of a specific topic within the theory covered by the course"
- "Discuss the relation between conditional expectations and conditional distributions"

Aim of courses – continued

Theory and proofs play an important part of the courses:

- Adds to the deep understanding of the subject
- Proof techniques is a useful skill that has to be learned
- Simply very beautiful mathematics – students should see at least some of it!

General approach to lectures

Mix of blackboard and slides/computer

- Use slides to present definitions, theorems and small examples: Saves some time compared to blackboard
- Use blackboard to proofs and longer examples – the lecture is slowed down: Arguments are easier to follow
- Use computer simulations, when something can be illustrated (and I have the time)

General approach to lectures – continued

Use of enthusiasm and energy:

- As a lecturer, you can use energy and enthusiasm to keep students awake and interested.
- In particular variation in enthusiasm: Think about what is the most interesting/fascinating part of the lecture?

Activation of students

Reasons to activate students during a lecture:

- Hard to stay concentrated and focused for 45 minutes
- Learning is easier if the students can relate to the problem. E.g think about small problems by themselves.
- Deeper learning if the students are part of the development process (or feel they are)
- Creates interaction between students and lecturer. Makes adjustments easier.

Think about:

- Should students be forced to be active?

Downside

- You put yourself in a vulnerable situation – socially awkward to ask a question and not get an answer. Much easier not to involve the students.
- Student activities take up much time: Just explaining something would be much faster.
- Student activities make planning more difficult:
    - You don't know how much time an activity takes.
    - If you realise that something was not understood, then you have to react.

Questions

A simple activity is just to ask a lot of questions.

Problems:

- At most one student answers
- Panic among students!
- Awkward if no-one answers

Instead:

- Use of slides with questions and exercises (examples will follow)
- Give the students time to think and discuss.
- If many students: Make them vote.

Activities developing intuition

Working with the intuition can be useful:

- Good as introduction: Easier to understand complicated theory if you have some intuition
- Helpful to have an intuitive reference later on, when the result is used again

Challenge:

- My intuition may be different from the way the students think
- That leads to confusion

Extension of Tonelli's Theorem

Theorem (Theorem 1.3.2, Extended Tonelli)

Let μ be a probability measure on $(\mathcal{X}, \mathbb{E})$, and assume that $(P_x)_{x \in \mathcal{X}}$ is a $(\mathcal{X}, \mathbb{E})$–Markov kernel on $(\mathcal{Y}, \mathbb{K})$. Let λ be the integration of $(P_x)_{x \in \mathcal{X}}$ with respect to μ. For every $\mathbb{E} \otimes \mathbb{K}$–measurable function $f : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ it holds that

$$\int f(x,y) \, \mathrm{d}\lambda(x,y) = \iint f(x,y) \, \mathrm{d}P_x(y) \, \mathrm{d}\mu(x)$$

Extension of Tonelli's Theorem

Theorem (Theorem 1.3.2, Extended Tonelli)

Let μ be a probability measure on $(\mathcal{X}, \mathbb{E})$, and assume that $(P_x)_{x \in \mathcal{X}}$ is a $(\mathcal{X}, \mathbb{E})$–Markov kernel on $(\mathcal{Y}, \mathbb{K})$. Let λ be the integration of $(P_x)_{x \in \mathcal{X}}$ with respect to μ. For every $\mathbb{E} \otimes \mathbb{K}$–measurable function $f : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ it holds that

$$\int f(x,y)\, \mathrm{d}\lambda(x,y) = \iint f(x,y)\, \mathrm{d}P_x(y)\, \mathrm{d}\mu(x)$$

Discussion: Compare with example 1.2.3 and discuss why this theorem is called an extension of Tonelli!

Extension of Tonelli's Theorem

Theorem (Theorem 1.3.2, Extended Tonelli)

Let μ be a probability measure on $(\mathcal{X},\mathbb{E})$, and assume that $(P_x)_{x\in\mathcal{X}}$ is a $(\mathcal{X},\mathbb{E})$–Markov kernel on $(\mathcal{Y},\mathbb{K})$. Let λ be the integration of $(P_x)_{x\in\mathcal{X}}$ with respect to μ. For every $\mathbb{E}\otimes\mathbb{K}$–measurable function $f:\mathcal{X}\times\mathcal{Y}\to[0,\infty]$ it holds that

$$\int f(x,y)\,\mathrm{d}\lambda(x,y) = \iint f(x,y)\,\mathrm{d}P_x(y)\,\mathrm{d}\mu(x)$$

Discussion: Compare with example 1.2.3 and discuss why this theorem is called an extension of Tonelli!

Measures like λ can be seen as a generalisation of product measures $\mu\otimes\nu$ on $\mathcal{X}\times\mathcal{Y}$.

We will show that

$$\frac{1}{\sqrt{n}} M_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

We will show that

$$\frac{1}{\sqrt{n}} M_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Discuss in 2 minutes:

- What will a plot of $(\frac{1}{n} M_n)_{n \geq 1}$ look like?
- What will a plot of $(\frac{1}{\sqrt{n}} M_n)_{n \geq 1}$ look like?
- What will a plot of $(M_n)_{n \geq 1}$ look like?

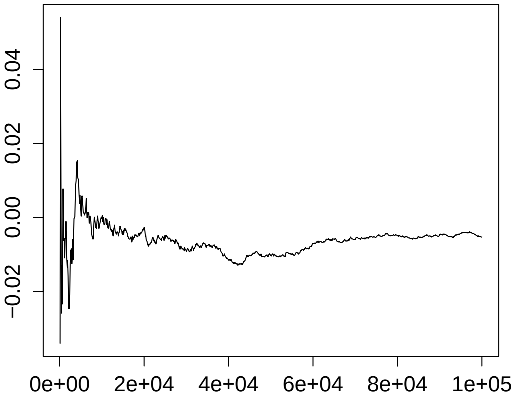


Figure: Simulation of $(\frac{1}{n}M_n)_{n\geq 1}$

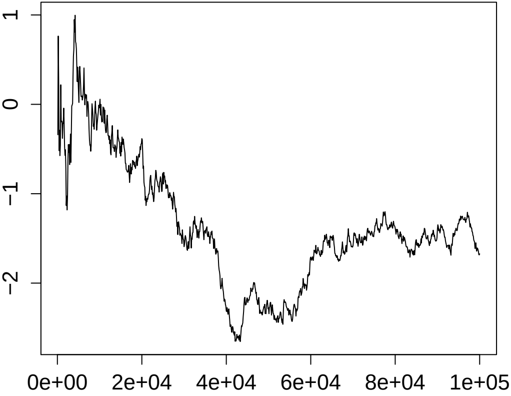


Figure: Simulation of $(\frac{1}{\sqrt{n}} M_n)_{n \geq 1}$

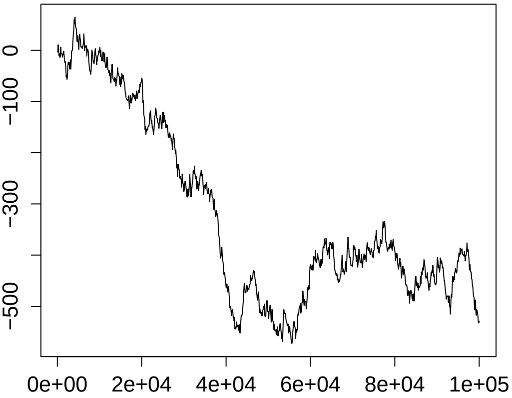
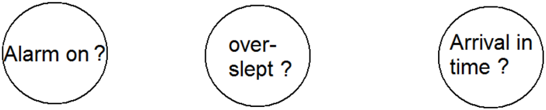


Figure: Simulation of $(M_n)_{n\geq 1}$

Discussion about your morning situation

Consider the following 3 events you could experience in the morning



- Do we have ordinary independence here?
- Try to identify a conditional independence between two events given a third
- Explain this using the asymmetric theorem

Relating to and combining known theory

Student activities can be used to remind students of material they (in principle) already know:

- You get to check whether the students have understood previous lectures
- Repeating helps understanding
- Remind students of some known theory that is about to be applied

Discussion – Memory game:-)

Consider the Markov kernels

$$1: \quad P^{*k} \qquad \text{and} \qquad 2: \qquad \mathcal{P}^k.$$

Combine them with the following conditional distributions, where $X_0, X_1, \ldots$ is a time homogeneous Markov chain with transition prob. $(P_x)_{x \in \mathcal{X}}$:

a) $(X_1, \ldots, X_k) \,|\, X_0$

b) $X_{n+k} \,|\, X_n$

c) $(X_n, \ldots, X_{n+k-1}) \,|\, X_{n-1}$

d) $X_k \,|\, X_0$

e) $X_{n+k} \,|\, X_0, \ldots, X_n$

f) $(X_n, \ldots, X_{n+k-1}) \,|\, X_0, \ldots, X_{n-1}$

Markov chains so far

Think about the arguments and the order:

- we defined a Markov chain
- we derived the finite–dimensional distributions for a Markov chain (if it exists)
- we argued that a probability measure exists on $(\mathcal{X}^\infty, \mathbb{E}^\infty)$ with finite–dimensional distributions like this
- we went back and saw that a process with this distribution IS a Markov chain

Think about: Could we have changed the order?

Involvement in proofs

- A proof could easily last 45 minutes: Breaks may be needed to avoid one–man lecturer shows and sleeping students!
- Can use activities to point out/repeat an important step - change of speed.
- Although the proof takes place on the blackboard, I (again) use slides to formulate questions/exercises.

First steps of proof of 5.5.1

Let $B = \{x \in \mathcal{X} \ : \ V(x) \leq r\}$ and

$$h(y) = \inf_{x \in B} k_x^{(m)}(y)$$

First steps of proof of 5.5.1

Let $B = \{x \in \mathcal{X} \ : \ V(x) \leq r\}$ and

$$h(y) = \inf_{x \in B} k_x^{(m)}(y)$$

Then we have shown

$$P^{*(m+n)}(f)(y) \geq h(y) \int_B P^{*n}(f)(x) \, \mathrm{d}\nu(x)$$

First steps of proof of 5.5.1

Let $B = \{x \in \mathcal{X} \ : \ V(x) \leq r\}$ and

$$h(y) = \inf_{x \in B} k_x^{(m)}(y)$$

Then we have shown

$$P^{*(m+n)}(f)(y) \geq h(y) \int_B P^{*n}(f)(x) \, \mathrm{d}\nu(x)$$

Suppose we furthermore can show that for some $\delta > 0$

$$\int_B P^{*n}(f)(x) \, \mathrm{d}\nu(x) > \delta$$

for n big enough.

First steps of proof of 5.5.1

Let $B = \{x \in \mathcal{X} \;:\; V(x) \leq r\}$ and

$$h(y) = \inf_{x \in B} k_x^{(m)}(y)$$

Then we have shown

$$P^{*(m+n)}(f)(y) \geq h(y) \int_B P^{*n}(f)(x)\, \mathrm{d}\nu(x)$$

Suppose we furthermore can show that for some $\delta > 0$

$$\int_B P^{*n}(f)(x)\, \mathrm{d}\nu(x) > \delta$$

for n big enough.

Discuss: Why will the proof in that case be complete?

Theorem (Theorem 4.3.5)

A time homogeneous Markov chain $X_0, X_1, \dots$ is weakly time homogeneous.

Theorem (Theorem 4.3.5)

A time homogeneous Markov chain $X_0, X_1, \ldots$ is weakly time homogeneous.

Proof.

On the blackboard! □

Theorem (Theorem 4.3.5)

A time homogeneous Markov chain $X_0, X_1, \ldots$ is weakly time homogeneous.

Proof.

On the blackboard! □

Discuss: Why can we conclude from the proof that

$$X_{\tau+1} \,|\, X_\tau \quad \overset{\mathcal{D}}{=} \quad X_{n+1} \,|\, X_n \ ?$$

Theorem (Theorem 3.4.3)

Let $\mathbb{A}$, $\mathbb{B}$, $\mathbb{G}$ and $\mathbb{H}$ be σ-algebras. It holds that

$$\mathbb{A} \perp\!\!\!\perp \mathbb{B} \,|\, \mathbb{H} \text{ and } \mathbb{A} \perp\!\!\!\perp \mathbb{G} \,|\, \mathbb{B} \vee \mathbb{H} \quad \Rightarrow \quad \mathbb{A} \perp\!\!\!\perp (\mathbb{B} \vee \mathbb{G}) \,|\, \mathbb{H}.$$

Theorem (Theorem 3.4.3)

Let $\mathbb{A}$, $\mathbb{B}$, $\mathbb{G}$ *and* $\mathbb{H}$ *be* σ*-algebras. It holds that*

$$\mathbb{A} \perp\!\!\!\perp \mathbb{B} \mid \mathbb{H} \text{ and } \mathbb{A} \perp\!\!\!\perp \mathbb{G} \mid \mathbb{B} \vee \mathbb{H} \quad \Rightarrow \quad \mathbb{A} \perp\!\!\!\perp (\mathbb{B} \vee \mathbb{G}) \mid \mathbb{H}.$$

Proof.

We show the asymmetric characterisation: Let $A \in \mathbb{A}$, then

$$P\big(A \mid (\mathbb{B} \vee \mathbb{G}) \vee \mathbb{H}\big) = P(A \mid \mathbb{B} \vee \mathbb{H}) = P(A \mid \mathbb{H}) \text{ a.s.}$$

Discuss: How is this equality obtained? □

Exercises

Activities could be small exercises:

- A way to give examples of applications of the theory
- As a lecturer you can check, whether a definition or result is understood

Potential problem:

- Party–killer if the exercise is too hard
- The exercise can/should be VERY simple: Should be possible to solve for most students within 2-3 minutes
- If the answer is very simple, the students should answer by voting

3–minute exercise:

Assume that $(X_n, \mathcal{F}_n)$ is adapted and that τ is a stopping time.

- Is X_4 $\mathcal{F}_6$–measurable?
- Is X_6 $\mathcal{F}_4$–measurable?
- Is $X_{\tau+1}$ $\mathcal{F}_\tau$–measurable?

3–minute exercise:

Assume that $(X_n, \mathcal{F}_n)$ is adapted and that τ is a stopping time.

- Is X_4 $\mathcal{F}_6$–measurable? Yes!
- Is X_6 $\mathcal{F}_4$–measurable?
- Is $X_{\tau+1}$ $\mathcal{F}_\tau$–measurable?

3–minute exercise:

Assume that $(X_n, \mathcal{F}_n)$ is adapted and that τ is a stopping time.

- Is X_4 $\mathcal{F}_6$–measurable? Yes!
- Is X_6 $\mathcal{F}_4$–measurable? No! (=not necessarily...)
- Is $X_{\tau+1}$ $\mathcal{F}_\tau$–measurable?

3–minute exercise:

Assume that $(X_n, \mathcal{F}_n)$ is adapted and that τ is a stopping time.

- Is X_4 $\mathcal{F}_6$–measurable? Yes!
- Is X_6 $\mathcal{F}_4$–measurable? No! (=not necessarily...)
- Is $X_{\tau+1}$ $\mathcal{F}_\tau$–measurable? No! (=not necessarily...)

Example (/Exercise Continued)

Explain the calculation

$$E(S_{n+1}|\mathcal{F}_n) = S_n + E(Y_{n+1}) = S_n + (2p-1)$$

and decide the type of $(S_n, \mathcal{F}_n)$:

- ?-martingale, if $p > \frac{1}{2}$
- ?-martingale, if $p = \frac{1}{2}$
- ?-martingale, if $p < \frac{1}{2}$

Example (/Exercise Continued)

Explain the calculation

$$E(S_{n+1}|\mathcal{F}_n) = S_n + E(Y_{n+1}) = S_n + (2p-1)$$

and decide the type of $(S_n, \mathcal{F}_n)$:

- ?-martingale, if $p > \frac{1}{2}$ (subMG)
- ?-martingale, if $p = \frac{1}{2}$
- ?-martingale, if $p < \frac{1}{2}$

Example (/Exercise Continued)

Explain the calculation

$$E(S_{n+1}|\mathcal{F}_n) = S_n + E(Y_{n+1}) = S_n + (2p-1)$$

and decide the type of $(S_n, \mathcal{F}_n)$:

- ?-martingale, if $p > \frac{1}{2}$ (subMG)
- ?-martingale, if $p = \frac{1}{2}$ (MG)
- ?-martingale, if $p < \frac{1}{2}$

Example (/Exercise Continued)

Explain the calculation

$$E(S_{n+1}|\mathcal{F}_n) = S_n + E(Y_{n+1}) = S_n + (2p-1)$$

and decide the type of $(S_n, \mathcal{F}_n)$:

- ?-martingale, if $p > \frac{1}{2}$ (subMG)
- ?-martingale, if $p = \frac{1}{2}$ (MG)
- ?-martingale, if $p < \frac{1}{2}$ (superMG)

Example 5.5.3: ARCH(1)–process

Define $X_0, X_1, X_2, \dots$ recursively by

$$X_{n+1} = \sqrt{(\gamma + \alpha X_n^2)}\epsilon_{n+1},$$

where (again) $\epsilon_1, \epsilon_2, \dots$ are iid $\mathcal{N}(0,1)$ (independent of X_0).

Example 5.5.3: ARCH(1)–process

Define $X_0, X_1, X_2, \ldots$ recursively by

$$X_{n+1} = \sqrt{(\gamma + \alpha X_n^2)}\epsilon_{n+1},$$

where (again) $\epsilon_1, \epsilon_2, \ldots$ are iid $\mathcal{N}(0,1)$ (independent of X_0).

Discuss:

- Why is this a Markov chain?
- How do you expect a simulation of $X_0, X_1, X_2, \ldots$ to look like?
- Which values (big/small) of α and γ do you think will make the Markov chain asymptotically stable?

Final comments – does it make a difference?

- No proof that it actually makes a difference, but I believe so...
- Students have seemed happy.
- Lecturing is more fun this way.
- Does not work without nice students:-)

Final comments – does it make a difference?

- No proof that it actually makes a difference, but I believe so...
- Students have seemed happy.
- Lecturing is more fun this way.
- Does not work without nice students:-)

Thank you!