Seminar on Probability and Computer Science

An Introduction to Martingales and Their Application

Nils Müller

Bochum, August 1, 2020

Table of Contents

Conditional Expectations and Martingales

Conditional Expectations

Definition (Conditional Expectation)
Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra over Ω. The conditional expectation of X with respect to $\mathcal{F}$ is a $\mathcal{F}$-$\mathcal{B}(\mathbb{R})$-measurable random variable $X^{\mathcal{F}} : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for which we have

$$\int_F X \, \mathrm{d}\mathbb{P} = \int_F X^{\mathcal{F}} \, \mathrm{d}\mathbb{P}_{|\mathcal{F}} \text{ for all } F \in \mathcal{F}. \tag{CE}$$

Remark
Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ *and* $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ *measurable. We will use the abbreviations* $\mathbb{E}(X \mid Y) := \mathbb{E}(X \mid Y^{-1}(\mathcal{E}))) := X^{Y^{-1}(\mathcal{E})}$.

Existence & Uniqueness of Conditional Expectations

Theorem (Existence & Uniqueness of Conditional Expectations)
Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ *and* $\mathcal{F} \subseteq \mathcal{A}$ *be a* σ*-algebra over* Ω *.*

(a) If $X^{\mathcal{F}}, Y^{\mathcal{F}} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ *satisfy*

$$\mathbb{E}(X \mathbb{1}_F) = \mathbb{E}(X^{\mathcal{F}} \mathbb{1}_F) = \mathbb{E}(Y^{\mathcal{F}} \mathbb{1}_F) \text{ for all } F \in \mathcal{F},$$

then $X^{\mathcal{F}} = Y^{\mathcal{F}}$ $\mathbb{P}$*-almost surely.*

(b) A random variable $X^{\mathcal{F}} : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *fulfilling* (CE) *exists.*

Proof Sketch.

- The positive and negative part of X are densities with respect to $\mathbb{P}$ and thus each induce a measure on $(\Omega, \mathcal{F})$.
- As both of these measures are absolutely continuous with respect to $\mathbb{P}_{|\mathcal{F}}$, by the Radon-Nikodym-Theorem there exist densities with respect to $\mathbb{P}_{|\mathcal{F}}$.
- Subtracting these densities, one obtains $X^{\mathcal{F}}$.
- $X^{\mathcal{F}}$ is unique as any absolute difference of two candidates is integrable and must almost surely be zero due to (CE) .

□

Definition (Discrete Filtration)
Let $(\Omega, \mathcal{A})$ be a measurable space. A sequence $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$ of sub σ-fields of $\mathcal{A}$ over Ω is called a filtration over $(\Omega, \mathcal{A})$.

Definition (Discrete Martingale)
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(\mathcal{F}_n)_{n \in \mathbb{N}}$ be a filtration over $(\Omega, \mathcal{A})$. The random variables $(X_n)_{n \in \mathbb{N}}$ are called a martingale with respect to $(\mathcal{F}_n)_{n \in \mathbb{N}}$ $:\Longleftrightarrow$

(a) $X_n \in L^1(\Omega, \mathcal{F}_n, \mathbb{P})$ and

(b) $\mathbb{E}\,(X_{n+1} \mid \mathcal{F}_n) = X_n$ $\mathbb{P}$-almost surely for all $n \in \mathbb{N}$.

Example (Gambler's Ruin 1)

Let $(Z_n)_{n\in\mathbb{N}}$ be i.i.d. random variables such that the associated probability measure is

$$\mathbb{P}_{Z_1} = \frac{\delta_1 + \delta_{-1}}{2}\,.$$

We define

$$X_n := \sum_{j=1}^{n} Z_j \text{ for all } n \in \mathbb{N}\,.$$

$(X_n)_{n\in\mathbb{N}}$ is a martingale with respect to σ-algebra $\mathcal{F}_n := \mathcal{I}(Z_1, \ldots, Z_n)$, as

(a) $\mathbb{E}|X_n| = n\mathbb{E}|Z_1| = n$ and X_n is $\mathcal{F}_n$-$\mathcal{B}(\mathbb{R})$-measurable and

(b) $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) + X_n = \mathbb{E}Z_{n+1} + X_n = X_n\,.$

A Simple Martingale

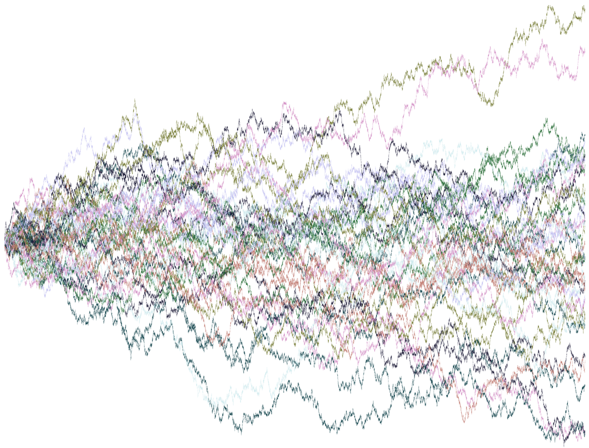

Figure 1: A realization of 50 independent copies of $(X_k)_{k\leq n}$, where $n = 7 \cdot 10^4$.

Stopping Times

Lemma (Stopping Lemma)
For a martingale $(X_n)_{n\in\mathbb{N}}$ *with respect to filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *we have*

$$\mathbb{E}X_n = \mathbb{E}X_1 \text{ for all } n \in \mathbb{N}\,.$$

Proof.
Due to property (b) of martingales we have

$$\begin{aligned} & \mathbb{E}\left(X_{n+1} \mid \mathcal{F}_n\right) = X_n && \mathbb{P}\text{-almost surely for all } n \in \mathbb{N} \\ \Longrightarrow\ & \mathbb{E}\left(\mathbb{E}\left(X_{n+1} \mid \mathcal{F}_n\right)\right) = \mathbb{E}X_n && \text{for all } n \in \mathbb{N} \\ \Longleftrightarrow\ & \mathbb{E}X_{n+1} = \mathbb{E}X_n && \text{for all } n \in \mathbb{N}\,. \end{aligned}$$

□

Example (Gambler's Ruin 2)
We consider the simple example of a martingale we encountered in the previous section – the random walk $(X_n)_{n\in\mathbb{N}}$. By the stopping lemma we conclude

$$\mathbb{E}X_n = \mathbb{E}X_1 = \mathbb{E}Z_1 = 0 \text{ for all } n \in \mathbb{N}.$$

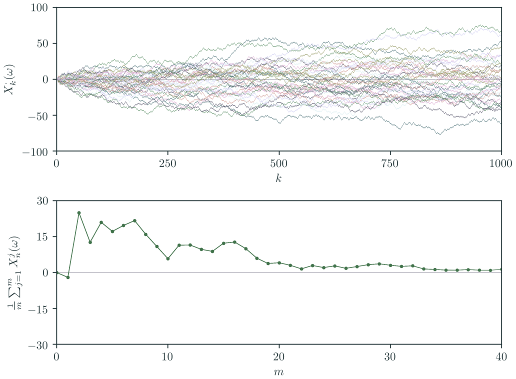

Figure 2: A realization of $s = 40$ independent copies $(X^j_k)_{k\le n}$ of $(X_k)_{k\le n}$ and their m-th sample mean at n, where $n = 10^3, m \le s, j \le s$.

Stopped Processes

Definition (Stopping Time)
Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a filtration over a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. A random variable $T : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{N}_{\leq\infty}, 2^{\mathbb{N}_{\leq\infty}})$ such that

$$\{T \leq n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}$$

is called a stopping time with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

Remark
In the above setting,

$$T \text{ is a stopping time} \iff \{T = n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}.$$

Definition (Stopped Process)
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), n \in \mathbb{N}$ be a process adapted to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega, \mathcal{A})$ and let $T : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{N}_{\leq\infty}, 2^{\mathbb{N}_{\leq\infty}})$ be a stopping time with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$. We call

$$(X_n^T)_{n\in\mathbb{N}}, \text{ where } X_n^T(\omega) := X_{\min\{T(\omega),n\}}(\omega) \text{ for all } \omega \in \Omega, n \in \mathbb{N}$$

the stopped process.

Definition (σ-algebra of the T-past)
Let $T : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{N}_{\leq\infty}, 2^{\mathbb{N}_{\leq\infty}})$ be a stopping time with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega, \mathcal{A})$. We call

$$\mathcal{A}_T := \{A \in \mathcal{A} \mid A \cap \{T \leq n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}\}$$

the σ-algebra of the T-past.

Theorem
Let $T : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{N}_{\leq\infty}, 2^{\leq\infty})$ be finite and a stopping time with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega, \mathcal{A})$. Let further $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ be a stochastic process adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Then the random variable

$$X_T : \Omega \to \mathbb{R}\,,\ \omega \mapsto X_T(\omega) := X_{T(\omega)}(\omega)$$

is $\mathcal{A}_T$-$\mathcal{B}(\mathbb{R})$-measurable.

Proof.
Let $B \in \mathcal{B}(\mathbb{R})$ and $n \in \mathbb{N}$, then due to finiteness of T, we have

$$X_T^{-1}(B) = \bigcup_{j\in\mathbb{N}} X_j^{-1}(B) \cap \{T = j\} \in \mathcal{A}$$

and

$$X_T^{-1}(B) \cap \{T \leq n\} = \bigcup_{j=0}^{n} X_j^{-1}(B) \cap \{T = j\} \in \mathcal{F}_n \implies X_T^{-1}(B) \in \mathcal{A}_T\,.$$

□

Martingale Stopping Theorem

Theorem (Martingale Stopping Theorem)
If $(X_n)_{n\in\mathbb{N}}$ is a martingale with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ and if T is a finite stopping time with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$, then

$$\mathbb{E}X_T = \mathbb{E}X_1$$

if one of the following holds:

- $(X_n)_{n\in\mathbb{N}}$ *is bounded* $\mathbb{P}$*-almost surely;*
- T *is bounded* $\mathbb{P}$*-almost surely; or*
- $\mathbb{E}T < \infty$, *and there is a constant* $c \in \mathbb{R}$ *such that*

$$\mathbb{E}(|X_{n+1} - X_n| \mid \mathcal{F}_n) < c \text{ for all } n \in \mathbb{N}.$$

Proof Sketch.

- In all cases we have a finite T. Therefore, X_T is $\mathcal{A}_T$-$\mathcal{B}(\mathbb{R})$-measurable and in particular $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$-measurable.
- There is a pointwise limit $X_n^T \to X_T$.
- By the requirements, one can uniformly bound the expected value $(\mathbb{E}X_n^T)_{n\in\mathbb{N}}$.
- Due to Lebesgue's Theorem we obtain $\lim_{n\to\infty} \mathbb{E}X_n^T = \mathbb{E}X_T$.
- By the Stopping Lemma, $\mathbb{E}X_n^T = \mathbb{E}X_1$ holds for all $n \in \mathbb{N}$.

□

Example (Gambler's Ruin 3)
We extend the previous example toward a more interesting stopping time. We consider the same martingale - the random walk $(X_n)_{n\in\mathbb{N}}$ in conjunction with the stopping time

$$T^b(\omega) := \min\{n \in \mathbb{N} \mid X_n(\omega) \in \{\ell_1, -\ell_2\} \text{ or } n = b\} \text{ for all } \omega \in \Omega\,,$$

where $b \in \mathbb{N}_{\leq\infty}$ and $\ell_1, \ell_2 \in \mathbb{N}$.

- Clearly, T^b is a stopping time with respect to $\mathcal{F}_n$, as for $n \neq b$

$$\{T^b = n\} = \{X_n \in \{\ell_1, -\ell_2\}\} \cap \left(\bigcap_{k<n} \{X_k \notin \{\ell_1, -\ell_2\}\}\right) \in \mathcal{F}_n$$

  and in case $n = b < \infty$

$$\{T^b = b\} = \bigcap_{k<b} \{X_k \notin \{\ell_1, -\ell_2\}\} \in \mathcal{F}_b\,.$$

An Application of the Martingale Stopping Theorem

$$T^b(\omega) := \min\{n \in \mathbb{N} \mid X_n(\omega) \in \{\ell_1, -\ell_2\} \text{ or } n = b\} \text{ for all } \omega \in \Omega\,.$$

- We prove that $\mathbb{P}\{T^b = \infty\} = 0$. For this let $n \in \mathbb{N}$, define $\ell := \ell_1 + \ell_2$ and pick $r_n := \max\{k \in \mathbb{N} \mid k\ell \leq n\}$. For all $b \in \mathbb{N}_{\leq\infty}$ it follows, that

$$\begin{aligned}\{T^b \geq n\} \subseteq \{T^\infty \geq n\} \subseteq \{T^\infty \geq r_n \ell\} &\subseteq \bigcap_{k<r_n} \{|X_{(k+1)\ell} - X_{k\ell}| < \ell\} \\ &= \bigcap_{k<r_n} \{|X_{(k+1)\ell} - X_{k\ell}| = \ell\}^C \\ &= \bigcap_{0\leq k<r_n} \left\{\left|\sum_{j=k\ell+1}^{(k+1)\ell} Z_j\right| = \ell\right\}^C\end{aligned}$$

and therefore

$$\mathbb{P}\{T^b \geq n\} \leq (1 - 1/2^{\ell-1})^{r_n} \xrightarrow{n\to\infty} 0\,.$$

An Application of the Martingale Stopping Theorem

$T^b(\omega) := \min\{n \in \mathbb{N} \mid X_n(\omega) \in \{\ell_1, -\ell_2\} \text{ or } n = b\}$ for all $\omega \in \Omega$.

$\mathbb{P}\{T^b \geq n\} \xrightarrow{n\to\infty} 0$ for all $b \in \mathbb{N}_{\leq\infty}$.

- It remains to show that $\mathbb{E}X_{T^b} = 0$, where $b < \infty$. As T^b is bounded, we have

$$\mathbb{E}T^b \leq b < \infty$$

and due to the fact that

$$\mathbb{E}(|X_{n+1} - X_n| \mid \mathcal{F}_n) = \mathbb{E}(|Z_{n+1}| \mid \mathcal{F}_n) \leq \mathbb{E}(1 \mid \mathcal{F}_n) = 1\,,$$

we can apply the Martingale Stopping Theorem. Setting

$$q_b := \mathbb{P}(X_{T^b} = \ell_1) \quad \text{and} \quad w_b := \mathbb{P}(X_{T^b} \notin \{\ell_1, -\ell_2\})$$

we obtain

$$\ell_1 q_b - \ell_2(1 - q_b) \leq \mathbb{E}X_{T^b} = 0 \leq \ell_1(q_b + w_b) - \ell_2(1 - q_b - w_b)\,.$$

By the previous derivation, it follows additionally that

$$0 = \lim_{b\to\infty} \mathbb{E}X_{T^b} = \lim_{b\to\infty} \ell_1 q_b - \ell_2(1 - q_b) \iff \lim_{b\to\infty} q_b = \frac{\ell_2}{\ell_1 + \ell_2}\,.$$

An Application of the Martingale Stopping Theorem

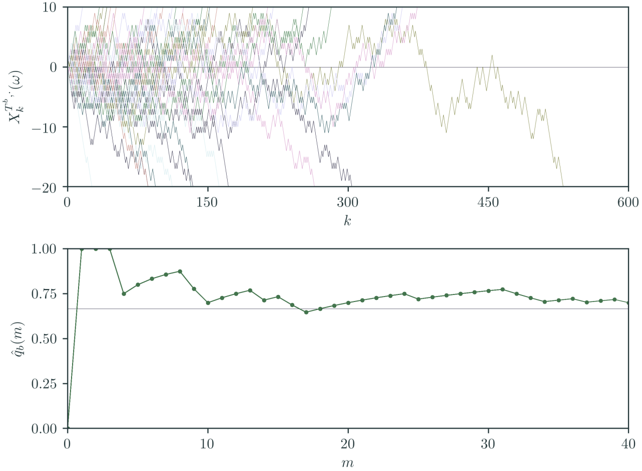

Figure 3: A realization of $s = 40$ independent copies $\left(X_k^{T_b,j}\right)_{k \leq n}$ of $\left(X_k^{T_b}\right)_{k \leq n}$ and the m-th sample mean $\hat{q}_b(m)$ of the corresponding realization of $\mathbb{1}\{X_T = \ell_1\}$, where $\ell_1 = 10; \ell_2 = 20; m, j \leq s; 0 \ll b$.

Wald's Equation

Wald's Equation

Theorem (Wald's Equation)
Let $X_n \in L^1(\Omega, \mathcal{A}, \mathbb{P}), n \in \mathbb{N}$ *be independent, identically distributed random variables and let* T *be a finite stopping time with respect to*

$$(\mathcal{F}_n := \mathcal{I}(X_1, X_2, \dots, X_n))_{n \in \mathbb{N}} .$$

If T *has bounded expectation, then*

$$\mathbb{E}\left(\sum_{i=1}^{T} X_i\right) = \mathbb{E}T \cdot \mathbb{E}X_1 .$$

Proof.

- $(Z_n := \sum_{j=1}^{n}(X_j - \mathbb{E}X_j))_{n \in \mathbb{N}}$ is a martingale with respect to $(\mathcal{F}_n)_{n \in \mathbb{N}}$.
- In conjunction with T , the martingale $(Z_n)_{n \in N}$ fulfills the third version of the Martingale Stopping Theorem.
- Due to linearity of the expected value and by the former, we obtain the result.

Wald's Equation

- $(Z_n := \sum_{j=1}^{n}(X_j - \mathbb{E}X_j))_{n\in\mathbb{N}}$ is a martingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

Clearly, Z_n is $\mathcal{F}_n$-$\mathcal{B}(\mathbb{R})$-measurable and

$$\mathbb{E}|Z_n| \le \mathbb{E}\left(\sum_{j=1}^{n}|X_j| + \mathbb{E}|X_j|\right) = 2n\mathbb{E}X_1 \text{ for all } n \in \mathbb{N}.$$

Also,

$$\begin{aligned}\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) &= \mathbb{E}\left(X_{n+1} - \mathbb{E}X_{n+1} + Z_n \mid \mathcal{F}_n\right)\\ &= \mathbb{E}\left(X_{n+1} \mid \mathcal{F}_n\right) - \mathbb{E}X_{n+1} + Z_n\\ &= Z_n \text{ for all } n \in \mathbb{N}.\end{aligned}$$

Wald's Equation

- In conjuction with T, the martingale $(Z_n)_{n\in N}$ fulfills the third version of the Martingale Stopping Theorem.

We have

$$\begin{aligned}\mathbb{E}\left(|Z_{n+1}-Z_n| \mid \mathcal{F}_n\right) &= \mathbb{E}\left(|X_{n+1}-\mathbb{E}X_{n+1}| \mid \mathcal{F}_n\right)\\ &= \mathbb{E}\left(|X_{n+1}-\mathbb{E}X_{n+1}|\right)\\ &\leq 2\mathbb{E}|X_1| \text{ for all } n\in\mathbb{N}\end{aligned}$$

and as T is finite and $\mathbb{E}T<\infty$ by assumption, the Martingale Stopping Theorem tells us

$$\mathbb{E}Z_T = \mathbb{E}Z_1\,.$$

Wald's Equation

- Due to linearity of the expected value and by the former, we obtain the result.

Clearly,

$$\mathbb{E}Z_1 = \mathbb{E}(X_1 - \mathbb{E}X_1) = \mathbb{E}X_1 - \mathbb{E}X_1 = 0\,.$$

Therefore, we have

$$\begin{aligned} 0 = \mathbb{E}Z_1 = \mathbb{E}Z_T &= \mathbb{E}\left(\sum_{j=1}^{T}(X_j - \mathbb{E}X_1)\right) \\ &= \mathbb{E}\left(\left(\sum_{j=1}^{T} X_j\right) - T\mathbb{E}X_1\right) \\ &= \mathbb{E}\left(\sum_{j=1}^{T} X_j\right) - \mathbb{E}T \cdot \mathbb{E}X_1\,, \end{aligned}$$

which gives the result.

□

Concentration of Martingales

The Azuma-Hoeffding Inequality

Theorem (Azuma-Hoeffding Inequality)
Let $(X_n)_{n\in\mathbb{N}}$ *be a martingale with respect to filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *and assume that for all* $k \in \mathbb{N}$, *there exists* $c_k \in \mathbb{R}$ *such that* $|X_{k+1} - X_k| \leq c_k$.

Then, for all $n \in \mathbb{N}$ *and for all* $\lambda > 0$, *we have*

$$\mathbb{P}\left(|X_n - X_1| \geq \lambda\right) \leq 2\exp\left(-\frac{\lambda^2}{2\sum_{k=1}^{n-1} c_k^2}\right).$$

Proof.

- For every $\alpha > 0$, there is an upper bound

$$\mathbb{E}\left(e^{\alpha(X_n - X_1)}\right) \leq \exp\left(\frac{\alpha^2}{2}\sum_{k=1}^{n-1} c_k^2\right) \text{ for all } n \in \mathbb{N}.$$

- By Markov's Inequality and by subadditivity of $\mathbb{P}$, we therefore obtain a bound

$$\mathbb{P}(|X_n - X_1| \geq \lambda) \leq 2\exp\left(\frac{\alpha^2}{2}\sum_{k=1}^{n-1} c_k^2 - \alpha\lambda\right) \text{ for all } n \in \mathbb{N}.$$

- We minimize the upper bound over $\alpha > 0$ to obtain the result.

The Azuma-Hoeffding Inequality

· For every $\alpha > 0$, there is an upper bound

$$\mathbb{E}\left(e^{\alpha(X_n - X_1)}\right) \leq \exp\left(\frac{\alpha^2}{2}\sum_{k=1}^{n-1} c_k^2\right) \text{ for all } n \in \mathbb{N}.$$

For all $n \in \mathbb{N}$, we define $Y_{n+1} := X_{n+1} - X_n$ and given $\alpha > 0$, we divide this step into the three substeps

- ◇ There is an upper bound

$$e^{\alpha Y_{n+1}} \leq \frac{e^{-\alpha c_n} + e^{\alpha c_n}}{2} + \frac{Y_{n+1}}{2c_n}(e^{\alpha c_n} - e^{-\alpha c_n});$$

- ◇ There is an upper bound

$$\mathbb{E}\left(e^{\alpha Y_{n+1}} \mid \mathcal{F}_n\right) \leq \exp\left((\alpha c_n)^2/2\right) \text{ ; and}$$

- ◇ There is an upper bound

$$\mathbb{E}\left(e^{\alpha(X_{n+1} - X_1)}\right) \leq \mathbb{E}\left(e^{\alpha(X_n - X_1)}\right) \exp\left((\alpha c_n)^2/2\right)$$

from which the statement can be concluded by induction.

The Azuma-Hoeffding Inequality

◇ There is an upper bound

$$e^{\alpha Y_{n+1}} \leq \frac{e^{-\alpha c_n} + e^{\alpha c_n}}{2} + \frac{Y_{n+1}}{2c_n}(e^{\alpha c_n} - e^{-\alpha c_n});$$

By assumption $|Y_{n+1}| = |X_{n+1} - X_n| \leq c_n$, therefore we can find a convex combination

$$Y_{n+1} = -c_n \frac{1 - {}^{Y_{n+1}}/_{c_n}}{2} + c_n \frac{1 + {}^{Y_{n+1}}/_{c_n}}{2}.$$

Using the convexity of $t \mapsto e^{\alpha t}$, we obtain that

$$\begin{aligned} e^{\alpha Y_{n+1}} &\leq \frac{1 - {}^{Y_{n+1}}/_{c_n}}{2} e^{-\alpha c_n} + \frac{1 + {}^{Y_{n+1}}/_{c_n}}{2} e^{\alpha c_n} \\ &= \frac{e^{-\alpha c_n} + e^{\alpha c_n}}{2} + \frac{Y_{n+1}}{2c_n}(e^{\alpha c_n} - e^{-\alpha c_n}). \end{aligned}$$

◇ There is an upper bound

$$\mathbb{E}\left(e^{\alpha Y_{n+1}} \mid \mathcal{F}_n\right) \le \exp\left({(\alpha c_n)^2}/{2}\right) \;;$$

Note that $(X_n)_{n\in\mathbb{N}}$ is a martingale with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, thus

$$\begin{aligned}\mathbb{E}\,(Y_{n+1} \mid \mathcal{F}_n) &= \mathbb{E}\,(X_{n+1} - X_n \mid \mathcal{F}_n) \\ &= \mathbb{E}\,(X_{n+1} \mid \mathcal{F}_n) - X_n = 0\,.\end{aligned}$$

Using this and the previous result, we obtain

$$\begin{aligned}\mathbb{E}\left(e^{\alpha Y_{n+1}} \mid \mathcal{F}_n\right) &\le \mathbb{E}\left(\frac{e^{\alpha c_n} + e^{-\alpha c_n}}{2} + \frac{Y_{n+1}}{2c_n}\left(e^{\alpha c_n} - e^{-\alpha c_n}\right) \mid \mathcal{F}_n\right) \\ &= \frac{e^{\alpha c_n} + e^{-\alpha c_n}}{2} = \tfrac{1}{2}\sum_{j=0}^{\infty} \frac{(\alpha c_n)^j + (-\alpha c_n)^j}{j!} \\ &= \sum_{j=0}^{\infty} \frac{(\alpha c_n)^{2j}}{(2j)!} \le \exp\left({(\alpha c_n)^2}/{2}\right)\,.\end{aligned}$$

The Azuma-Hoeffding Inequality

◇ There is an upper bound

$$\mathbb{E}\left(e^{\alpha(X_{n+1}-X_1)}\right) \le \mathbb{E}\left(e^{\alpha(X_n-X_1)}\right)\exp\left((\alpha c_n)^2/2\right) ;$$

From the previous statement and by the tower property it follows that

$$\begin{aligned}\mathbb{E}\left(e^{\alpha(X_{n+1}-X_1)}\right) &= \mathbb{E}\left(\mathbb{E}\left(e^{\alpha(X_n-X_1)}e^{\alpha Y_{n+1}} \mid \mathcal{F}_n\right)\right)\\ &= \mathbb{E}\left(e^{\alpha(X_n-X_1)}\mathbb{E}\left(e^{\alpha Y_{n+1}} \mid \mathcal{F}_n\right)\right)\\ &\le \mathbb{E}\left(e^{\alpha(X_n-X_1)}\right)\exp\left((\alpha c_n)^2/2\right) .\end{aligned}$$

By induction and the fact that $e^{\alpha(X_1-X_1)} = 1$, we obtain

$$\mathbb{E}\left(e^{\alpha(X_n-X_1)}\right) \le \exp\left(\frac{\alpha^2}{2}\sum_{k=1}^{n-1} c_k^2\right) \quad \text{for all } n \in \mathbb{N} .$$

The Azuma-Hoeffding Inequality

- By Markov's Inequality and by subadditivity of $\mathbb{P}$, we therefore obtain a bound

$$\mathbb{P}(|X_n - X_1| \geq \lambda) \leq 2\exp\left(\frac{\alpha^2}{2}\sum_{k=1}^{n-1} c_k^2 - \alpha\lambda\right) \text{ for all } n \in \mathbb{N}\,.$$

Using the previous result, we have for all $\alpha > 0$, for all $\lambda > 0$, and for all $n \in \mathbb{N}$ that

$$\begin{aligned}\mathbb{P}(X_n - X_1 \geq \lambda) &= \mathbb{P}\left(e^{\alpha(X_n - X_1)} \geq e^{\alpha\lambda}\right)\\ &\leq \frac{\mathbb{E}\left(e^{\alpha(X_n - X_1)}\right)}{e^{\alpha\lambda}}\\ &\leq \exp\left(\alpha^2\sum_{k=1}^{n-1} c_k^2/2 - \alpha\lambda\right)\,.\end{aligned}$$

By the same argument applied to $(-X_n)_{n\in\mathbb{N}}$ we obtain the bound

$$\mathbb{P}(-X_n + X_1 \geq \lambda) \leq \exp\left(\alpha^2\sum_{k=1}^{n-1} c_k^2/2 - \alpha\lambda\right)\,.$$

Subadditivity of $\mathbb{P}$ then yields the result.

The Azuma-Hoeffding Inequality

- We minimize the upper bound over $\alpha > 0$ to obtain the result.

For all $n \in \mathbb{N}$ and $\lambda > 0$, minimizing the upper bound of

$$\mathbb{P}(|X_n - X_1| \geq \lambda) \leq 2\exp\left(\frac{\alpha^2}{2}\sum_{k=1}^{n-1} c_k^2 - \alpha\lambda\right)$$

over $\alpha > 0$ is equivalent to minimizing the function

$$\alpha \longmapsto \frac{\alpha^2}{2}\sum_{k=1}^{n-1} c_k^2 - \alpha\lambda\,.$$

The minimum of the above function is found in its only local minimum and as the unique root

$$\alpha \sum_{k=1}^{n-1} c_k^2 - \lambda = 0 \iff \alpha = \frac{\lambda}{\sum_{k=1}^{n-1} c_k^2}\,.$$

of its derivative with respect to α. Plugging the minimum into the exponent of the above bound, we obtain

$$\frac{\alpha^2}{2}\sum_{k=1}^{n-1} c_k^2 - \alpha\lambda = \frac{1}{2}\frac{\lambda^2}{\sum_{k=1}^{n-1} c_k^2} - \frac{\lambda^2}{\sum_{k=1}^{n-1} c_k^2} = -\frac{\lambda^2}{2\sum_{k=1}^{n-1} c_k^2}$$

and conclude the result. $\square$

Definition (Graphs)
Define a dense graph $G_m := (V_m, E_m)$, where $m \in \mathbb{N}$,

- the set $V_m := \{v_j \mid j \in \mathbb{N},\ j \leq m\}$ is called vertices and
- the set
$$E_m := \{e : \{0,1\} \to V_m \mid e(0) \neq e(1)\} /{\sim}\,,$$
where
$$e \sim e' :\iff e\{0,1\} = e'\{0,1\}\,,$$
is called edges.

We say $G := (V, E)$ is a graph k-subordinate to G_m if

- $V = V_m$,
- $E \subseteq E_m$, and
- $|E| = k$.

We write $\mathcal{G}_m^k := \{G \mid G \text{ is } k\text{-subordinate to } G_m\})$. Further, we call a vertice $v \in V$ isolated if
$$v \notin \bigcup_{e \in E} e\{0,1\}\,.$$

Example (Concentration of the Number of Isolated Vertices)
We will now take a look at random variable

$$G \sim \mathrm{Unif}(\mathcal{G}_m^k)\,.$$

Clearly, as the range of G is finite, we can consider the associated power set as a σ-algebra. Of special interest to us are the number of isolated vertices $\mathrm{iso}(G)$, constructed with the map

$$\begin{cases} \mathrm{iso} : \mathcal{G}_m^k \longrightarrow \mathbb{N}_{[0,m]} \\ (V, E) \longmapsto |\{v \in V \mid v \text{ is isolated}\}|\,. \end{cases}$$

We will now

- Compute $\mathbb{E}\,\mathrm{iso}(G)$ and
- Prove that $\mathbb{P}\{|\mathrm{iso}(G) - \mathbb{E}\,\mathrm{iso}(G)| \geq \lambda\} \leq 2\exp\left(-\frac{\lambda^2}{8k}\right)$.

An Application of the Azuma-Hoeffding Inequality

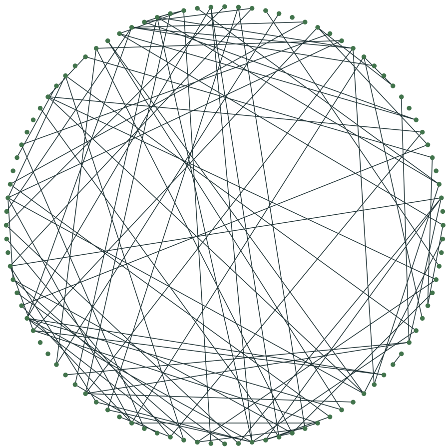

Figure 4: A realization of $G \sim \text{Unif}(\mathcal{G}_m^k)$, where $m = k = 100$.

An Application of the Azuma-Hoeffding Inequality

- Compute $\mathbb{E}\,\mathrm{iso}(G)$;

We have the identity

$$\begin{aligned}\mathbb{E}\,\mathrm{iso}(G) &= \mathbb{E}\sum_{v\in V} \mathbb{1}\{v \text{ is isolated}\}\\ &= \sum_{v\in V} \mathbb{E}\,\mathbb{1}\{v \text{ is isolated}\}\\ &= |V| \cdot \mathbb{P}\{v_1 \text{ is isolated}\}\,.\end{aligned}$$

Clearly, $|V| = |V_m| = m$. As the random variable G takes exactly the values of every graph containing k edges, we have

$$|\mathcal{G}_m^k| = \binom{\binom{m}{2}}{k}\,.$$

For the probability of a vertice being isolated we count the number of graphs that fulfill this property and obtain the associated probability by division

$$\mathbb{P}\{v_1 \text{ is isolated}\} = \frac{|\mathcal{G}_{m-1}^k|}{|\mathcal{G}_m^k|}\,.$$

An Application of the Azuma-Hoeffding Inequality

· Prove that $\mathbb{P}\{|\mathrm{iso}(G) - \mathbb{E}\,\mathrm{iso}(G)| \geq \lambda\} \leq 2\exp\left(-\frac{\lambda^2}{8k}\right)$.

This follows, as we can

- Build an edge exposure martingale $(X_n)_{n\in\mathbb{N}}$ of $\mathrm{iso}(G)$ with respect to G;
- Bound $|X_{n+1} - X_n| \leq 2$ for all $n \in \mathbb{N}$; and
- Show that the result can be obtained by applying the Azuma-Hoeffding Inequality to the martingale $(X_n)_{n\in\mathbb{N}}$.

An Application of the Azuma-Hoeffding Inequality

◇ Build an edge exposure martingale $(X_n)_{n\in\mathbb{N}}$ of iso(G) with respect to G;

Consider an enumeration $(e)_{i\leq j}$, $j = \frac{m(m-1)}{2}$ of the edges E_m of a dense graph with m vertices. We will now define a random variable describing the vertices that are connected to e_i in the random graph G. For $1 \leq i \leq j$, we define

$$\begin{cases} \tilde{e}_i : \mathcal{G}_n^k \longrightarrow 2^V \\ (V, E) \longmapsto e_i\{0,1\} \text{ if } e_i \in E\,,\ \emptyset \text{ else}\,. \end{cases}$$

and therefore $\tilde{e}_i(G)$ describes just that.
Now for all $n \in \mathbb{N}$, we define $\mathcal{F}_n := \mathcal{I}(\tilde{e}_1(G), \ldots, \tilde{e}_n(G))$ and

$$X_0 := \mathbb{E}\,\mathrm{iso}(G)\,,\ X_n := \mathbb{E}(\mathrm{iso}(G) \mid \mathcal{F}_n)\,.$$

Clearly, $(X_n)_{n\in\mathbb{N}_{\geq 0}}$ is a martingale as for all $n \in \mathbb{N}_{\geq 0}$, we have

$$\begin{aligned} \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) &= \mathbb{E}(\mathbb{E}(\mathrm{iso}(G) \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n) \\ &= \mathbb{E}(\mathrm{iso}(G) \mid \mathcal{F}_n) = X_n\,. \end{aligned}$$

An Application of the Azuma-Hoeffding Inequality

◇ Bound $|X_{n+1} - X_n| \leq 2$ for all $n \in \mathbb{N}$;

By definition, we have for $G = (V, E) \sim \text{Unif}(\mathcal{G}_m^k)$, that

$$\begin{aligned} \text{iso}(G) &= \left|\left(\bigcup_{e\in E} e\{0,1\}\right)^C\right| = \left|\left(\bigcup_{i\leq j} \tilde{e}_i(G)\right)^C\right| \\ &\leq \left|\left(\bigcup_{\substack{i\leq j \\ i\neq n+1}} \tilde{e}_i(G)\right)^C\right| . \end{aligned}$$

Therefore, for all $n \in \mathbb{N}$, we obtain

$$\begin{aligned} X_{n+1} - X_n &= \mathbb{E}(\text{iso}(G) \mid \mathcal{F}_{n+1}) - \mathbb{E}(\text{iso}(G) \mid \mathcal{F}_n) \\ &\leq \mathbb{E}\left(\left|\bigcap_{\substack{i\leq j \\ i\neq n+1}} \tilde{e}_i(G)^C\right| \,\middle|\, \mathcal{F}_{n+1}\right) - \mathbb{E}(\text{iso}(G) \mid \mathcal{F}_n) \\ &= \mathbb{E}\left(\left|\bigcap_{\substack{i\leq j \\ i\neq n+1}} \tilde{e}_i(G)^C\right| - \text{iso}(G) \,\middle|\, \mathcal{F}_n\right) \leq \mathbb{E}\left(|\tilde{e}_{n+1}(G)| \mid \mathcal{F}_n\right) \\ &\leq 2 . \end{aligned}$$

By a similar derivation we get $X_{n+1} - X_n \geq -2$ and conclude the result.

An Application of the Azuma-Hoeffding Inequality

◇ Show that the result can be obtained by applying the Azuma-Hoeffding Inequality to the martingale $(X_n)_{n\in\mathbb{N}}$.

Clearly, $X_n = \operatorname{iso}(G)$ if $n \geq \frac{m(m-1)}{2}$ and $X_0 = \mathbb{E}\operatorname{iso}(G)$. By the fact that G has only k edges and by the Azuma-Hoeffding Inequality, we get

$$\mathbb{P}\{|\operatorname{iso}(G) - \mathbb{E}\operatorname{iso}(G)| \geq \lambda\} \leq 2\exp\left(-\frac{\lambda^2}{2\sum_{i=1}^{m(m-1)/2} c_i^2}\right) = 2\exp\left(-\frac{\lambda^2}{8k}\right) .$$

An Application of the Azuma-Hoeffding Inequality

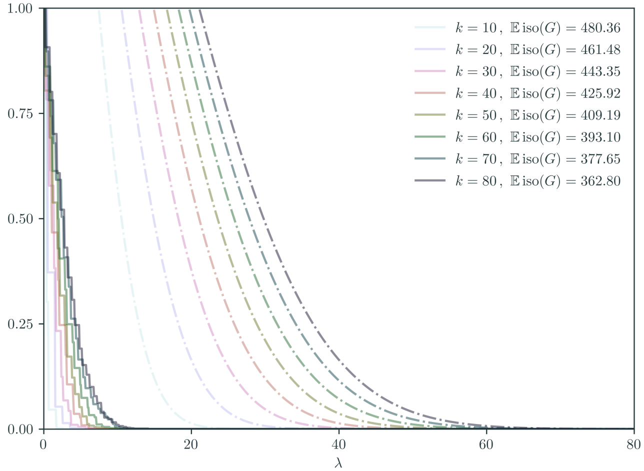

Figure 5: An estimation of $\lambda \mapsto \mathbb{P}\{|\mathrm{iso}(G) - \mathbb{E}\,\mathrm{iso}(G)| \geq \lambda\}$, created from 1000 independent samples of $\mathrm{Unif}(\mathcal{G}_m^k)$, and the associated bounds (dash-dotted) for $m = 500$ and different values of k .

Thank You!

Find the full source-code of this presentation on my Github soon:

github.com/NiMlr

The content of this presentation is mainly based on and influenced by:

- M. Mitzenmacher and E. Upfal (2017). *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis.* Cambridge University Press. URL: `https://books.google.de/books?id=E9UlDwAAQBAJ`
- N. Henze (2016). *Vorlesung Wahrscheinlichkeitstheorie.* KIT Bibliothek. URL: `https://publikationen.bibliothek.kit.edu/1000113898`

The source-code to generate the visualizations was written in *Python 3* and makes use of the software:

- P. Virtanen et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, pp. 261–272
- F. Pedregosa et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830
- A. Hagberg, D. Schult, and P. Swart (2008). "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th Python in Science Conference.* Pasadena, CA USA, pp. 11–15