Wahrscheinlichkeitstheorie, Statistik und Stochastische Prozesse 2021W

In diesem Dokument wurden die Wiederholungsfragen des Skripts vom März 2022 ausgearbeitet, sowie mit zusätzlichen nicht in den Wiederholungsfragen erwähnten Sätzen und Definitionen ergänzt. Die meisten Antworten beruhen auf Erklärungen aus dem Skriptum, aber es gibt auch einige die auf anderen Quellen wie Wikipedia oder Studyflix aufbauen.

Das Themengebiet Informationstheorie wurde nicht behandelt, da es zum jetzigen Zeitpunkt nicht abgefragt wurde.

Ein empfehlenswerter Kurs für das Verständnis der Basics von Wahrscheinlichkeitstheorie sowie teilweise von Markovketten bietet der folgende Kurs aus dem MIT-OPECOURSEWARE

https://ocw.mit.edu/resources/res-6-012-introduction-to-probability-spring-2018/index.htm

Grundlagen der Wahrscheinlichkeitstheorie

Grundlagen

1. <u>Was ist ein Wahrscheinlichkeitsraum?</u>

Ein Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ besteht aus der *Grundmenge* Ω und dem Wahrscheinlichkeitsmaß $\mathbb{P}$

. Die Elemente $\omega \in \Omega$ heißen *Elementarereignisse,* Teilmengen $A \subseteq \Omega$ heißen *Ereignisse*. Das Wahrscheinlichkeitsmaß $\mathbb{P}$ ist auf der Menge dieser Ereignisse definiert.

2. Was ist ein Ereignis?

Eine Teilmenge $A \subseteq \Omega$ der Grundmenge.

3. Was besagen die Axiome von Kolmogorov?

1. $0 \leq \mathbb{P}(A) \leq 1$.
2. $\mathbb{P}(\emptyset) = 0$.
3. $\mathbb{P}(\Omega) = 1$.
4. (Additivität) Wenn A und B disjunkte Ereignisse ($A \cap B = \emptyset$) sind, dann gilt

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

4a. (Abzählbare Additivität, Sigmaadditivität) Wenn $A_n, n \in \mathbb{N}$ disjunkte Ereignisse sind (d.h., $A_i \cap A_j = \emptyset, i \neq j$), dann gilt

$$\mathbb{P}(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

4. Welche Eigenschaften von Wahrscheinlichkeitsmaßen kennen Sie?

Zum einen die Eigenschaften, welche sich aus den Axiomen von Kolmogorov ergeben,
$0 \leq \mathbb{P}(A) \leq 1$,
$\mathbb{P}(0) = 0$,
$\mathbb{P}(\Omega) = 1$,
$\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$
und wenn $A \subseteq B$, dann gilt $\mathbb{P}(A) \leq \mathbb{P}(B)$
zum anderen, sollte es sich um einen Laplaceschen Wahrscheinlichkeitsraum handeln, dann gilt
$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$

5. Wie kann man die Wahrscheinlichkeit einer Vereinigung berechnen?

Sollten die beiden Ereignisse disjunkt sein: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
Sollten diese nicht disjunkt sein: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
Allgemein kann also das Additionstheorem angewendet werden:

Satz 2.2: Additionstheorem

$A_1, \ldots A_n$ seien beliebige Ereignisse. Dann gilt

$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} (-1)^{i-1} S_i$$

mit

$$S_i = \sum_{1 \leq j_1 < j_2 < \ldots < j_i \leq n} \mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_i}).$$

6. Wie ist die bedingte Wahrscheinlichkeit definiert?

Nimmt man an, das in N Versuchen B $N \cdot \mathbb{P}(B)$ mal eintritt und bei diesen $N \cdot \mathbb{P}(A \cap B)$ mal A eintritt so kann auf folgende Definition gefolgert werden.

Definition 2.4

A und B seien zwei Ereignisse mit $\mathbb{P}(B) > 0$. Dann heißt

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

die bedingte Wahrscheinlichkeit von A unter (der Bedingung) B.

7. <u>Wie kann man die Wahrscheinlichkeit eines Durchschnitts berechnen?</u>

Der Durchschnitt kann zum einen durch die Umformung der bedingten Wahrscheinlichkeit berechnet werden:

$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B) = \mathbb{P}(A)\mathbb{P}(B|A)$

Wodurch auf den Multiplikationssatz gefolgert werden kann:

Satz 2.3: Multiplikationssatz

$A_1, \ldots A_n$ seien Ereignisse mit $\mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$. Dann gilt

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \ldots \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

Dieser kann auch verstanden werden wie das Produkt aller bedingten Wahrscheinlichkeiten nachdem ein oder mehrere Ereignisse eingetroffen sind. (Einfach darstellbar als ein Pfad in einem Baumdiagramm)

8. <u>Wann heißen zwei Ereignisse unabhängig?</u>

Zwei Ereignisse heißen unabhängig wenn der Durchschnitt dieser das Produkt ihrer Wahrscheinlichkeiten ist:

Definition 2.5: Unabhängigkeit

Zwei Ereignisse A und B heißen unabhängig, wenn

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Die Ereignisse $A_1, \ldots, A_n$ heißen unabhängig, wenn für alle $k \leq n$ und $1 \leq i_1 < i_2 \ldots < i_k \leq n$ gilt

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_k}).$$

Die Ereignisse $A_1, \ldots, A_n$ heißen paarweise unabhängig, wenn für alle $1 \leq i < j \leq n$

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

Unendlich viele Ereignisse nennen wir unabhängig, wenn jede endliche Auswahl daraus unabhängig ist.

8. <u>Wann heißen mehr als zwei Ereignisse unabhängig?</u>

Siehe oben

9. <u>Wann heißen mehr als zwei Ereignisse paarweise unabhängig?</u>

Siehe oben

Vollständige Wahrscheinlichkeit und Bayes

10. Was besagt der Satz von der vollständigen Wahrscheinlichkeit?

Mithilfe dieses Satzes kann die vollständige Wahrscheinlichkeit eines Ereignisses A über die Ereignisse B_i berechnet werden. Anschaulich kann man sich vorstellen, dass alle "Pfadwahrscheinlichkeiten" eines Baumes, welche zum Ereignis A führen addiert werden müssen.

Satz 2.4: Satz von der vollständigen Wahrscheinlichkeit

B_i seien disjunkte Ereignisse mit $\mathbb{P}(B_i) > 0$ und $\bigcup_i B_i = \Omega$ und A ein beliebiges Ereignis. Dann gilt

$$\mathbb{P}(A) = \sum_i \mathbb{P}(B_i)\mathbb{P}(A|B_i).$$

11. Was besagt der Satz von Bayes?

Der Satz von Bayes erlaubt in gewissem Sinn das Umkehren einer Schlussfolgerung. Ein Wert $\mathbb{P}(A|B_j)$ ist bekannt, man will aber den eigentlich den Wert $\mathbb{P}(B_j|A)$ wissen, so kann das mithilfe des Satzes von Bayes gemacht werden, wenn die folgenden drei Werte bekannt sind:
- $\mathbb{P}(B_j)\ldots$ Die vollständige Wahrscheinlichkeit des Ereignisses B_j,
- $\mathbb{P}(A) = \sum_i \mathbb{P}(B_i)\mathbb{P}(A|B_i)\ldots$ Die vollständige Wahrscheinlichkeit des Ereignisses A,
- $\mathbb{P}(A|B_j)\ldots$Die bedingte Wahrscheinlichkeit von A zu B_j.

Satz 2.5: Satz von Bayes

Unter denselben Voraussetzungen wie im vorigen Satz gilt (wenn der Nenner positiv ist)

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j)\mathbb{P}(A|B_j)}{\sum_i \mathbb{P}(B_i)\mathbb{P}(A|B_i)}.$$

Man kann ihn auch mit der bedingten Wahrscheinlichkeit sowie dem Satz der vollständigen Wahrscheinlichkeit herleiten.

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j \cap A)}{\mathbb{P}(A)} = \frac{\frac{\mathbb{P}(B_j \cap A)}{\mathbb{P}(B_j)} \cdot \mathbb{P}(B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}{\mathbb{P}(A)}$$

Zufallsvariablen

12. Was ist eine Zufallsvariable?

Eine Zufallsvariable ist eine Art Funktion die jedem Ergebnis ω eines Zufallsexperiments eine Zahl x zuordnet. Dabei beschreibt X die tatsächliche Zufallsvariable, welche noch keinen Wert hat und x das Ergebnis nach dem Experiment. Eine Zufallsvariable kann auch mehrdimensional sein (auch Zufallsvektor genannt): $X = (X_1, \ldots, X_d)$

13. Welche Typen von Zufallsvariablen gibt es?

Es gibt stetige sowie diskrete Zufallsvariablen.

14. Wie kann man die Verteilung einer Zufallsvariable angeben?

Die Verteilung einer Zufallsvariable bezeichnet das Wahrscheinlichkeitsmaß mit welchem die von der Zufallsvariablen bestimmten Ereignisse auftreten.

Definition 2.7

Die Verteilung einer Zufallsvariable ist das Wahrscheinlichkeitsmaß

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega : X(\omega) \in A\})(A \subseteq \mathbb{R}^d).$$

Definition 2.8

Wenn der Wertebereich von X endlich oder höchstens abzählbar ist, dann nennen wir X diskret.

In diesem Fall kann die Verteilung von X durch die Wahrscheinlichkeitsfunktion

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}_X(\{x\})$$

angegeben werden.

15. <u>Welche Eigenschaften hat eine Verteilungsfunktion?</u>

Ist eine Zufallsvariable nicht diskret so wird die Verteilungsfunktion von X wie folgt angegeben

$F_X(x) = \mathbb{P}((-\infty, x]) = \mathbb{P}(X \leq x)$

und kann für eine Dimension wie folgt charakterisiert werden:

Satz 2.6

$F : \mathbb{R} \to \mathbb{R}$ ist genau dann eine Verteilungsfunktion, wenn

1. $0 \leq F(x) \leq 1$ für alle x,
2. F ist monoton nichtfallend,
3. F ist rechtsstetig,
4. $\lim_{x \to -\infty} F(x) = 0$,
5. $\lim_{x \to \infty} F(x) = 1$.

Handelt es sich um eine mehrdimensionale Funktion, so bleiben annähernd alle Eigenschaften gleich, wobei noch eine zusätzliche dazukommt:

Satz 2.7

$F : \mathbb{R}^2 \to \mathbb{R}$ ist genau dann eine Verteilungsfunktion, wenn

1. $0 \leq F(x_1, x_2) \leq 1$ für alle x_1, x_2,
2. F ist monoton nichtfallend in jeder Argumentvariable,
3. F ist rechtsstetig,
4. $\lim_{x_1 \to -\infty} F(x_1, x_2) = \lim_{x_2 \to -\infty} F(x_1, x_2) = 0$,
5. $\lim_{x_1, x_2 \to \infty} F(x_1, x_2) = 1$,
6. Für $a_1 < b_1, a_2 < b_2$ gilt

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0.$$

Dabei beschreibt Punkt 6 die Wahrscheinlichkeit
$\mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2)$

16. Welche Eigenschaften hat eine Wahrscheinlichkeitsfunktion?
Summiert man die bei allen Abszissenwerten vorkommenden Werte muss 1 herauskommen.

Die Funktion p ist genau dann eine Wahrscheinlichkeitsfunktion, wenn

$$p(x) \geq 0 \text{ für alle } x \in \mathbb{R}^d \text{ und } \sum_x p(x) = 1.$$

Dichtefunktion

17. Welche Eigenschaften hat eine Dichtefunktion?
Eine Dichtefunktion beschreibt das selbe wie eine Wahrscheinlichkeitsfunktion nur für stetige Zufallsvariablen und besitzt damit die selben Eigenschaften.

Die Funktion f ist genau dann eine Dichtefunktion, wenn

$$f(x) \geq 0 \text{ für alle } x \in \mathbb{R}^d \text{ und } \int f(x)dx = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1 \dots, x_d)dx_1 \dots dx_d = 1.$$

Weiters kann die Verteilungsfunktion einer Zufallsvariable auf folgende Weise aus der Dichte entnommen werden.

Definition 2.14

Wenn F_X in der Form

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

bzw.

$$F_X(x) = \int_{-\infty}^{x_d} \dots \int_{-\infty}^{x_1} f_X(u_1, \dots, u_d)du_1 \dots du_d$$

(falls $X = (X_1, \dots, X_d)$ mehrdimensional ist) darstellbar ist, dann heißt f_X die Dichte der Verteilung von X, und wir nennen X stetig (verteilt).

18. Wann heißen zwei/mehrere Zufallsvariable unabhängig?
Mehrere Zufallsvariablen heißen unabhängig, wenn die gemeinsame Verteilungsfunktion das Produkt der einzelnen Randverteilungen ist.

Definition 2.17

Die Zufallsvariablen $(X_1, \dots, X_n)$ heißen unabhängig, wenn für alle $x_1, \dots x_n$

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i).$$

Die unendliche Folge $(X_n, n \in \mathbb{N})$ heißt unabhängig, wenn jede endliche Teilfolge unabhängig ist.

19. Was versteht man unter einer "Randverteilung"?

Wenn die Verteilung einer Zufallsvariable X (oder Y) aus der gemeinsamen Verteilung ermittelt wird, nennt man sie Randverteilung. Diese können beispielsweise über die gemeinsame Dichte oder Wahrscheinlichkeitsfunktion bestimmt werden.

Wenn X und Y eine gemeinsame Verteilung mit der Dichte $f_{X,Y}(x,y)$ haben, dann können wir die Verteilungsfunktion von X wie folgt berechnen:

$$F_X(x) = \mathbb{P}(X \leq x, -\infty < Y < \infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx,$$

und durch Differenzieren ergibt sich die Dichte von X (und analog für Y) als

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy, f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$

Für diskrete Verteilungen gilt

$$p_X(x) = \sum_y p_{X,Y}(x,y), p_Y(y) = \sum_x p_{X,Y}(x,y).$$

Wahrscheinlichkeitsfunktion der Summe unabhängiger Zufallsvariablen

Satz 2.13

X und Y seien unabhängig mit Wahrscheinlichkeitsfunktion p_X und p_Y. Dann ist die Wahrscheinlichkeitsfunktion von $X+Y$ die (diskrete) Faltung von p_X und p_Y:

$$p_{X+Y}(z) = p_X * p_Y(z) = \sum_x p_X(x)p_Y(z-x).$$

20. Wie kann man die Dichte einer transformierten Zufallsvariable bestimmen?

Satz 2.10: Transformationssatz für Dichten

$X = (X_1, \ldots, X_n)$ sei stetig verteilt mit der Dichte f_X. $g : \mathbb{R}^n \to \mathbb{R}^n$ sei stetig differenzierbar und eindeutig umkehrbar. $Y = g(X)$ (d.h. $Y_i = g_i(X_1, \ldots, X_n)$) ist dann ebenfalls stetig verteilt mit der Dichte

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{\partial g^{-1}}{\partial y}(y)| = f_X(g^{-1}(y))\frac{1}{|\frac{\partial g}{\partial x}(g^{-1}(y))|} & \text{wenn } y \in g(\mathbb{R}^n), \\ 0 & \text{sonst.} \end{cases}$$

Dabei ist

$$\frac{\partial g}{\partial x} = \det((\frac{\partial g_i}{\partial x_j})_{n \times n})$$

die Funktionaldeterminante.

21. Wie berechnet man die bedingte Dichte?

Definition 2.18

X, Y seien stetig verteilt mit Dichte $f_{X,Y}$. Die bedingte Dichte von X unter $Y = y$ ist

$$f_X(x|Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Damit erhalten wir die bedingte Wahrscheinlichkeit als

$$\mathbb{P}(X \leq a|Y=y) = \int_{-\infty}^{a} f_X(x|Y=y)dx.$$

Unabhängige Dichten

Satz 2.11

X und Y seien unabhängig mit Dichte f_X und f_Y. Dann ist die Dichte von $X+Y$

$$f_{X+Y}(z) = f_X * f_Y(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx.$$

Definition 2.19

f und g seien zwei Dichten. Die Dichte $f * g$ mit

$$f * g(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx.$$

heißt die Faltung von f und g.

Satz 2.12

X und Y seien unabhängig mit Dichte f_X und f_Y. Dann hat $Z = X/Y$ die Dichte

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y)f_X(zy)|y|dy.$$

p-Quantil

Inverse Verteilungsfunktion und Quantil

Das p-Quantil ist der Mittelpunkt (oder einer der Endpunkte) in denen $F(x) = p$ gilt.

Definition 2.20: Quantil

x heißt p-Quantil $(0 > p > 1)$ der Verteilung mit der Verteilungsfunktion F, wenn

$$F(x-0) \leq p \leq F(x)$$

(d.h., wenn $X \sim F$, dann ist $\mathbb{P}(X < x) \leq p \leq \mathbb{P}(X \leq x)$).

Definition 2.21

Die verallgemeinerte Inverse der Verteilungsfunktion F ist gegeben durch

$$F^{-1}(y) = \inf\{x : F(x) \geq y\}.$$

Satz 2.14

Wenn U auf $[0,1]$ gleichverteilt ist, dann hat

$$X = F^{-1}(U)$$

die Verteilungsfunktion F.

Erwartungswert und Varianz

22. <u>Wie ist der Erwartungswert einer Zufallsvariable definiert?</u>

Der Erwartungswert ist das Ergebnis eines Zufallsexperiments, welches man im Mittel erwarten kann.

Definition 2.22

Der Erwartungswert einer Zufallsvariable X ist

$$\mathbb{E}(X) = \sum_x x p_X(x)$$

für diskrete Zufallsvariable, und

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

für stetige Zufallsvariable. Falls X gemischt verteilt ist, gilt

$$\mathbb{E}(X) = \sum_x x p_X(x) + \int_{-\infty}^{\infty} x f_X(x) dx.$$

23. <u>Wie ist die Varianz einer Zufallsvariable definiert?</u>

Die Varianz bezeichnet die quadratische Standardabweichung vom Mittelwert.

Definition 2.23

Die Varianz von X ist

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

Eine andere Art und Weise wie man die Varianz berechnen kann ergibt sich aus der Anwendung der **Steinerschen Verschiebungssatzes**

Satz 2.17: Steinerscher Verschiebungssatz

Für beliebiges reelles a gilt

$$\mathbb{E}((X-a)^2) = \mathbb{V}(X) + (\mathbb{E}(X) - a)^2.$$

Setzt man in dieser Gleichung das $a = 0$ ergibt sich

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

24. Wie ist die Kovarianz von zwei Zufallsvariablen definiert?

Die Kovarianz beschreibt den linearen Zusammenhang zwischen zwei Zufallsvariablen (wie sich die Vorzeichen der beiden Zufallsvariablen Durchschnittlich zu einander liegen). Es wird der Erwartungswert des Produktes der beiden Differenzen zwischen den jeweiligen Zufallsvariablen und ihren Erwartungswerten gebildet.

Definition 2.25

X und Y seien Zufallsvariable mit endlicher Varianz. Dann heißt

$$\mathbf{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

die Kovarianz von X und Y.

Teile der Kovarianz lassen sich besser mit der **Cauchy-Schwarz Ungleichung** bestimmen.

Satz 2.20: Cauchy-Schwarz Ungleichung

Wenn $\mathbb{E}(X^2)$ und $\mathbb{E}(Y^2)$ endlich sind, dann gilt

$$\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Gleichheit gilt genau dann, wenn es ein $a > 0$ gibt, sodass $Y = aX$.

Außerdem kann die Kovarianz dazu verwendet werden den Korrelationskoeffizienten von zwei Zufallsvariablen zu bestimmen.

Definition 2.26

X und Y seien zwei Zufallsvariable mit positiver endlicher Varianz. Dann heißt

$$\rho(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

der Korrelationskoeffizient von X und Y.

Satz 2.19

Es gilt

$$-1 \leq \rho(X,Y) \leq 1.$$

$\rho(X,Y) = 1$ genau dann, wenn es Konstante $a > 0$ und b gibt mit $Y = aX + b$.
$\rho(X,Y) = -1$ genau dann, wenn es Konstante $a < 0$ und b gibt mit $Y = aX + b$.
$\rho(X,Y) = 0$ genau dann, wenn X und Y unkorreliert sind.

Das heißt wenn der Betrag des Korrelationskoeffizienten 1 ist, gibt es einen linearen Zusammenhang zwischen den beiden Zufallsvariablen.

25. Welche Eigenschaften des Erwartungswerts kennen Sie?

Satz 2.15: Eigenschaften des Erwartungswerts

1. Linearität: $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$,
2. Additivität: $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,
3. Monotonie: wenn $X \leq Y$, dann ist auch $\mathbb{E}(X) \leq \mathbb{E}(Y)$,
4. wenn X und Y unabhängig sind, dann gilt $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Um das Rechnen mit mehreren Zufallsvariablen zu erleichtern kann weiters der **Satz des Unachtsamen Statistikers** verwendet werden.

X sei eine Zufallsvariable, $g : \mathbb{R} \to \mathbb{R}$ eine Funktion und$Y = g(X)$. Dann ist

1. Wenn X diskret verteilt ist mit Wahrscheinlichkeitsfunktion p_X,

$$\mathbb{E}(Y) = \sum_x g(x) p_X(x).$$

2. Wenn X stetig verteilt ist mit Dichte f_X,

$$\mathbb{E}(Y) = \int g(x) f_X(x) dx.$$

Diese Formel gilt sinngemäß auch für mehrdimensionales $X \in \mathbb{R}^k$ und $g : \mathbb{R}^k \to \mathbb{R}$:

$$\mathbb{E}(Y) = \int g(x) f_X(x) dx = \int \dots \int g(x_1, \dots, x_k) f_{X_1,\dots,X_k}(x_1, \dots, x_k) dx_1 \dots dx_k.$$

3. Wenn X gemischt verteilt ist mit Wahrscheinlichkeitsfunktion p_X und Dichte f_X,

$$\mathbb{E}(Y) = \sum_x g(x) p_X(x) + \int g(x) f_X(x) dx.$$

Weiters gibt es für den Erwartungswert (sowie die Varianz Standardwerte):

Definition 2.24

Wenn $\mathbb{E}(X) = 0$, dann heißt X zentriert.
Wenn $\mathbb{E}(X^2) = 1$, dann heißt X normiert.
Wenn $\mathbb{E}(X) = 0$ und $\mathbb{V}(X) = 1$, dann heißt X standardisiert.
Wenn $\mu = \mathbb{E}(X)$ endlich ist, dann ist $X^o = X - \mu$ zentriert (die Zentrierung von X).
Wenn $\mathbb{E}(X^2)$ endlich und positiv ist, dann ist $X^* = \frac{X}{\sqrt{\mathbb{E}(X^2)}}$ normiert (die Normierung von X).
Wenn $\sigma^2 = \mathbb{V}(X)$ endlich und positiv ist, dann ist $X^{o*} = \frac{X-\mu}{\sigma}$ standardisiert (die Standardisierung von X).

26. Welche Eigenschaften der Varianz kennen Sie?

Satz 2.18: Eigenschaften der Varianz

1. $\mathbb{V}(X) \geq 0$,
2. $\mathbb{V}(X) = 0$ genau dann, wenn $\mathbb{P}(X = \mathbb{E}(X)) = 1$,
3. $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$,
4. wenn X und Y unabhängig sind, dann ist $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.

Markov, Chebychev und Kolmogorov

27. Was besagt die Ungleichung von Markov?

Sie beschreibt mit höchstens welcher Wahrscheinlichkeit eine Zufallsvariable größer als ein Wert λ ist.

Satz 2.21: Ungleichung von Markov

X sei eine nichtnegative Zufallsvariable, $\lambda > 0$. Dann ist

$$\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda}\mathbb{E}(X).$$

Diese kann auch über die Momente einer Zufallsvariablen bestimmt werden.

$$P(|X| \geq x) \leq \frac{M_k}{x^k}.$$

28. Was besagt die Ungleichung von Chebychev?

Sie beschreibt die höchste Wahrscheinlichkeit, dass die absolute Differenz zwischen einer Zufallsvariable und ihrem Erwartungswert über einem Wert λ liegt.

Satz 2.22: Ungleichung von Chebychev

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{\mathbb{V}(X)}{\lambda^2}.$$

Genauso wie im Fall der Markov-Ungleichung kann auch die Chebychev-Ungleichung über das zweite Moment (welches die Varianz beschreibt) beschrieben werden.

29. Was besagt die Ungleichung von Kolmogorov?

Sie besagt, dass die größte Wahrscheinlichkeit einer Partialsumme mehrerer unabhängiger Variablen größer als ein Wert λ ist, kleiner als die Varianz dieser Partialsumme geteilt durch λ^2 ist.

Satz 2.23: Ungleichung von Kolmogorov

$X_1, \ldots, X_n$ seien unabhängig mit Erwartungswert 0, $S_0 = 0$, $S_n = X_1 + \ldots + X_n$. Dann ist

$$\mathbb{P}(\max_{k \leq n} |S_k| \geq \lambda) \leq \frac{\mathbb{V}(S_n)}{\lambda^2}.$$

Momente

30. Wie sind die Momente einer Zufallsvariable definiert?

Das nte-Moment einer Zufallsvariablen X bezeichnet den Erwartungswert von X^n

Definition 2.27

$$M_n(X) = \mathbb{E}(X^n)$$

heißt das n-te Moment von X,

$$m_n(X) = \mathbb{E}((X - \mathbb{E}(X))^n)$$

das n-te zentrale Moment von X.

Betrachtet man beispielsweise die ersten 3 zentralen Momente einer Zufallsvariablen (das heißt dass die Verteilung dem Erwartungswert 0 besitzt), so können einige Eigenschaften bemerkt werden.

- $n = 0$: Erwartungswert, muss also 0 sein
- $n = 1$: Mittlere absolute Abweichung
- $n = 2$: Varianz
- $n = 3$: nach Normierung mit der Standardabweichung ergibt dieses die Schiefe

31. Was ist die momentenerzeugende Funktion einer Zufallsvariable?

Definition 2.29

Die Funktion

$$M_X(t) = \mathbb{E}(e^{Xt})$$

heißt die *momentenerzeugende Funktion* von X.

Sie kann mittels Ableitung zur Berechnung der Momente der Zufallsvariable verwendet werden. Die k-te Ableitung im Punkt $t = 0$ ist gleich dem k-ten Moment.

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k} M_X(t)\bigg|_{t=0} = E(X^k) = m_X^k.$$

Das sieht man besser indem man die Potenzreihe des Ausdrucks bildet.

$$M_X(t) = E\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n) = \sum_{n=0}^{\infty} \frac{t^n}{n!} m_X^n.$$

32. Was ist die wahrscheinlichkeitserzeugende Funktion einer diskreten Zufallsvariable?

Definition 2.28

X sei eine diskrete Zufallsvariable mit nichtnegativen ganzzahligen Werten. Dann heißt die Funktion

$$p_X^*(z) = \mathbb{E}(z^X) = \sum_{n=0}^{\infty} z^n p_X(n)$$

(für reelles oder auch komplexes z) die *wahrscheinlichkeitserzeugende Funktion* von X.

Jeder diskreten Verteilung (in $\mathbb{N}$) sowie jeder Zufallsvariablen (mit Werten in $\mathbb{N}$) kann eine wahrscheinlichkeitserzeugende Funktion zugeordnet werden. Dies gilt auch umgekehrt. Dadurch können einige Eigenschaften von Zufallsvariablen auf Funktionen übertrage werden.

- Faltung von Verteilungen entspricht Multiplikation von wahrscheinlichkeitserz. Funktionen
- Es gibt eine Beziehung zwischen Ableitung der Funktion und $\mathbb{E}$, $\mathbb{V}$ und den Momenten.

33. Was ist die charakteristische Funktion einer Zufallsvariable?

Definition 2.30

Die charakteristische Funktion ϕ_X der Zufallsvariable X ist gegeben durch

$$\phi_X(t) = \mathbb{E}(e^{iXt}) = \mathbb{E}(\cos(Xt)) + i\mathbb{E}(\sin(Xt)), t \in \mathbb{R}.$$

Eigenschaften der momentenerzeugenden und charakteristischen Funktion

Satz 2.24

F[r die momentenerzeugende Funktion und die charakteristische Funktion gilt:

1. Wenn X und Y unabhängig sind, dann gilt

$$M_{X+Y}(t) = M_X(t)M_Y(t),$$

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

2. Für $Y = aX + b$, $a, b \in \mathbb{R}$:

$$M_Y(t) = e^{bt}M_X(at),$$

$$\phi_Y(t) = e^{ibt}\phi_X(at).$$

3. Wenn $m_k = \mathbb{E}(X^k)$ endlich ist, dann ist ϕ_X k-mal differenzierbar, und

$$\phi_X^{(k)}(0) = i^k m_k.$$

Wenn wir zusätzlich annehmen, dass $M_X(t)$ für ein $t \neq 0$ endlich ist, dann gilt auch

$$M_X^{(k)}(0) = m_k.$$

(unter Umständen, wenn die Momentenerzeugende nur auf einer Seite der Null endlich ist, ist hier die einseitige Ableitung zu verwenden. Wenn das nicht der Fall ist, wenn es also negative und positive Werte von t gibt, für die $M_X(t)$ endlich ist, dann sind auch alle Momente m_k endlich).

Grenzwertsätze

34. Was besagt das schwache Gesetz der großen Zahlen?

Das Stichprobenmittel S_n/n einer Folge von Zufallsvariablen konvergiert in Wahrscheinlichkeit gegen den Erwartungswert μ.

Satz 2.25: schwaches Gesetz der großen Zahlen

$(X_n, n \in \mathbb{N})$ sei eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit Erwartungswert μ und endlicher Varianz, $S_n = X_1 + \ldots + X_n$. Dann gilt

$$\mathbb{P}(|\frac{S_n}{n} - \mu| \geq \epsilon) \to 0$$

für jedes $\epsilon > 0$.

Definition 2.31

Wir betrachten Zufallsvariable $X_n, n \in \mathbb{N}$ und X. Die Folge X_n konvergiert gegen X

1. in Wahrscheinlichkeit, wenn für alle $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

2. mit Wahrscheinlichkeit 1 (oder "fast sicher"), wenn

$$\mathbb{P}(X_n \text{ konvergiert gegen } X) = 1.$$

3. in Verteilung, wenn für alle Stetigkeitspunkte x von F_X

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

35. Was besagt das starke Gesetz der großen Zahlen?

Satz 2.26: Starkes Gesetz der großen Zahlen

$(X_n, n \in \mathbb{N})$ sei eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit endlichem Erwartungswert $\mu = \mathbb{E}(X_n)$, $S_n = X_1 + \ldots + X_n$. Dann konvergiert $\frac{S_n}{n}$ mit Wahrscheinlichkeit 1 gegen μ.

36. Was besagt der zentrale Grenzwertsatz?

Dass die Verteilung einer Folge an Zufallsvariablen mit genügender Größe durch eine Normalverteilung beschrieben werden kann.

Satz 2.27: Zentraler Grenzwertsatz

$(X_n, n \in \mathbb{N})$ sei eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $\mathbb{E}(X) = \mu$, $\mathbb{V}(X) = \sigma^2$. Dann ist S_n näherungsweise normalverteilt mit Mittel $n\mu$ und Varianz $n\sigma^2$, d.h., es gilt

$$\lim_{n\to\infty} \mathbb{P}(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

Satz von deMoivre-Laplace

Für große n kann die Binomialverteilung $B(n,p)$ durch die Normalverteilung $N(np, np(1-p))$ approximiert werden. Das gilt sowohl für die Verteilungs- als auch Wahrscheinlichkeitsfunktion.

Satz 2.28: deMoivre-Laplace

X_n sei binomialverteilt: $X_n \sim B(n,p)$. Dann gilt

$$\mathbb{P}(X_n = k) = \frac{1}{\sqrt{2n\pi p(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}} + o(\frac{1}{\sqrt{n}})$$

gleichmäßig in k für $n \to \infty$.

Verschiedene Verteilungen

Hypergeometrische Verteilung

Definition 2.9

Die hypergeometrische Verteilung $H(N,A,n)$ $(n, A, N \in \mathbb{N},\ 0 \leq n, A \leq N)$ hat die Wahrscheinlichkeitsfunktion

$$p(x) = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}}.$$

Binomialverteilung

Definition 2.10

Die Binomialverteilung $B(n,p)$:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Alternativverteilung

Definition 2.11

X heißt alternativverteilt $(X \sim A(p))$ wenn

$$\mathbb{P}(X=1) = p, \mathbb{P}(X=0) = 1-p.$$

Stetige Gleichverteilung

- Die stetige Gleichverteilung $U(a,b)$:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{wenn } a \leq x \leq b, \\ 0 & \text{sonst.} \end{cases}$$

Exponentialverteilung

- Die Exponentialverteilung $Ex(\lambda)$:

$$f(x) = \lambda e^{-\lambda x}[x \geq 0].$$

(Standard-)Normalverteilung

- Die Normalverteilung (eine der wichtigsten Verteilungen) $N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}).$$

Der Spezialfall $N(0,1)$ (also $\mu = 0, \sigma^2 = 1$) wird als Standardnormalverteilung bezeichnet, für ihre Dichte bzw. Verteilungsfunktion haben sich die Bezeichnungen φ und Φ eingebürgert, also

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \Phi(x) = \int_{-\infty}^{x} \phi(u) du.$$

Gemischte Verteilung

Definition 2.15

Wenn F_X sowohl Sprünge als auch eine nichtverschwindende Ableitung hat, dann nennen wir X gemischt verteilt. In diesem Fall gibt es sowohl eine Wahrscheinlichkeitsfunktion als auch eine Dichte.

Statistik

Statistisches Modell

1. Woraus besteht ein statistisches Modell?

Definition 5.1

Ein statistisches Modell ist eine Menge $\mathcal{P}$ von Verteilungen. Wenn diese Verteilungen durch endlich viele reelle Zahlen (die Parameter) beschrieben werden können, also

$$\mathcal{P} = \{P_\theta, \theta \in \Theta\}$$

mit $\Theta \subseteq \mathbb{R}^d$, dann sprechen wir von einem parametrischen Modell, sonst von einem nichtparametrischen Modell.

Eine Stichprobe ist eine Folge $(X_1, \ldots, X_n)$ von unabhängigen Zufallsvariablen mit einer (unbekannten) Verteilung aus $\mathcal{P}$.

2. Was ist der Unterschied zwischen einem parametrischen und einem nichtparametrischen Modell?

Lässt sich die Menge von Wahrscheinlichkeitsmaßen über einer Parametermenge beschreiben, dann spricht man von einem parametrischen Modell, ansonsten von einem nichtparametrischen.

3. Geben Sie Beispiele für statistische Modelle an.

Alternativverteilung und Normalverteilung.

4. Was ist eine Statistik?

Definition 5.2

Eine Statistik T ist eine Zufallsvariable, die aus der Stichprobe berechnet werden kann:

$$T = T(X_1, \ldots, X_n)$$

Insbesondere dürfen in dieser Funktion die unbekannten Parameter nicht vorkommen.

Schätzer und Konfidenzintervalle

5. Was ist ein Schätzer?

Definition 5.3

Ein Schätzer ist eine Folge $(\hat{\theta}_n)$ von Statistiken $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$.

6. Welche Eigenschaften können wir von einem Schätzer verlangen?

Definition 5.4

Ein Schätzer $\hat{\theta}_n$ heißt

- schwach konsistent, wenn $\hat{\theta}_n \to \theta$ in Wahrscheinlichkeit konvergiert (also $\mathbb{P}_\theta(|\hat{\theta}_n - \theta| > \epsilon) \to 0$ für alle $\epsilon > 0$).
- stark konsistent, wenn $\hat{\theta}_n \to \theta$ mit Wahrscheinlichkeit 1 konvergiert (also $\mathbb{P}_\theta(\hat{\theta}_n \to \theta) = 1$.
- erwartungstreu, wenn für alle $\theta \in \Theta$

$$\mathbb{E}_\theta(\hat{\theta}_n) = \theta.$$

- effizient, wenn er erwartungstreu ist und die kleinste Varianz unter allen erwartungstreuen Schätzern hat, also wenn für jeden weiteren erwartungstreuen Schätzer $\tilde{\theta}_n$

$$\mathbb{V}_\theta(\hat{\theta}_n) \leq \mathbb{V}_\theta(\tilde{\theta}_n) \text{ für alle } \theta. \tag{5.1}$$

7. Wie sind Stichprobenmittel und Stichprobenvarianz definiert? Warum wird bei der Stichprobenvarianz durch n − 1 dividiert und nicht durch n?

Beispiel 5.4: Das Stichprobenmittel

Wir betrachten eine Stichprobe aus einer Verteilung mit Erwartungswert μ (dieser ist im allgemeinen eine Funktion des Parameters θ). Das Stichprobenmittel

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

hat Erwartungswert μ, und das Gesetz der großen Zahlen sagt uns, das $\bar{X}_n$ für $n \to \infty$ mit Wahrscheinlichkeit eins gegen μ konvergiert. Das Stichprobenmittel ist also ein erwartungstreuer und konsistenter Schätzer für den Erwartungswert. Ob er effizient ist, lässt sich nicht in allgemeiner Form beantworten, die Antwort auf diese Frage hängt vom zugrundeliegenden Modell ab.

Beispiel 5.5: Die Stichprobenvarianz

Wir nehmen jetzt auch noch an, dass $\sigma^2 = \mathbb{V}(X)$ endlich ist.

Wie das Stichprobenmittel für den Erwartungswert ist

$$T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

ein erwartungstreuer und konsistenter Schätzer für $\mathbb{E}(X^2)$. Daraus folgt, dass

$$T_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = T_1 - \bar{X}_n^2$$

ein konsistenter Schätzer für die Varianz ist. Er ist allerdings nicht erwartungstreu. Sein Erwartungswert ist

$$\mathbb{E}(T_2) = \mathbb{E}(X^2) - \mathbb{E}(\bar{X}_n^2) = \mu^2 + \sigma^2 - (\mu^2 + \frac{\sigma^2}{n}) = \sigma^2\frac{n-1}{n}.$$

Für Stichprobenumfänge $n > 1$ kann man daraus einen erwartungstreuen Schätzer machen, indem man den störenden Faktor wegmultipliziert:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Diese Statistik heißt Stichprobenvarianz und ist ein erwartungstreuer Schätzer für die Varianz.

- Der Nenner $n-1$ ist für Stichproben richtig.
  Da ist der Mittelwert der Abgezogen wird nur ein Schätzwert und liegt näher an der Stichprobe als der genaue Erwartungswert.
- Der Nenner n ist korrekt, wenn der Mittelwert, der abgezogen wird, mit dem exakten Erwartungswert übereinstimmt.

Likelihoodfunktion

8. Wie ist die Likelihoodfunktion definiert?

Definition 5.6: Likelihoodfunktion

Die Likelihoodfunktion ist

$$L(X_1, \ldots, X_n; \theta) = \prod_{i=1}^{n} p_\theta(X_i),$$

wenn P_θ diskret mit Wahrscheinlichkeitsfunktion p ist, und

$$L(X_1, \ldots, X_n; \theta) = \prod_{i=1}^{n} f_\theta(X_i),$$

wenn P_θ stetig mit Dichte f ist.

9. Welche Methoden zur Konstruktion von Schätzern kennen Sie?

Definition 5.5: Momentenmethode

θ sei ein einzelner reeller Parameter (also der Parameterraum Θ eindimensional). Dann können wir den Erwartungswert von X als Funktion von θ schreiben:

$$\mathbb{E}_\theta(X) = m(\theta).$$

Wenn die Funktion m stetig umkehrbar ist, dann ist

$$\hat{\theta}_n = m^{-1}(\bar{X}_n)$$

ein (stark) konsistenter Schätzer, der Momentenschätzer.

Wenn es $d > 1$ Parameter gibt, dann verwendet man zusätzlich höhere Momente:

$$m_i(\theta) = m_i(\theta_1, \ldots, \theta_d) = \mathbb{E}_\theta(X^i).$$

Wir ersetzen die theoretischen Momente durch ihre Schätzer, das ergibt die d Gleichungen

$$m_i(\theta_1, \ldots, \theta_d) = \frac{1}{n} \sum_{j=1}^{n} X_j^i, \; i = 1, \ldots, d$$

in den d Variablen $\theta_1, \ldots, \theta_d$. Auflösen dieses Gleichungssytems nach θ ergibt den Momentenschätzer für $\theta = (\theta_1, \ldots, \theta_d)$. Wenn das Ergebnis stetig von den rechten Seiten des Gleichungssystems abhängt, erhalten wir dadurch wieder einen konsistenten Schätzer.

Definition 5.7: Maximum-Likelihood-Methode

Der Maximum-Likelihood (ML-) Schätzer ist der Wert von θ, der die Likelihoodfunktion maximiert.

Cramer-Rao und Suffizienz

10. Wie klein kann die Varianz eines erwartungstreuen Schätzers werden?

Mit der Cramer-Rao Schranke kann in vielen Fällen eine untere Abschätzung für die Varianz gefunden werden.

Satz 5.1: Cramér-Rao

Wenn p_θ bzw. f_θ zweimal nach θ differenzierbar ist und zusätzliche Regularitätsvoraussetzungen erfüllt, dann gilt für jeden Erwartungstreuen Schätzer $\hat{\theta}_n$

$$\mathbb{V}(\hat{\theta}_n) \geq \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}.$$

Dabei ist

$$I_n(\theta) = \mathbb{E}_\theta((\frac{\partial \log(L(X_1,\dots,X_n;\theta))}{\partial\theta})^2) = -\mathbb{E}_\theta(\frac{\partial^2}{\partial\theta^2}\log(L(X_1,\dots,X_n;\theta)))$$

und $I(\theta) = I_1(\theta)$, also

$$I(\theta) = -\mathbb{E}_\theta((\frac{\partial \log(f_\theta(X))}{\partial\theta})^2) = -\mathbb{E}_\theta(\frac{\partial^2}{\partial\theta^2}\log(f_\theta(X)))$$

bzw.

$$I(\theta) = -\mathbb{E}_\theta((\frac{\partial \log(p_\theta(X))}{\partial\theta})^2) = -\mathbb{E}_\theta(\frac{\partial^2}{\partial\theta^2}\log(p_\theta(X))).$$

$I(\theta)$ bzw. $I_n(\theta)$ wird als die Fisherinformation bezeichnet.

Suffizienz

Wenn eine einzelne Statistik alle Informationen enthält, welche die Stichprobe über einen Parameter liefern kann.

Definition 5.8

Die Statistik $T = T(X_1,\dots,X_n)$ ist *suffizient* für den Parameter θ, wenn die bedingte Verteilung von $(X_1,\dots,X_n)$ unter T nicht von θ abhängt.

Satz 5.2: Faktorisierungssatz von Neyman

T ist genau dann suffizient für θ, wenn die Likelihoodfunktion in der Form

$$L(X_1,\dots,X_n;\theta) = g(T,\theta)h(X_1,\dots,X_n)$$

dargestellt werden kann.

Satz 5.3

Zu jedem Schätzer $\hat{\theta}$ gibt es einen Schätzer $\tilde{\theta}$, der eine Funktion der suffizienten Statistik ist und

$$\mathbb{E}_\theta(\tilde{\theta}) = \mathbb{E}_\theta(\hat{\theta}), \mathbb{V}_\theta(\tilde{\theta}) \leq \mathbb{V}_\theta(\hat{\theta})\forall\theta$$

erfüllt. Insbesondere ist ein effizienter Schätzer eine Funktion der suffizienten Statistik.

Kofidenzintervall

11. Wie ist ein Konfidenzintervall definiert?

Das Intervall, in welchem sich ein Parameter mit einer vorgegebenen Wahrscheinlichkeit befindet.

Definition 5.9

$a = a(X_1, \dots, X_n) \leq b = b(X_1, \dots, X_n)$ seien zwei Statistiken. Das Intervall $[a, b]$ heißt Konfidenzintervall für θ mit Überdeckungswahrscheinlichkeit γ, wenn

$$\mathbb{P}_\theta(a \leq \theta \leq b) \geq \gamma.$$

Wenn in dieser Ungleichung Gleichheit gilt, sprechen wir von einem exakten Konfidenzintervall.

12. Geben Sie ein Konfidenzintervall für den Mittelwert μ einer Normalverteilung an.

Satz 5.5

Konfidenzintervalle für die Normalverteilung $N(\mu, \sigma^2)$:
Für μ, wenn σ^2 bekannt ist:

$$[\bar{X}_n - z_{\frac{1+\gamma}{2}}\sqrt{\frac{\sigma^2}{n}}, , \bar{X}_n + z_{\frac{1+\gamma}{2}}\sqrt{\frac{\sigma^2}{n}}],$$

Für μ, wenn σ^2 unbekannt ist:

$$[\bar{X}_n - t_{n-1;\frac{1+\gamma}{2}}\sqrt{\frac{S_n^2}{n}}, , \bar{X}_n + t_{n-1;\frac{1+\gamma}{2}}\sqrt{\frac{S_n^2}{n}}],$$

für σ^2:

$$[\frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{1+\gamma}{2}}}, \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{1-\gamma}{2}}}].$$

13. Geben Sie ein Konfidenzintervall für die Varianz σ2 einer Normalverteilung an.

Siehe oben.

14. Geben Sie ein approximatives Konfidenzintervall für die Erfolgswahrscheinlichkeit p einer Alternativverteilung an.

Satz 5.6

$(X_1, \dots, X_n)$ sei eine Stichprobe einer Alternativverteilung $A(p), 0 < p < 1$.

$$\hat{p} = \bar{X}_n$$

sei der (erwartungstreue und effiziente) Schätzer für p. Dann ist ein approximatives Konfidenzintervall gegeben durch

$$[\hat{p} - z_{\frac{1+\gamma}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, , \hat{p} + z_{\frac{1+\gamma}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}].$$

Tests

15. Was ist ein statistischer Test?

Man versucht mithilfe einer Stichprobe eine Aussage über die Gültigkeit einer Hypothese zu überprüfen.

16. Was ist eine Hypothese?

Definition 5.10

Eine *Hypothese* ist eine Teilmenge des Parameterraums Θ.

17. Welche Typen von Hypothesen gibt es?

Es gibt

- einseitige Hypothesen ($\theta \leq c\,\text{oder}\,\theta > c$)
- zweiseitige Hypothesen ($\theta \neq c$)
- einfache Hypothesen ($\theta = c$)

Andererseits gibt es bei Tests die Nullhypothese H_0 sowie die Gegenhypothese H_1

18. Welche Fehler können beim Testen auftreten?

- Fehler erster Art:
  Die Nullhypothese wird verworfen obwohl sie zutrifft.
- Fehler zweiter Art:
  Die Nullhypothese wird angenommen obwohl sie nicht zutrifft.

19. Was ist das Niveau (Signifikanzniveau) eines Tests?

Die Schranke für die Wahrscheinlichkeit des Eintreffens eines Fehlers erster Art.

Definition 5.11

Ein Test heißt vom Niveau α, wenn die Wahrscheinlichkeit für einen Fehler erster Art (die bei zusammengesetzten Hypothesen eine Funktion davon θ ist) nicht größer als α ist.

20. Was ist der p-Wert und wie wird er verwendet?

Der *p*-Wert gibt an, wie wahrscheinlich die Ergebnisse der Stichprobe (oder extremere Ergebnisse) sind, unter der Annahme, dass die Nullhypothese stimmt.

Definition 5.12: p-Wert

T sei eine Teststatistik, T_{stp} sei der Wert von T, der aus einer konkreten Stichprobe berechnet wurde. Der p-Wert ist die Wahrscheinlichkeit, einen Wert für T zu erhalten, der mindestens so stark gegen die Nullhypothese spricht wie der aus der Stichprobe. Was das konkret bedeutet, hängt von der Semantik von T ab:

- Wenn T einseitig ist, und H_0 verworfen wird, wenn $T > t_c$ ist, dann ist der p-Wert $\mathbb{P}(T \geq T_{stp})$,
- Wenn T in der anderen Richtung einseitig ist, also H_0 verworfen wird, wenn $T < t_c$ ist, dann ist der p-Wert $\mathbb{P}(T \leq T_{stp})$,
- Wenn T zweiseitig ist, dann ist der p-Wert $2\min(\mathbb{P}(T \geq T_{stp}), \mathbb{P}(T \leq T_{stp}))$.

Die Nullhypothese wird auf Niveau α verworfen, wenn der p-Wert kleiner als α ist.

Likelihoodquotiententest

Definition 5.13

Die Likelihoodquotientenstatistik ist

$$\ell = \frac{\sup_{\theta \in H_0} L(X_1, \ldots, X_n, \theta)}{\sup_{\theta \in H_1} L(X_1, \ldots, X_n, \theta)}$$

bzw.

$$\ell = \frac{\sup_{\theta \in H_0} L(X_1, \ldots, X_n, \theta)}{\sup_{\theta \in \Theta} L(X_1, \ldots, X_n, \theta)}$$

(die zweite Form ist oft einfacher zu berechnen, und für große Stichprobenumfänge sind die Tests identisch).

Der Likelihoodquotiententest verwirft die Nullhypothese, wenn ℓ kleiner ist als ein kritischer Wert.

Satz 5.7: Neyman-Pearson

Falls sowohl H_0 als auch H_1 einfach ist,dann ist der Likelihoodquotiententest optimal, d.h., er hat unter allen Tests mit demselben Niveau die minimale Wahrscheinlichkeit für einen Fehler zweiter Art (in diesem Fall wird der Likelihoodquotient einfach als

$$L(X_1, \ldots X_n, \theta_0)/L(X_1, \ldots X_n, \theta_1)$$

berechnet)

21. Was versteht man unter einem Anpassungstest?

Das sind Tests, welche überprüfen ob eine Stichprobe zu einer gegebenen Verteilung passt.

22. Wie wird der Chi-Quadrat-Anpassungstest durchgeführt?

1. Testhypothesen aufstellen und Signifikanzniveau festlegen
2. Mit Freiheitsgraden und Signifikanzniveau $\chi^2_{k-1;1-\alpha}$ bestimmen
3. Gewichtete Quadratsumme bilden

$$T = \sum_{i=1}^{k} \frac{(Y_i - np_i)^2}{np_i}.$$

4. Überprüfen ob Nullhypothese verworfen werden soll

$$T > \chi^2_{k-1;1-\alpha}.$$

23. Wie hängen Tests und Konfidenzintervalle zusammen?

Satz 5.8

Wenn $I = [a(X_1, \ldots, X_n), b(X_1, \ldots, X_n)]$ ein Konfidenzintervall mit Überdeckungswahrscheinlichkeit $\gamma = 1 - \alpha$ ist, dann ist für jedes $\theta_0 \in \Theta$ durch die Regel "verwerfe, wenn $\theta_0 \notin I$" ein Test mit Niveau α für $H_0 : \theta = \theta_0$ gegeben.

Ist umgekehrt für jedes θ_0 ein Test mit Niveau α für die Nullhypothese $H_0 : \theta = \theta_0$ gegeben, dann ist die Menge I aller θ_0, für die $H_0 : \theta = \theta_0$ nicht verworfen wird, ein Konfidenzbereich mit Überdeckungswahrscheinlichkeit $\gamma = 1 - \alpha$.

Stochastische Prozesse

Verschiedene Prozesse

1. Was ist ein stochastischer Prozess?

Eine Menge von Zufallsvariablen, welche von dem Parameter t abhängen.

Definition 3.1

Ein stochastischer Prozess ist eine Familie $(X_t, t \in T)$ von Zufallsvariablen. Die Indexmenge T wird Parameterraum genannt und soll eine Teilmenge der reellen Zahlen sein. Der Wertebereich Ω_X von X_t heißt Zustandsraum oder Phasenraum. Wenn T endlich oder abzählbar (etwa $\mathbb{N}$) ist, sprechen wir von einem Prozess in diskreter Zeit, wenn T ein ganzes (endliches oder unendliches) Intervall ist, von einem Prozess in stetiger Zeit,

2. Was ist ein Erneuerungsprozess?

Ein Zählprozess, dessen Zwischenankunftszeiten unabhängige, identisch verteilte, nichtnegative Zufallsvariablen sind.

Definition 3.2

(T_n) sei eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit $T_n > 0$. Wir setzen $S_0 = 0$, $S_n = T_1 + \ldots + T_n$, und für $n = 0, \ldots$

$$X(t) = n \text{ für } S_n \leq t < S_{n+1}.$$

Wir nennen X einen Erneuerungsprozess.

3. Wie ist ein stationärer Prozess definiert?

Wenn die gemeinsame Verteilung eines Prozesses unabhängig von der Teilfolge der Zufallsvariablen gleich bleibt.

Definition 3.4

Der Prozess $X_t, t \in T$ heißt stationär, wenn wenn für $t_1 < t_2 < \ldots < t_n$ und $h > 0$ die gemeinsame Verteilung von $(X(t_1), \ldots, X(t_n))$ mit der von $(X(t_1 + h), \ldots, X(t_n + h))$ übereinstimmt.

Damit kann der Ergodensatz von Birkhoff bewiesen werden.

Satz 3.1: Ergodensatz von Birkhoff

Wenn die Folge (X_n) stationär ist und endlichen Erwartungswert $m = \mathbb{E}(X_n)$ hat, dann existiert

$$X_\infty = \lim_{n\to\infty} \bar{X}_n$$

mit Wahrscheinlichkeit 1 und

$$\mathbb{E}(X_\infty) = m.$$

4. Wie ist ein Prozess mit unabhängigen Zuwächsen definiert?

Wenn man die Partialsummenfolge von unabhängigen Zufallsvariablen betrachtet, dann hängt die Differenz zwischen diesen nicht von den früheren Partialsummen ab, sondern nur von den momentan betrachteten.

Definition 3.5

Der Prozess $(X_t, t \in T)$ heißt Prozess mit unabhängigen Zuwächsen, wenn für $t_1 < t_2 < \ldots < t_n$ die Zufallsvariablen

$$X(t_1), X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$$

unabhängig sind.

5. Wie ist die Markoveigenschaft definiert?

Die Wahrscheinlichkeit einer Zufallsvariable hängt nur von dem letzten Zustand ab nicht von der gesamten Vergangenheit.

Definition 3.6

Der Prozess $X(t)$ heißt Markovprozess, wenn für $t_1 < t_2 < \ldots < t_n$

$$\mathbb{P}(X(t_n) \le x_n | X(t_1) = x_1, \ldots, X(t_{n-1}) = x_{n-1}) = \mathbb{P}(X(t_n) \le x_n | X(t_{n-1}) = x_{n-1}).$$

Poisson Prozess

Erneuerungsprozess mit poissonverteilten Zuwächsen.

Definition 3.7

Der Poissonprozess mit Rate $\lambda > 0$ ist ein Prozess mit unahängigen Zuwächsen, mit $X(0) = 0$ und $X(t) - X(s) \sim P(\lambda(t-s))(s < t)$.

Markovketten

6. Was ist eine Markovkette?

Ein Markovprozess mit diskretem Zustandsraum.

Definition 3.8

Die Wahrscheinlichkeiten

$$\mathbb{P}(X_{n+1} = j | X_n = i)$$

nennen wir die Übergangswahrscheinlichkeiten der Markovkette. Wenn diese nicht von n abhängen, sprechen wir von einer homogenen Markovkette und setzen

$$p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$$

Die Wahrscheinlichkeiten

$$p_{ij}(t) = \mathbb{P}(X_{n+t} = j | X_n = i)$$

nennen wir die t-stufigen Übergangswahrscheinlichkeiten.

Sie kann auch als ein dynamisches System betrachtet werden.

Definition 3.9

Wir sagen, dass die Folge (X_n) ein (stochastisches) dynamisches System bildet, wenn es eine Folge (Y_n) von unabhängigen Zufallsvariablen und eine Funktion $f : \Omega_X \times \mathbb{R} \to \Omega_X$ gibt, sodass

$$X_{n+1} = f(X_n, Y_n)$$

gilt.

7. Wie lauten die Chapman-Kolmogorov-Gleichungen?

Sie beschreibt die Wahrscheinlichkeit des Eintreffen eines Zustandes j nach $s + t$ Schritten vom Ausgangszustand i und stellt die Summe aller möglichen Wege zwischen dem Eingangs- und Ausgangszustand dar.

Satz 3.2: Chapman-Kolmogorov Gleichungen

$$p_{ij}(s+t) = \sum_{k \in \Omega_X} p_{ik}(s) p_{kj}(t).$$

in Matrixnotation mit den *t-stufigen Übergangsmatrizen*

$$P(t) = (p_{ij}(t))_{\Omega_X \times \Omega_X}$$

lauten die Chapman-Kolmogorov Gleichungen

$$P(t+s) = P(t)P(s).$$

Hat man also eine Übergangsmatrix mit t-Schritten und eine mit s-Schritten, können diese Multipliziert werden um eine Übergangsmatrix mit $s+t$ Schritten zu bekommen.

8. Wie ist der einfache Random Walk definiert, und wie lauten seine Übergangswahrscheinlichkeiten?

Der Random Walk beschreibt einen Prozess, bei welchem zufällige Schritte über eine Menge gemacht werden. Beispielsweise ist ein Random Walk über $\mathbb{Z}$ ein Prozess bei welchem mit jedem Schritt zur Partialsumme mit einer gewissen Wahrscheinlichkeit +1 oder -1 addiert wird.

Definition 3.10

Der einfache Random Walk (auch: einfache Irrfahrt) ist der Partialsummenprozess

$$X(t) = X(0) + \sum_{i=1}^{t} Y_i,$$

wobei $(Y_1, Y_2, \ldots)$ unabhängige Zufallsvariable mit

$$\mathbb{P}(Y_i = 1) = p, \mathbb{P}(Y = -1) = 1 - p$$

sind. Der Spezialfall $p = 1/2$ heißt einfacher symmetrischer Random Walk.

9. Wann heißt ein Zustand Nachfolger eines anderen?

Wenn der Zustand j von i $(i \to j)$ mit positiver Wahrscheinlichkeit erreicht werden kann.

Definition 3.11

Der Zustand j heißt Nachfolger von i $(i \to j)$, wenn j von i aus mit positiver Wahrscheinlichkeit erreicht werden kann, also, wenn es ein $t \geq 0$ gibt, sodass $p_{ij}(t) > 0$.

Wenn sowohl $i \to j$ als auch $j \to i$ gilt, dann heißen i und j verbunden oder kommunizierend.

10. Wann heißen zwei Zustände kommunizierend?

Siehe Definition oben.

Klasseneigenschaften

11. Was ist eine Klasseneigenschaft?

Definition 3.12

Eine Eigenschaft heißt Klasseneigenschaft, wenn sie entweder für alle Zustände einer Klasse oder für keinen gilt.

12. Wann heißt eine Markovkette irreduzibel?

Wenn alle Zustände miteinander kommunizieren.

13. Wie ist die Periode eines Zustands definiert?

Die Periode eines Zustandes beschreibt die Anzahl an Schritten welche benötigt werden um wieder zu diesem Zustand zurückzukehren.

Definition 3.13

Die Periode eines Zustandes ist

$$d(i) = \mathrm{ggT}\{t \geq 0 : p_{ii}(t) > 0\}.$$

Wenn $d(i) = 1$ gilt, dann heißt der Zustand i aperiodisch, sonst periodisch.

14. Wann heißt ein Zustand rekurrent?

Ein Zustand i ist rekurrent, wenn der Prozess, ausgehend von i, mit Wahrscheinlichkeit 1 wieder nach i zurückkehren muss.

Satz 3.3

Die folgenden Bedingungen sind äquivalent:

1. $\mathbb{P}_i(\tau_i < \infty) = 1$,
2. $\mathbb{P}_i(\nu_i = \infty) = 1$,
3. $\mathbb{E}_i(\nu_i) = \infty$,
4. $\sum_t p_{ii}(t) = \infty$.

Wenn diese Bedingungen erfüllt sind, dann heißt i rekurrent, sonst transient. Rekurrenz und Transienz sind Klasseneigenschaften.

Wobei τ_i die Übergangs- bzw. Rückkehrzeit und ν_i die Anzahl der Besuche in i ist.

15. Wie ist positive Rekurrenz definiert?

Wenn der Erwartungswert der Übergangs- bzw. Rückkehrzeit des Zustandes i kleiner unendlich ist.

Definition 3.14

i sei ein rekurrenter Zustand. Wenn

$$\mathbb{E}_i(\tau_i) < \infty$$

gilt, dann heißt i positiv rekurrent, sonst nullrekurrent.

16. Wie verhalten sich die Übergangswahrscheinlichkeiten pij (t) in einer irreduziblen aperiodischen Markovkette für t → ∞?

Dann gibt es eine Wahrscheinlichkeit, dass die Markovkette einen stationären Zustand erreicht.

Satz 3.4

Wenn (X_n) irreduzibel und aperiodisch ist, dann existieren die Grenzwerte

$$\lim_{n\to\infty} p_{ij}(t) = \pi_j = \frac{1}{\mathbb{E}_j(\tau_j)}.$$

Im periodischen Fall gilt

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} p_{ij}(t).$$

Wenn (π_i) nicht identisch verschwindet (also wenn die Kette positiv rekurrent ist), dann ist es eine stationäre Verteilung. Umgekehrt folgt aus der Existenz einer stationären Verteilung die positive Rekurrenz. Die positive Rekurrenz bzw. Nullrekurrenz ist ebenfalls eine Klasseneigenschaft.

17. Was ist eine stationäre Verteilung?

Von einer stationären Verteilung spricht man genau dann, wenn eine Markov-Kette eine Startverteilung hat, welche sich mit der Zeit nicht ändert.

Definition 3.15

$(\pi_i, i \in \Omega_X)$ heißt stationäre Verteilung, wenn

$$\pi_i \geq 0,$$

$$\sum_i \pi_i = 1$$

und

$$\sum_i \pi_i p_{ij} = \pi_j.$$

18. Wie kann man die positive Rekurrenz anders charakterisieren?

Aus der Existenz einer stationären Verteilung folgt die positive Rekurrenz.

19. Was kann man über die Rekurrenz des Random Walk aussagen?

Quelle

- Symmetrischer Random Walk auf $\mathbb{Z}$ $p = q = \frac{1}{2}$ (Nullrekurrent)

Für ein N und alle $n \geq N$ gilt $p_{00}^{(2n)} \sim \frac{1}{\sqrt{n\pi}}$

Also: $\sum_{n=1}^{\infty} p_{ii}^{(2n)} \geq \sum_{n=N}^{\infty} p_{ii}^{(2n)} \sim \frac{1}{\sqrt{\pi}} \sum_{n=N} \frac{1}{\sqrt{n}} = +\infty$, d.h. rekurrent.

- Asymmetrischer Random Walk $\mathbb{Z}$

Wie vorher

$$p_{00}^{(2n)} = \frac{(2n)!}{n!n!} p^n q^n \sim \frac{(4pq)^n}{\sqrt{n\pi}}.$$

Hier gilt $4pq = \alpha < 1$. Also:

$$\sum_{n=1}^{\infty} p_{ii}^{(2n)} = K + \sum_{n=N}^{\infty} p_{ii}^{(2n)} \sim K + \frac{1}{\sqrt{\pi}} \sum_{n=N} \frac{\alpha^n}{\sqrt{n}} < K + \frac{1}{\sqrt{\pi}} \frac{1}{1-\alpha},$$

d.h. transient.

20. Was kann man über dir Rekurrenzeigenschaften von endlichen Markovketten haben.

Ist eine Markovkette irreduzibel, dann ist diese immer positiv rekurrent. Nullrekurrenz wäre nur möglich, wenn die entsprechende Klasse unendlich viele Zustände hätte.

21. Wie sind Absorptionswahrscheinlichkeiten definiert?

Absorptionswahrscheinlichkeiten beschreibt die Wahrscheinlichkeit a_i durch welche bei dem Start in einem Zustand i ein absorbierender Zustand i_0 erreicht wird.

$$a_i = \mathbb{P}_i(\tau_{i_0} < \infty) = \mathbb{P}_i(X \text{wird in } i_0 \text{ absorbiert}).$$

Satz 3.5

Die Absorptionswahrscheinlichkeiten sind die kleinste nichtnegative Lösung des Gleichungssystems

$$a_{i_0} = 1,$$
$$a_i = \sum_j p_{ij} a_j, i \neq i_0.$$

22. Wie kann man die Absorptionswahrscheinlichkeiten berechnen?

Siehe oben.

Markovketten in stetiger Zeit

23. Welche speziellen Annahmen machen wir für Markovketten in stetiger Zeit?

Es muss die Stetigkeit der Übergangswahrscheinlichkeiten bei 0 verlangt werden.

$$\lim_{t \to 0} p_{ij}(t) = p_{ij}(0) = \begin{cases} 1 & \text{wenn } i = j, \\ 0 & \text{sonst.} \end{cases}$$

Diese Annahme genügt schon für die Existenz der Ableitungen:

Satz 3.6

Wenn die Übergangswahrscheinlichkeiten bei 0 stetig sind, dann existieren die Grenzwerte

$$q_{ij} = \lim_{t \to 0} \frac{p_{ij}(t) - p_{ij}(0)}{t}.$$

Die Zahlen q_{ij} heißen die infinitesimalen Parameter, die Matrix

$$Q = (q_{ij})_{\Omega_X \times \Omega_X}$$

die infinitesimale Matrix oder der infinitesimale Erzeuger der Markovkette.

24. Wie kann man Rekurrenz, Transienz, positive Rekurrenz und Nullrekurrenz für Markovketten in stetiger Zeit definieren?

Prinzipiell genauso wie die Eigenschaften einer Markovkette in diskreter Zeit.

25. Was sind die infinitesimalen Parameter einer Markovkette?

Die Werte des infinitesimalen Erzeugers einer Markovkette. (Ableitungen der Übergangswahrscheinlichkeiten bei $t = 0$)

Satz 3.6

Wenn die Übergangswahrscheinlichkeiten bei 0 stetig sind, dann existieren die Grenzwerte

$$q_{ij} = \lim_{t \to 0} \frac{p_{ij}(t) - p_{ij}(0)}{t}.$$

Die Zahlen q_{ij} heißen die infinitesimalen Parameter, die Matrix

$$Q = (q_{ij})_{\Omega_X \times \Omega_X}$$

die infinitesimale Matrix oder der infinitesimale Erzeuger der Markovkette.

26. Wie lauten die Kolmogorovschen Differentialgleichungen?

Definition 3.16: Kolmogorovsche Differentialgleichungen

Die Gleichung

$$P'(t) = QP(t)$$

heißt die Rückwärtsgleichung,

$$P'(t) = P(t)Q$$

die Vorwärtsgleichung von Kolmogorov.

27. Was kann man über die Gültigkeit der Kolmogorovschen Differentialgleichungen sagen?

Satz 3.7

Wenn die Markovkette X_t konservativ ist, dann erfüllen die Übergangswahrscheinlichkeiten $P(t)$ die Rückwärtsgleichung.

Wenn die infinitesimalen Parameter beschränkt sind, dann erfüllen die Übergangswahrscheinlichkeiten $P(t)$ die Vorwärtsgleichung.

Definition 3.17

Die Markovkette X_t bzw. ihr Erzeuger Q heißt konservativ, wenn für alle i

$$q_{ii} > -\infty$$

und

$$\sum_j q_{ij} = 0$$

gilt.

Definition 3.19

Eine Markovkette in stetiger Zeit heißt regulär, wenn sie in endlicher Zeit nur endlich viele Sprünge ausführt, im anderen Fall singulär..

Satz 3.8

Ist X regulär, dann gelten die Vorwärts- und die Rückwärtsgleichung, und sie sind eindeutig lösbar.

Geburts- und Todesprozesse

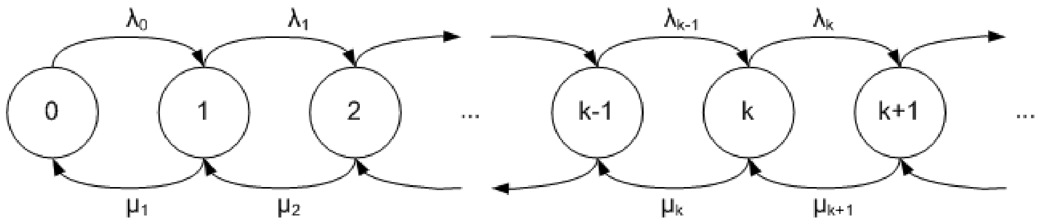
Definition 3.18

Ein Geburts- und Todesprozess ist eine Markovkette mit Zustandsraum $\{0, 1, \ldots\}$ und infinitesimalen Parametern

$$q_{00} = -\lambda_0, q_{01} = \lambda_0,$$

$$q_{i,i-1} = \mu_i, q_{ii} = -(\lambda_i + \mu_i), q_{i,i+1} = \lambda_i.$$

Die Zahlen λ_i heißen die Geburtsraten, μ_i die Todesraten. Sind alle $\mu_i = 0$, dann sprechen wir von einem reinen Geburtsprozess, sind alle $\lambda_i = 0$, von einem reinen Todesprozess.



28. Wie kann man für endliche Markovketten die Übergangsmatrizen aus dem infinitesimalen Erzeuger erhalten?

Allgemein kann die Übergangsmatrix mit folgender Formel berechnet werden.

$$P(t) = e^{Qt} = E\Lambda^t E^{-1} = E \begin{pmatrix} e^{t\lambda_1} & & \\ & \ddots & \\ & & e^{t\lambda_d} \end{pmatrix} E^{-1}.$$

Wobei E eine Matrix mit den Eigenvektoren von Q als Spaltenvektoren gebildet wird.
Daher müssen in erster Linie die Eigenwerte und Eigenvektoren von Q berechnet werden und anschließend die inverse Matrix E^{-1} bestimmt werden. Der Rest ist Matrixmultiplikation.

29. Wie kann man die infinitesimalen Parameter anschaulich (?) interpretieren?

?

30. Wie kann mithilfe des infinitesimalen Erzeugers die Absorptionswahrscheinlichkeiten berechnen?

Für die Absorptionswahrscheinlichkeiten: es wird wieder $a_i = 1$ gesetzt, wenn i ein "guter" absorbierender Zustand ist, $a_i = 0$ für "schlechte", und wenn i nicht absorbierend ist, dann

$$\sum_j q_{ij} a_j = 0.$$

Wie im diskreten Fall kann es sein, dass diese Gleichung nicht eindeutig lösbar ist. Dann ist wieder die kleinste mit $a_i \geq 0$ die richtige.

Mittlere Absorptionszeiten: m_i ist wieder 0, wenn i absorbierend ist, sonst gilt

$$\sum_j q_{ij} m_j = 0.$$

31. Wie kann mithilfe des infinitesimalen Erzeugers die stationäre Verteilung berechnen?

Entweder über die eingebettete Markovkette, oder aus der Gleichung für die diskretisierte Kette, deren Schrittweite wir gegen 0 gehen lassen, erhalten wir die Formeln

Für die stationäre Verteilung:

$$\pi Q = 0.$$