3) Konvergenzsätze und Monte Carlo Methoden.

3.1) Kovarianz.

. Wir haben schon die Varianz definiert. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein W-raum und $X: \Omega \to S$ eine diskrete reelle ZV.

Def. 1) Falls $\mathbb{E}(X^2) < \infty$, dann

$$Var(X) := \mathbb{E}((X - \mathbb{E}(X))^2) \in [0, \infty)$$

ist die Varianz und

$$\sigma(X) := \sqrt{Var(X)}$$

ist die Standardabweichung von X.

Eigenschaften: . $Var(X + a) = Var(X)$, $\forall a \in \mathbb{R}$ fest.

. $Var(\lambda X) = \lambda^2 Var(X)$

. Falls $X_1, \ldots, X_n$ unabhängig sind, dann $Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n)$.

. $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

. $Var(X) = 0 \iff \mathbb{P}(X = \mathbb{E}(X)) = 1$.

Def. 2) (Kovarianz)

ⓐ Seien X, Y mit $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$.

Dann $Cov(X, Y) := \mathbb{E}(X \cdot Y) - \mathbb{E}(X)\mathbb{E}(Y)$

heisst die Kovarianz von X und Y

(b) Falls $\sigma(X), \sigma(Y) \neq 0$, dann
$$\rho(X,Y) := \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$$
heisst den Korrelationskoeffizient von X und Y.

(c) X und Y heissen unkorreliert, falls $Cov(X,Y) = 0$.

18. Mai

Bemerkung: (a) $|\rho(X,Y)| \leq 1$. (Siehe Übungen).

(b) Falls $\mathbb{E}(X^2) < \infty$ und $\mathbb{E}(Y^2) < \infty$, dann

aus Cauchy-Schwarz: $\mathbb{E}(|X|) < \infty$, $\mathbb{E}(|Y|) < \infty$ und $\mathbb{E}(|X \cdot Y|) < \infty \Rightarrow Cov(X,Y)$ ist wohldefiniert.

(c) $Cov(X,X) = Var(X)$

(d) Unabhängigkeit $\Rightarrow$ Unkorreliertheit
$$\not\Leftarrow$$

In der Tat: $\mathbb{E}(X \cdot Y) = \sum_{\substack{x \in S_1 \\ y \in S_2}} x \cdot y \, \mathbb{P}(X=x, Y=y)$
$$\overset{X,Y \text{ unab.}}{=} \sum_{x \in S_1} x \cdot \mathbb{P}(X=x) \sum_{y \in S_2} y \, \mathbb{P}(Y=y)$$
$$= \mathbb{E}(X)\mathbb{E}(Y).$$

Andererseits: Sei $X = +1, 0, -1$ mit jeweils W'-keit $\frac{1}{3}$, und $Y := X^2$.
$$\Rightarrow \mathbb{E}(X \cdot Y) = \mathbb{E}(X^3) = 0 \quad \text{(wegen Antisymmetrie)}$$
und $\mathbb{E}(X)\mathbb{E}(Y) = 0$ (weil $\mathbb{E}(X) = 0$).

Aber, $\mathbb{P}(X \in \{-1,1\}, Y=1) = \frac{2}{3} \neq \mathbb{P}(X \in \{-1,1\}) \cdot \mathbb{P}(Y=1) = \frac{4}{9}$.

Def. 3) $\mathcal{L}^2 = \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) := \{X: \Omega \to \mathbb{R} \mid \mathbb{E}(X^2) < \infty\}$.

Lem. 4) ⓐ Falls $X, Y \in \mathcal{L}^2$, $\mathbb{E}(X \cdot Y) \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$

ⓑ $\mathcal{L}^2$ ist ein Vektorraum mit

$$(X, Y)_{\mathcal{L}^2} := \mathbb{E}(X \cdot Y)$$

ein Skalarprodukt auf $\mathcal{L}^2$.

(ohne Beweis → siehe Skript).

Jetzt untersuchen wir den Zusammenhang zwischen Unabhängigkeit und Unkorreliertheit.

Satz 3) Seien $X: \Omega \to S \subset \mathbb{R}$, $Y: \Omega \to T \subset \mathbb{R}$ (diskrete) ZV. auf $(\Omega, \mathcal{F}, \mathbb{P})$. Dann haben wir zwei äquivalente Aussagen:

ⓐ X und Y sind unabhängig

(d.h. $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$, $\forall A, B \in \mathcal{F}$)

ⓑ $f(X)$ und $g(Y)$ sind unkorreliert für alle Funktionen $f: S \to \mathbb{R}$, $g: T \to \mathbb{R}$ mit $f(X), g(Y) \in \mathcal{L}^2$.

Beweis: ⓐ ⇒ ⓑ: Trivial: $\mathbb{E}(f(X)g(Y)) =$

$$= \sum_{a \in S} \sum_{b \in T} f(a) g(b)\, \mathbb{P}(X=a, Y=b)$$

$$\overset{\text{unab.}}{=} \sum_{a \in S} f(a)\, \mathbb{P}(X=a) \sum_{b \in T} g(b)\, \mathbb{P}(Y=b) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).$$

(b) $\Rightarrow$ (a): $\forall\, a \in S, b \in T$,

$$\mathbb{P}(X=a, Y=b) = \mathbb{E}(\mathbb{1}_{\{a\}}(X)\,\mathbb{1}_{\{b\}}(Y))$$

$$\overset{\text{unkorr.}}{=} \mathbb{E}(\mathbb{1}_{\{a\}}(X))\,\mathbb{E}(\mathbb{1}_{\{b\}}(Y))$$

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

#

. Die Kovarianz spielt eine Rolle in die Summe von Z.V.

Satz 4) Seien $X_1, \dots, X_n \in \mathcal{L}^2$. Dann,

$$\mathrm{Var}(X_1 + \dots + X_n) = \sum_{k=1}^{n} \mathrm{Var}(X_k) + 2 \sum_{1 \le k < l \le n} \mathrm{Cov}(X_k, X_l)$$

Beweis: Die Kovarianz ist bilineär und $\mathrm{Var}(Y) = \mathrm{Cov}(Y,Y)$.

So, $\mathrm{Var}(X_1 + \dots + X_n) = \mathrm{Cov}\left(\sum_{k=1}^{n} X_k, \sum_{l=1}^{n} X_l\right)$

$$= \sum_{k,l=1}^{n} \mathrm{Cov}(X_k, X_l) = \sum_{k=1}^{n} \mathrm{Var}(X_k) + 2 \sum_{k<l} \mathrm{Cov}(X_k, X_l).$$

#

3.2) Gesetze der grossen Zahlen.

. Falls wir unabhängige Versuche $X_1, X_2, \dots$ mit $X_k \sim \mu$, $\forall k \geq 1$, dann ist

$$m_n := \frac{1}{n} \sum_{k=1}^{n} X_k ,$$

das arithmetische Mittelwert von $X_1, \dots, X_n$ nicht so weit von $\mathbb{E}(m_n)$.

. Falls aber $X_1, X_2, \dots$ sind 100% korreliert, d.h.,

$X_1 = X_2 = \dots = X_n$, dann $m_n \not\approx \mathbb{E}(m_n)$.

· Wann ist $m_n \approx \mathbb{E}(m_n)$?

Satz 5) (Schwaches Gesetz der grossen Zahlen).

· Seien $X_1, X_2, \dots, X_n \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, mit

ⓐ $\mathrm{Cov}(X_k, X_\ell) = 0, \ \forall k \neq \ell$, (Unabhängigkeit ist nicht vorausgesetzt).

ⓑ $v := \sup_{n \geq 1} \mathrm{Var}(X_n) < \infty$.

· Sei $S_n := X_1 + \dots + X_n$. Dann, $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{S_n}{n} - \frac{\mathbb{E}(S_n)}{n} \right| \geq \varepsilon \right) \leq \frac{v}{\varepsilon^2 n} \xrightarrow{n\to\infty} 0$$

Bem.: · Falls (dazu) $\mathbb{E}(X_k) = m \ \forall k \geq 1$, $\frac{\mathbb{E}(S_n)}{n} = m$ und man sagt, dass $\frac{S_n}{n}$ konvergiert stochastisch gegen m, d.h.,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{S_n}{n} - m \right| \geq \varepsilon \right) = 0 \quad , \forall \varepsilon > 0.$$

. Zum Beweis benötigen wir die Tchebishev-Ungleichung.

Lemma 6) Sei $X \in \mathcal{L}^2$ und $a > 0$. Dann,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

Beweis: $\mathbb{P}(|X - \mathbb{E}(X)| \geq a) = \mathbb{E}\left(\mathbb{1}_{\{|X-\mathbb{E}(X)| \geq a\}}\right)$.

. Aber $\mathbb{1}_{\{|X-\mathbb{E}(X)| \geq a\}} \cdot \frac{a^2}{a^2} \leq \frac{|X - \mathbb{E}(X)|^2}{a^2}$

$$\Rightarrow \mathbb{P}(|X-\mathbb{E}(X)| \geq a) \leq \mathbb{E}\left(\frac{|X-\mathbb{E}(X)|^2}{a^2}\right) = \frac{Var(X)}{a^2}. \quad \#$$

Beweis von Satz 5) Sei $\varepsilon > 0$. Dann,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right| \geq \varepsilon\right) \leq$$

$$\underset{\text{Lemma 6}}{\leq} \frac{Var}{\varepsilon^2}\left(\frac{S_n}{n}\right) \underset{(a)}{=} \frac{1}{\varepsilon^2 n^2} Var\left(\sum_{k=1}^{n} X_k\right)$$

$$\underset{(b)}{\leq} \frac{n \cdot v}{n^2 \varepsilon^2} = \frac{v}{n^2 \varepsilon}.$$

$\#$

. Es gibt noch zwei wichtige Ungleichungen, die wir der Vollständigkeit halber geben.

. Markov Ungleichung: . Falls $\mathbb{E}(X)$ existiert, $\forall a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}.$$

. Jensen Ungleichung: Sei $\varphi: \mathbb{R} \to \mathbb{R}$ konvex, dann falls $\mathbb{E}(|X|) < \infty$ und $\mathbb{E}(|\varphi(X)|) < \infty$, dann

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

(Bem.: Statt $Cov(X_k, X_\ell) = 0$ für $k \neq \ell$, kann man die Bedingung schwächer (z.B. $|Cov(X_k, X_\ell)| \leq f(|k-\ell|)$ mit $f(x) \xrightarrow[x\to\infty]{} 0$ ersetzen.)

3.3) Monte-Carlo Verfahren.

Ziel: Wir wollen unabhängige, identische Verteilte Versuche benutzen um

$$\Theta := \mathbb{E}_\mu(f) = \sum_{x \in S} f(x)\mu(x)$$

zu approximieren. Für $f(x) = x$, Satz 5 sagt uns, dass wenn $X_1, X_2, \dots, X_n \sim \mu$, dann

$$\mathbb{E}_\mu(x) = \mathbb{E}(X) \simeq \frac{1}{n}\sum_{k=1}^{n} X_k.$$

. Sei S abzählbar, μ eine W-Verteilung auf S.

Notation: . $\mu(x) := \mu(\{x\})$.

. Sei $f: S \to \mathbb{R}$ mit

$$\mathbb{E}_\mu(f^2) = \sum_{x \in S} f^2(x)\,\mu(x) < \infty.$$

. Ein Schätzer von Θ ist, nach Satz 5,

$$\hat{\Theta}_n := \frac{1}{n}\sum_{k=1}^{n} f(X_k)$$

wobei $X_1, X_2, \dots \sim \mu$ sind unabhängige Z.V.

Lemma 7) $\forall \varepsilon > 0$, $\mathbb{P}\left(|\hat{\Theta}_n - \Theta| \geq \varepsilon\right) \leq \dfrac{Var_\mu(f^2)}{n\varepsilon^2} \xrightarrow{n\to\infty} 0$

Bem.: Eine Folge von Schätzern $\hat{\Theta}_n$ (von Θ) mit der Eigenschaft $\lim_{n\to\infty} \mathbb{P}(|\hat{\Theta}_n - \Theta| \geq \varepsilon) = 0$, $\forall \varepsilon > 0$, heisst Konsistent.

Beweis: . X_k un ab. $\underset{\text{Satz 3}}{\Rightarrow}$ $f(X_k)$ sind unkorreliert.

$$\cdot\ \mathbb{E}(f(X_k)) = \sum_{x \in S} f(x)\mu(x) = \mathbb{E}_\mu(f) = \theta, \quad \forall k$$

und $$\mathrm{Var}(f(X_k)) = \sum_{x\in S} (f(x)-\theta)^2 \mu(x) = \mathrm{Var}_\mu(f) < \infty.$$

$\Rightarrow$ Folgt aus Satz 5. $(X_k \to f(X_k))$ #

Bem.: • $\hat{\theta}_n$ ist ein erwartungstreu Schätzer für θ, falls $\mathbb{E}(\hat{\theta}_n) = \theta$.

. Das mittleren quadratische Fehler von $\hat{\theta}_n$, definiert als

$$\mathbb{E}(|\hat{\theta}_n - \theta|^2) = \frac{1}{n} \mathrm{Var}_\mu(f).$$

$$\Rightarrow \|\hat{\theta}_n - \theta\|_{L^2} := \sqrt{\mathbb{E}(|\hat{\theta}_n-\theta|^2)} = O\left(\frac{1}{\sqrt{n}}\right).$$

. MC nützlich, wenn wir μ nicht explizit genug wissen.

23. Mai

Beispiel: (1) Monte-Carlo-Schätzung von $\theta = \int_0^1 f(x)dx$

. Seien $u_1, u_2, \dots, u_n \in [0,1)$ Pseudozahlen mit $U_k \sim \mathrm{Unif}([0,1))$.

Dann $$\hat{\theta}_n := \frac{1}{n}\sum_{k=1}^n f(u_k)$$

ist ein Konsistent Schätzer für $\theta = \int_0^1 f(x)dx$,

und $\sqrt{\mathbb{E}(|\hat{\theta}_n - \theta|^2)} = O(1/n)$.

(2) Schätzung von W-keiten.

. Sei S diskret und $B \subseteq S$.

Gesucht ist $p := \mu(B)$ (unbekannt)

Da $\mu(B) = \mathbb{E}_\mu(\mathbb{1}_B)$, ein MC Schätzer von p ist

$$\hat{p}_n := \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_B(X_k)$$

mit $X_k \sim \mu$ unabhängig.

. Wie gut ist die Schätzung von p?

Tchebishev:

$$\mathbb{P}(|\hat{p}_n - p| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} Var(\hat{p}_n)$$

$$= \frac{1}{n\varepsilon^2} Var_\mu(\mathbb{1}_B) = \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

3.3.1) Varianzreduktion durch Importance Sampling

Idee: Wenn wir weniger Punkte (Sample), wo $f(x)$ klein ist, nehmen, wird die Varianz kleiner.

. Sei μ die W-keit auf S, $f: S \to \mathbb{R}$ s.d.

$$\mathbb{E}_\mu(f^2) = \sum_{x \in S} f^2(x)\mu(x) < \infty.$$

. Sei ν eine weitere W-keit auf S mit

$$\nu(x) = \nu(\{x\}) > 0, \forall x \in S.$$

Dann, haben wir:

$$\Theta := \mathbb{E}_\mu(x) = \sum_{x \in S} f(x)\mu(x)$$

$$= \sum_{x \in S} f(x) \frac{\mu(x)}{\nu(x)} \cdot \nu(x) = \mathbb{E}_\nu(f g)$$

wobei $g(x) = \frac{\mu(x)}{\nu(x)}$.

Dann, kann man statt

$$\hat{\Theta}_n = \frac{1}{n}\sum_{k=1}^{n} f(X_k), \quad X_k \sim \mu,$$

den Schätzer $\tilde{\Theta}_n := \frac{1}{n}\sum_{k=1}^{n} f(Y_k) g(Y_k)$, $Y_k \sim \nu$.

- $\mathbb{E}_\nu(\tilde{\Theta}_n) = \mathbb{E}_\nu(f g) = \Theta$; $\mathbb{E}_\mu(\hat{\Theta}_n) = \Theta$: Beide Erwartungstreu.

$$\begin{cases} \operatorname{Var}_\nu(\tilde{\Theta}_n) = \frac{1}{n}\operatorname{Var}_\nu(fg) = \frac{1}{n}\left(\sum_{x \in S} f^2(x) g^2(x)\nu(x) - \Theta^2\right). \\ \operatorname{Var}_\mu(\hat{\Theta}_n) = \frac{1}{n}\left(\sum_{x \in S} f^2(x)\mu(x) - \Theta^2\right) \end{cases}$$

- Ziel: Wahl von ν s.d. $\operatorname{Var}_\nu(\tilde{\Theta}_n) \ll \operatorname{Var}_\mu(\hat{\Theta}_n)$.

Da $\sum_{x \in S} f^2(x) g^2(x)\nu(x) = \sum_{x \in S} f^2(x)\mu(x)\frac{\mu(x)}{\nu(x)}$

wird man $\nu(x)$ gross wählen für $f^2(x)$ gross.

Beispiel: Siehe Skript / Übung.

3.4) Gleichgewichte von Markovketten

. Die Übergangsmatrizen einer MK sind stochastische Matrizen (siehe Kap 2).

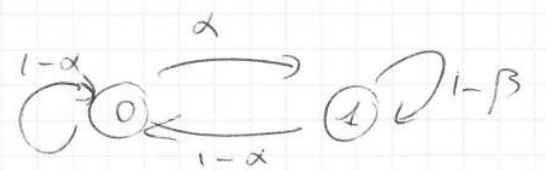
Def. 8) Eine MK heisst homogen (zeitlich homogen) falls die Übergangsmatrizen P_k, $k\geq 1$, sind unabhängig von der Zeit k, d.h., $P_k = P$, $\forall k\geq 1$.

. Wir haben schon gesehen wie erhält man die Verteilung μ_n zur Zeit n, falls die Verteilung zur Zeit $t=0$ ist μ_0:

$$\mu_n = \mu_0 P^n$$

(siehe Kap 2, Satz 10).

. Im Beispiel



, $\alpha,\beta\in(0,1)$

haben wir gesehen, dass $\lim_{n\to\infty}\mu_0 P^n = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$ unabhängig von μ_0.

. Falls $\mu_\infty = \lim_{n\to\infty}\mu_n$ existiert, dann $\mu_{n+1}\approx\mu_n$ für $n\gg1$

$\mu_{n+1}=\mu_n P \quad \Rightarrow \quad \underline{\mu_\infty = \mu_\infty P}$.

. μ_∞ ist eine stationäre Verteilung, weil sie entwickelt sich nicht mehr.

Def. 9) (a) Stationäre Verteilung.

. Eine W-Verteilung μ auf S heisst eine stationäre Verteilung einer MK mit Übergangsmatrix P falls $\underline{\mu = \mu P}$.

(b) Detailed-Balance

- μ erfüllt die Detailed-Balance Bedingung bzgl. P falls
$$\mu(x)P(x,y) = \mu(y)P(y,x), \quad \forall x,y \in S.$$

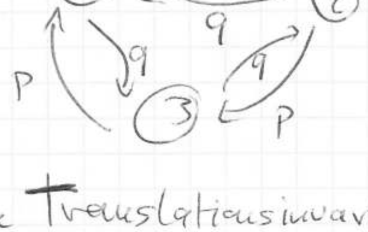
↳ Das ist die lokale Gleichgewichtsbedingung, d.h., den Flux von "Mass" (W-mass) zwischen jede 2 Zustände ist neutral (Null).

Satz 10) Detailed Balance von μ erfüllt
$$\Rightarrow \mu \text{ ist stationär.}$$

Beweis: Zu zeigen: $\mu P = \mu$.
$$(\mu P)(x) = \sum_{y\in S} \mu(y)P(y,x) \overset{\text{Det. Bal.}}{=} \sum_{y \in S} \mu(x) P(x,y)$$
$$\overset{P \text{ stoch.}}{=} \mu(x). \quad \#$$

Bem: μ stationär $\not\Rightarrow$ μ erfüllt Detailed Balance.

Beispiel:



$q := 1-p \in (0,1)$

Wegen Translationsinvarianz ist es klar, dass
$$\mu = \left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right) \quad \text{ist stationär.}$$

Aber, $\mu(1)P(1,2) = \frac{1}{3}\cdot p \neq \frac{1}{3}\cdot q = \mu(2)P(2,1)$ für alle $p \neq \frac{1}{2}$.

Weitere Beispiele: Siehe Eberle's Skript.
(insbesonderes, Beispiel d: Random Walks auf Graphen)

- Der Rest dieser Kapitel ist nicht handgeschrieben:
Siehe Seiten 87 bis 96 von Eberle's Skript.
Die wichtige Inhalte sind:

Kap 3.4 { → Metropolis-Algorithmus
→ Gibbs-Sampler

Kap 3.5 { → Konvergenz ins Gleichgewicht (Satz 3.16)