Einführung in die Wahrscheinlichkeitstheorie

Elke Warmuth

WS 2005/ 2006

Contents

Einführung

Wie erstellt man ein Modell, um ein Problem der Elementaren Wahrscheinlichkeitstheorie zu lösen?

1. Was kann alles passieren?

   Die Ergebnismenge Ω enthält alle möglichen Ausgänge eines Zufallsversuches.

   Wir bezeichnen mit ω ein Ergebnis, $\omega \in \Omega$.

   Bsp. vierfacher Münzwurf

   $\Omega = \{WWWW, WWWZ, ..., ZZZZ\}$

   Wir können 16 verschiedene Ergebnisse erhalten.

2. Welche Wahrscheinlichkeiten haben die Ergebnisse?

   Bsp. Wie hoch ist die Wahrscheinlichkeit vier mal Wappen zu werfen?

   $\omega = (WWWW)$

   $P(\omega) = \frac{1}{16}$

3. Merkmal/Zufallsgröße

   Bsp. Sei X die Anzahl der Wappen im Viererblock.

   $P(X = 2) = \frac{6}{16}$

1. Deutung

Es ist möglich eine Vorhersage, für die relative Häufigkeit bei einer großen Anzahl von Beobachtungen, zu treffen.

2. Deutung

Man kann bei einer Wette einen fairen Einsatz ermitteln.

Chapter 1

Diskrete Wahrscheinlichkeitsraeume

1.1 Vorgaenge mit zufaelligem Ergebnis

Ein Vorgang mit zufälligem Ergebnis wird auch als Zufallsexperiment oder Zufallsversuch bezeichnet.

Bedingungen:

- es muss ein Komplex von Bedingungen gegeben sein
- bei jeder Beobachtung ist das Ergebnis nicht sicher vorhersehbar
- (der Vorgang ist im Prinzip beliebig oft wiederholbar)

1.2 Modelle fuer Vorgaenge mit zuf. Ergebnis

Ergebnismenge Ω:

Die Elemente von Ω repräsentieren die Ergebnisse des Vorgangs.

Die Wahl der Ergebnismenge ist nicht eindeutig.

Bsp. Würfeln mit 2 Würfeln

$\Omega_1 = \{(\omega_1, \omega_2) : \omega_i \in \{1, 2, ..., 6\}\}$

$\Omega_2 = \{[\omega_1, \omega_2] : \omega_i \in \{1, 2, ..., 6\}, \omega_i \leq \omega_j : fuer, i \leq j\}$

Diskrete Wahrscheinlichkeitsräume sind höchstens abzählbar,

d.h. Ω ist eineindeutig auf die Menge der natürlichen Zahlen abbildbar.

Bsp.

$\Omega_1 = \{\omega_1, \omega_2, ..., \omega_\nu\}$

$\Omega_2 = \{\omega_1, \omega_2, ...\}$

Ereignisse:

Ereignisse sind eine Aussage über die Eigenschaft des Ergebnisses.

Ereignisse werden im Modell durch Teilmengen von Ω repräsentiert.

besondere Ereignisse:

- Ω- das sichere Ereignis
- $\oslash$ - das unmögliche Ereignis

Sprechweise:

- ein Ereignis A tritt ein, wenn ein Ergebnis ω mit $\omega \in A$ eintritt
- $\omega \in A$ heißt günstig für das Ereignis A

Wiederholung wichtiger Begriffe der Mengenlehre:

Verknüpfung	**Beschreibung**
$A \cup B$	Ereignis A od. B tritt ein, od. beide
$A \cap B$	Ereignis A und B treten ein
$\overline{A}$	Ereignis A tritt nicht ein
$A \backslash B = A \cap \overline{B}$	Ereignis A tritt ein, aber B nicht
$\bigcup\limits_{k=1}^{\infty} A_k$	mind. eins der A_k tritt ein
$\overline{\bigcup\limits_{k=1}^{\infty} A_k} = \bigcap\limits_{k=1}^{\infty} \overline{A_k}$	keins der A_k tritt ein/ de Morgansche Regeln

Einführung einer Notation:

$A \dot{\cup} B$ für $A \cup B$, wenn gilt $A \cap B = \oslash$

Wir haben bis hier die folgenden Begriffe behandelt:

Ω	Ergebnismenge
ω	Ergebnis
$A \subseteq \Omega$	Ereignis

Die Wahrscheinlichkeit ist eine Abbildung, die jedem Ergebnis eine Wahrscheinlichkeit zuordnet:

$\Omega \ni \omega \mapsto P(\omega)$

Dabei gelten folgende Eigenschaften:

$$0 \leq P(\omega) \leq 1$$

$$\sum_{\omega \in \Omega} P(\omega) = 1$$

(Das ist eine endliche Summe oder eine absolut konvergente Reihe.)

Mit welcher Wahrscheinlichkeit tritt ein Ereignis A ein?

$$P(A) = \sum_{\omega \in A} P(\omega)$$

Weitere Begriffe:

$\Re(\Omega)$	Potenzmenge, d.h. die Menge aller Teilmengen von Ω
$P : \Re(\Omega) \to [0,1]$	Wahrscheinlichkeitsverteilung
(Ω, P)	Wahrscheinlichkeitsraum

Satz 1.1:

Geg. ein Wraum (Ω, P) mit $P(A) = \sum\limits_{\omega \in A} P(\omega)$ wobei $P(\omega) \in [0,1]$ und $\sum\limits_{\omega \in \Omega} P(\omega) = 1$.

Dann gilt:

1) P ist nichtnegativ, d.h. $P(A) \geq 0$

2) P ist normiert, d.h. $P(\Omega) = 1$

3) P ist additiv, d.h. $P(A \dot{\cup} B) = P(A) + P(B)$

4) P ist σ-additiv, d.h. $P(\dot{\bigcup\limits_{k=1}^{\infty}} A_k) = \sum\limits_{k=1}^{\infty} P(A_k)$.

Wie findet man $P(\omega)$?

1. statistische Grundlagen
2. Man nimmt eine Gleichwahrscheinlichkeit an und bildet das Verhältnis zwischen Treffer und Anzahl aller möglichen Ergebnisse.
3. Expertenschätzung

Wahrscheinlichkeit und relative Häufigkeit

Definition 1.1: relative Häufigkeit
Beobachtet man einen Vorgang n-mal unter den gleichen Bedingungen, ergibt sich als relative Häufigkeit $h_n(A) = \frac{n(A)}{n}$.
Dabei bezeichnet $n(A)$ die absolute Häufigkeit.

Eigenschaften:
- Stabilwerden der relativen Häufigkeit
- kein Grenzwert
- nicht beweisbar, in der Beobachtungsebene
- bei festem n gelten für den stabilen Wert die Eigenschaften 1)-3) aus Satz 1.1

Satz 1.2: RR für Wkt.
In einen Wraum (Ω, P) gilt:

1) $P(\bar{A}) = 1 - P(A)$
2) $P(A \backslash B) = P(A) - P(A \cap B)$
3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4) $P(A_1 \cup A_2 \cup ... \cup A_n) = \sum\limits_{k=1}^{n} (-1)^{k+1} \sum\limits_{1 \leq i_1 < i_2 < ... < i_k \leq n} P(A_{i_1} \cap A_{i_2} ... \cap A_{i_k})$.

1.3 Die Gleichverteilung auf einer endlichen Menge

Eigenschaften der Gleichverteilung:

$$\Omega = \{\omega_1, \omega_2, ..., \omega_r\}$$

$$P(\omega_1) = P(\omega_2) = ... = P(\omega_r)$$

$$P(A) = \sum_{\omega \in A} \frac{1}{r} = \frac{|A|}{r} = \frac{|A|}{|\Omega|}$$

Elemente der Kombinatorik

Zählalgorithmus:

Eine Auswahl vollziehe sich in k Schritten.

Gibt es im:

1. Schritt: n_1 Möglichkeiten

2. Schritt: n_2 Möglichkeiten

...

k. Schritt: n_k Möglichkeiten.

So gibt es insgesamt: $n_1 \cdot n_2 \cdot ... \cdot n_k$ Möglichkeiten.

Anzahl der Stichproben vom Umfang k aus $\{1, ..., n\}$:

	mit Zurücklegen	**ohne Zurücklegen**
mit Reihenfolge	n^k	$n \cdot (n-1) \cdot ... \cdot (n-(n-k))$ Spezialfall für $k = n : n!$
ohne Reihenfolge	$\binom{n+k-1}{k}$	$\binom{n}{k}$

Was bedeutet "Auswahl auf gut Glück"?

Jede der $\binom{n}{k}$ Teilmengen hat dieselbe Wahrscheinlichkeit.

Beispiele:

1. **Geburtstagsproblem:**

   Wie groß ist die Wahrscheinlichkeit, dass unter n zufällig ausgewählten Personen mindestens zwei am gleichen Tag Geburtstag haben?

   $\Omega = \{(g_1, g_2, ..., g_n) : g_i \in \{1, 2, ..., 365\}\}$

   $|\Omega| = 365^n$

Annahme: gleichwahrscheinlich

A: mindestens zwei am gleichen Tag

$\overline{A}$: alle an verschiedenen Tagen

$\left|\overline{A}\right| = 365 \cdot 364 \cdot \ldots \cdot (365 - (n-1))$

$P(\overline{A}) = \frac{365 \cdot 364 \cdot \ldots \cdot (365-(n-1))}{365^n} = \frac{365}{365} \cdot \frac{364}{365} \cdot \ldots \cdot \frac{365-(n-1)}{365} = 1 \cdot (1 - \frac{1}{365}) \cdot (1 - \frac{2}{365}) \cdot \ldots \cdot (1 - \frac{n-1}{365})$

Wir nehmen nun eine Abschätzung vor, durch: $e^x \approx (1-x)$ für kleine x

$P(\overline{A}) \approx e^{-\frac{1}{365}} \cdot e^{-\frac{2}{365}} \cdot \ldots \cdot e^{-\frac{n-1}{365}} = e^{-\frac{1}{365}(1+2+\ldots+(n-1))}$

$P(\overline{A}) \approx e^{-\frac{1}{365}\left(\frac{(n-1)n}{2}\right)} = e^{\frac{-(n-1)n}{730}}$

$P(A) \approx 1 - e^{\frac{-(n-1)n}{730}}$

2. **Ausgleich beim Spiel:**

- zwei gleichstarke Gegner, kein Remis
- 2n Spiele

Wie groß ist die Wahrscheinlichkeit, dass nach 2n Spielen Gleichstand herrscht?

$\Omega = \{(s_1, s_2, \ldots, s_{2n}) : s_i \in \{A, B\}\}$

$|\Omega| = 2^{2n}$

Annahme: gleichwahrscheinlich

A_{2n}: Gleichstand nach 2n Spielen

$|A_{2n}| = \binom{2n}{n}$

$$P(A_{2n}) = \frac{\binom{2n}{n}}{2^{2n}} =: a_{2n}$$

Hilfsmittel: Stirlingsche Formel

$$n! \approx_{as} \sqrt{2\pi n} \cdot n^n \cdot e^{-n} =: c_n$$

d.h.:

$$\lim_{n\to\infty} \frac{n!}{c_n} = 1$$

$$\Rightarrow \lim_{n\to\infty} \frac{n! - c_n}{c_n} = 0$$

durch Einsetzen erhält man:

$$\lim_{n\to\infty} a_{2n} = \lim_{n\to\infty} \frac{1}{\sqrt{\pi n}} = 0$$

3. **Fixpunkte einer zufälligen Permutation**

z.B. Juleklapp

Wie groß ist die Wahrscheinlichkeit, dass ein Fixpunkt auftritt?

$\Omega = \{(i_1, i_2, ..., i_n) : (i_1, i_2, ..., i_n)$Permutation von $(1, ..., n)\}$

$|\Omega| = n!$

Annahme: gleichwahrscheinlich

A_k: k ist Fixpunkt der Permutation

$$P(A_k) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

B: mindestens ein Fixpunkt

$$\begin{aligned} B &= \bigcup_{k=1}^{n} A_k \\ P(B) &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k \le n} P(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}) \\ &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k \le n} \frac{(n-k)!}{n!} \\ &= \sum_{k=1}^{n} (-1)^{k+1} \frac{(n-k)!}{n!} \binom{n}{k} \\ &= \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} \\ &= -\sum_{k=0}^{n} \frac{(-1)^k}{k!} + 1 \stackrel{n \to \infty}{\longrightarrow} 1 - \frac{1}{e} \approx 0,632 \end{aligned}$$

Berechnung der Anzahl der Fixpunkte:

a_n: Anzahl der fixpunktfreien Permutationen einer n-elementigen Menge

$a_n = (n-1)(a_{n-2} + a_{n-1})$

z.Bsp.

$a_4 = 3(1+2) = 9$, $a_5 = 4(2+9) = 44$

Sei n=5,

B_k : genau k Fixpunkte

b_k : Anzahl der Möglichkeiten für k Fixpunkte

$P(B_0) = \frac{a_5 \cdot b_0}{n!} = \frac{44 \cdot \binom{5}{0}}{5!} = \frac{44}{5!}$

$P(B_1) = \frac{a_4 \cdot b_1}{n!} = \frac{9 \cdot \binom{5}{1}}{5!}$

$P(B_2) = \frac{a_3 \cdot b_2}{n!} = \frac{2 \cdot \binom{5}{2}}{5!}$

$P(B_3) = \frac{a_2 \cdot b_3}{n!} = \frac{1 \cdot \binom{5}{3}}{5!}$

$P(B_4) = \frac{a_1 \cdot b_4}{n!} = \frac{0 \cdot \binom{5}{4}}{5!}$

$P(B_5) = \frac{a_0 \cdot b_5}{n!} = \frac{1 \cdot \binom{5}{5}}{5!}$ wobei $a_0 := 1$

4. **Ziehen ohne Zurücklegen aus "roten und schwarzen" Kugeln**

   - Skat mit 32 Karten, darunter 4 Buben

   - jeder der drei Spieler bekommt 10 Karten

   A: Spieler 1 bekommt 3 Buben

   Modell für Karten von Spieler 1 (Karten der anderen bleiben unberücksichtigt)

   $\Omega = \{\{k_1, k_2, .., k_{10}\} : \{k_1, ..., k_{10}\} \subset \{1, ..., 32\}\}$
   $|\Omega| = \binom{32}{10}$

   Annahme: gleichwahrscheinlich

   rote Kugeln - 4 Buben

   schwarze Kugeln - 28 Nichtbuben
   $|A| = \binom{4}{3}\binom{28}{7}$
   $P(A) = \frac{|A|}{|\Omega|} \approx 0,07$

5. **Wer zuerst zieht,...**

   (a) - Urne mit 1 weißen und 9 roten Kugeln

   - 2 Spieler ziehen abwechselnd und ohne Zurücklegen

   - weiß gewinnt

   - Hat der zuerst Ziehende einen Vorteil?

   A: der zuerst Ziehende gewinnt

$P(A) = \frac{1}{2}$

Fazit: Es besteht nur ein Vorteil für den ersten Spieler, wenn die Anzahl der Kugeln ungerade ist.

(b) - eine weiße und eine schwarze Kugel

- 2 Spieler ziehen abwechselnd mit Zurücklegen
- weiß gewinnt
- Hat der zuerst Ziehende einen Vorteil?

$|\Omega| = \{W, RW, RRW, ...\}$ nicht gleichwahrscheinliche Ergebnisse

$$P(\underbrace{RR...R}_{k-mal}W) = \frac{1}{2^{k+1}} = \left(\frac{1}{2}\right)^{k+1}$$

$$P(A) = P\left(\dot{\bigcup_{k=0}^{\infty}} A_{2k+1}\right) = \sum_{k=0}^{\infty} P(A_{2k+1}) = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{2k+1} = \frac{1}{2} \cdot \frac{1}{1-\frac{1}{4}} = \frac{2}{3}$$

1.4 Zufallsgroessen und ihre Verteilung

Definition 1.2: Zufallsgröße

Sei (Ω, P) ein diskreter WRaum. Dann heißt jede Funktion $X : \Omega \longrightarrow \Re$ eine Zufallsgröße (ZG).

Bsp.: Anton und Pünktchen Übung 1.1

T: Anzahl der Spiele

$APA \mapsto T(APA) = 3$

$PAA \mapsto T(PAA) = 3$

$PPP \mapsto T(PPP) = 3$

Die Wahrscheinlichkeiten von $P(\omega)$ übertragen sich auf die Zufallsgröße.

$$P(T = 3) = \frac{3}{8}$$

$$P(\{\omega \in \Omega : T(\omega) = 3\})$$

$$P(\{APA, PAA, PPP\})$$

Definition 1.3: Verteilung

Sei (Ω, P) ein diskreter WRaum und X eine ZG mit der Wertemenge $\{x_1, x_2, \ldots\}$.

Die durch $p_k = P(X = x_k) = P(\{\omega \in \Omega : X(\omega) = x_k\})$ gegebene Verteilung auf $\{x_1, x_2, \ldots\}$ heißt (Wahrscheinlichkeits-) Verteilung der ZG X.

Beispiele:

1. Würfeln bis zur ersten Sechs/ geometrische Verteilung

   $\Omega = \{6, 06, 006, ...\}$

   $$P(\underbrace{00..0}_{(k-1)-mal}6) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6}, k = 1, 2, ...$$

   X- Anzahl der Würfe bis zur ersten Sechs (einschließlich)

   Y- Anzahl der Würfe bis vor der ersten Sechs

   $$P(X = k) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6}, k = 1, 2, ...$$

   $$P(Y = k) = \left(\frac{5}{6}\right)^{k} \cdot \frac{1}{6}, k = 0, 1, 2, ...$$

   Wie groß ist die Wahrscheinlichkeit, spätestens im n-ten Wurf die erste Sechs zu würfeln?

   $$P(X \leq n) = \sum_{k=1}^{n} P(X = k) = \sum_{k=1}^{n} \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6} = \frac{1}{6} \cdot \frac{1 - \left(\frac{5}{6}\right)^n}{1 - \frac{5}{6}} = 1 - \left(\frac{5}{6}\right)^n$$

2. geg.: Sicherheitswahrscheinlichkeit von $1 - \alpha, \alpha$-klein

   ges.: Anzahl der Würfe n, so dass

   $$P(X \leq n) \geq 1 - \alpha$$

   $$1 - \left(\frac{5}{6}\right)^n \geq 1 - \alpha$$

   $$\Longleftrightarrow$$

   $$n \geq \frac{\ln \alpha}{\ln \frac{5}{6}}$$

Indikator- und Zählvariablen

- diese dienen als technisches Hilfsmittel

$$A \subset \Omega : I_A(\omega) = \begin{cases} 1 & : \quad \omega \in A \\ 0 & : \quad \omega \notin A \end{cases}$$

Eigenschaften:

1. $I_{A \cap B}(\omega) = I_A(\omega) \cdot I_B(\omega)$
2. $I_{A \dot{\cup} B}(\omega) = I_A(\omega) + I_B(\omega)$
3. $I_{\bar{A}}(\omega) = 1 - I_A(\omega)$
4. für die Ereignisse: $A_1, A_2, ..., A_n$ $\qquad X(\omega) = I_{A_1}(\omega) + ... + I_{A_n}(\omega)$
   $X(\omega)$: gibt an, wie viele der A_k eingetreten sind.

1.5 Statistische Qualitaetskontrolle

Begriffe:

- Posten: N Erzeugnisse
- Stichprobe: n Erzeugnisse
- D: defekte Erzeugnisse im Posten
- N-D: intakte Erzeugnisse im Posten

Urteil aufgrund der Stichprobe:

1. Fehler 1.Art:
   Ein "guter" Posten wird abgelehnt.
   Das passiert,wenn zufällig zu viele defekte Teile in die Stichprobe gelangen.
2. Fehler 2.Art:
   Ein "schlechter" Posten wird angenommen.
   Das passiert, wenn zufällig zu wenige defekte Teile in die Stichprobe gelangen.

Modell für Stichprobenentnahme:

Man denkt sich die Teile durchnummeriert.

$\Omega = \{\{t_1, t_2, ..., t_n\} : \{t_1, t_2, ..., t_n\} \subset \{1, ..., N\}\}$

Annahme: Auswahl auf Gut Glück

$|\Omega| = \binom{N}{n}$

$X(\{t_1, ..., t_n\}) =$ Anzahl der defekten Teile in $\{t_1, ..., t_n\}$

Hypergeometrische Verteilung:

$$P(X = k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$$

$$k = 0, 1, ..., min(n, D)$$

Didaktisches Beispiel:

$N = 100$ $D = 5$ $n = 20$

x_k	0	1	2	3	4	5
p_k	0,32	0,42	0,21	0,05	0,00	0,00

Wie groß ist die Wahrscheinlickeit, dass mindestens 10% Ausschuss in der Stichprobe enthalten sind?

$P(X \geq 2) = 0,26$

Folie 1

Entscheidungsregel

- Wenn $X \leq d$, nimm den Posten an.
- Wenn $X > d$, lehne den Posten ab.

Wie wählt man d?

Die Konsequenz der Wahl von d veranschaulicht die Annahmekennlinie.

Welcher funktionale Zusammenhang besteht zwischen d und D und der Annahmewahrscheinlichkeit?

Wie groß ist die Wahrscheinlichkeit, den Posten anzunehmen wenn der Ausschussanteil $\frac{D}{N}$

beträgt?

$$A\left(\frac{D}{N}\right) = P(X \leq d) = \frac{\binom{D}{0}\binom{N-D}{n} + \binom{D}{1}\binom{N-D}{n-1} + ... + \binom{D}{d}\binom{N-D}{n-d}}{\binom{N}{n}}$$

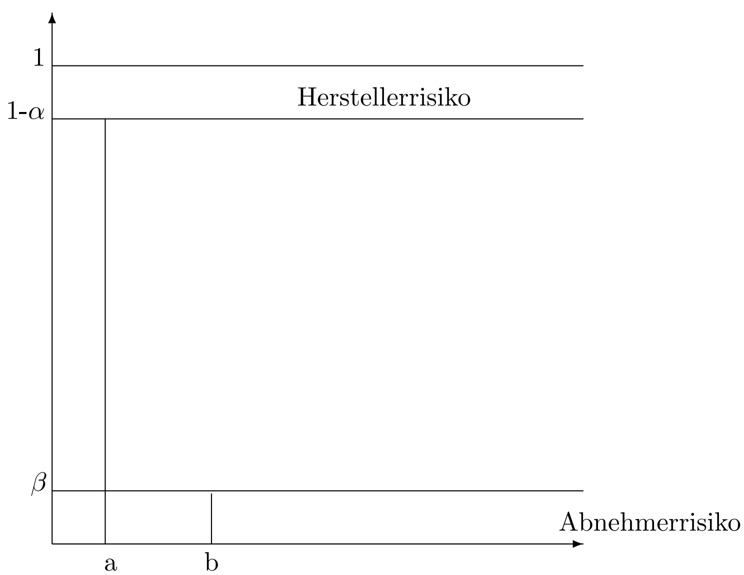
Der Graph dieser Funktion heißt Annahmekennlinie bzw. Operationscharakteristik.

Was sind die Konsequenzen einer Entscheidungsregel?

Folie 2

- man kann nicht gleichzeitig beide Fehlerarten reduzieren
- man kann die Annahmekennlinie im kritischen Bereich steiler gestalten, indem man den Stichprobenumfang erhöht

Folie 3



Vereinbart werden a, b, α, β.

Hersteller:

Wenn der Posten $a \cdot 100\%$ Ausschussanteil enthält, wird er mit einer Wahrscheinlichkeit von mindestens $1 - \alpha$ abgenommen.

Abnehmer:

Wenn der Ausschussanteil $b \cdot 100\%$ überschreitet, muss er den Posten höchstens mit der Wahrscheilichkeit β abnehmen.
α und β sind kleine Zahlen.

1.6 Bedingte Wahrscheinlichkeiten, Kopplung von Experimenten, Unabhaengigkeit

Beispiele:

1. Urnenmodell
   zweimaliges Ziehen aus einer Urne mit 5 roten und 3 schwarzen Kugeln
   R_2: die zweite Kugel ist rot
   $P(R_2) = \frac{5}{8}$
   Information: R_1: die erste Kugel rot ist eingetreten
   $\Rightarrow$ Änderung der Wahrscheinlichkeit von R_2
   $P(R_2|R_1) = \frac{4}{7} < \frac{5}{8}$

2. Skatspiel
   A: 2 Buben im Skat
   $P(A) = \frac{3}{248} \approx 0,012$
   B_k: ich habe k Buben auf der Hand
   $P(A|B_3) = 0$
   $P(A|B_0) > 0,012$

3. Lebensversicherung
   Sei X die zufällige Lebensdauer eines männlichen Neugeborenen.
   $P(X \geq 50) \approx 0,944$
   Ein 40-jähriger Mann will eine Lebensversicherung mit 10 Jahren Laufzeit abschließen.
   Information: $X \geq 40$
   $P(X \geq 50 | X \geq 40) > 0,944$

4. Zweimal Würfeln
   Information: 1. Wurf ist eine 6

$P(W_2 = 6|W_1 = 6) = \frac{1}{6}$

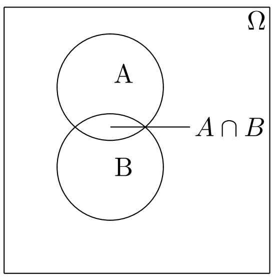
keine Änderung

Allgemeine Situtation:

(Ω, P), B Ereignis mit $P(B) > 0$

Information: B ist eingetreten

Daraus folgt eine neue Chancenverteilung für A.



Eine bedingte Wahrscheinlichkeit ist der "Anteil" der günstigen Ergebnisse $A \cap B$ an allen möglichen Ergebnissen in B.

Der "Anteil" wird als Wahrscheinlichkeit gemessen.

Definition 1.4: bedingte Wahrscheinlichkeit

Sei (Ω, P) ein WRaum, A, B $\subseteq \Omega$ Ereignisse und $P(B) > 0$.

Die bedingte Wahrscheinlichkeit von A unter der Bedingung B ist: $P(A|B) = \frac{P(A \cap B)}{P(A)}$.

Satz 1.3:

Durch $P(\bullet|B)$ wird eine Wahrscheinlichkeitsverteilung auf Ω definiert.

Eigenschaften:

1. $P(A|B) \geq 0$
2. $P(\Omega|B) = 1$
3. $P(A_1 \dot{\cup} A_2|B) = P(A_1|B) + P(A_2|B)$
4. σ-Additivität
5. $P(\overline{A}|B) = 1 - P(A|B)$

Motivation für den nachfolgenden Satz:

Wir haben bis jetzt keine Regel für: $P(A \cap B)$.

Bedingte Wahrscheinlichkeiten helfen uns, soetwas zu berechnen.

Satz 1.4:

Seien $A, A_1, A_2, ..., A_n, B$ Ereignisse mit $P(B) > 0$ und $P(A_1 \cap \ldots \cap A_{n-1}) > 0$.

Dann gilt:

1) $P(A \cap B) = P(B) \cdot P(A|B)$
2) $P(A_1 \cap ... \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot ... \cdot P(A_n|A_1 \cap ... \cap A_{n-1})$

   Multiplikationsformel

Kopplung von Experimenten

Rahmen:

- n: Teilvorgänge
- Ω_i: Ergebnismenge für i-ten Teilvorgang
- P_1: Wahrscheinlichkeitsverteilung für 1. Teilvorgang
- Übergangswahrscheinlichkeiten

  $P_i(\omega_i|\omega_1, ..., \omega i - 1)$

  Übergangswahrscheinlichkeiten für ω, zum i-ten Teilvorgang, wenn $\omega_1, ..., \omega_{i-1}$ die Ergebnisse der Teilvorgänge $1, 2, ..., i - 1$ waren, mit $i = 1, 2, ..., n$.

Modell für den Gesamtvorgang:

$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n = \{(\omega_1, \omega_2, ..., \omega_n) : \omega_i \in \Omega_i\}$

Zufallsgröße:

$X_i(\omega) = X_i((\omega_1, ..., \omega_n)) = \omega_i$, $X_i : \Omega \to \Omega_i$

(Was ist im i-ten Teilexperiment passiert?)

Ereignis:

im i-ten Teilvorgang ist ω_i eingetreten

$\Leftrightarrow \{\omega \in \Omega : X_i(\omega) = \omega_i\} \subset \Omega$

kurz: $\{X_i(\omega) = \omega_i\}$ oder: $\{X_i = \omega_i\}$

Satz 1.5:

In dem beschriebenen Rahmen wird durch:

$P(\omega) = P(\omega_1, \ldots, \omega_n) = P(\omega_1) \cdot P(\omega_2|\omega_1) \cdot P(\omega_3|\omega_1, \omega_2) \cdot \ldots \cdot P(\omega_n|\omega_1, \ldots, \omega_{n-1})$

eine Wahrscheinlichkeitsverteilung auf Ω definiert mit folgenden Eigenschaften:

1) $P(X_1 = a_1) = P_1(a_1)$ für alle $a_1 \in \Omega_1$

2) $P(X_i = a_i|X_1 = a_1, X_2 = a_2, ..., X_{i-1} = a_{i-1}) = P_i(a_i|a_1, ..., a_{i-1})$

mit $(X_1 = a_1, X_2 = a_2, ..., X_{i-1} = a_{i-1}) = (\{X_1 = a_1\} \cap ... \cap \{X_{i-1} = a_{i-1}\})$

Der obige Ansatz ist der einzige mit den Eigenschaften 1) und 2).

Mehrstufige Vorgänge in der Schule:

1. Pfadregel:

Die Wahrscheinlichkeit eines Pfades ist gleich dem Produkt der Wahrscheinlichkeiten entlang dieses Pfades.

2. Pfadregel:

Die Wahrscheinlichkeit eines Ereignisses ist gleich der Summe der Wahrscheinlichkeiten derjenigen Pfade, die für dieser Ereignis günstig sind.

Beispiel: Polyasches Urnenmodell

- einfaches Modell zur Ausbreitung ansteckender Krankheiten
- Urne mit w weißen und s schwarzen Kugeln
- Kugel ziehen, Farbe notieren und Kugel und c Kugeln derselben Farbe zurücklegen
- n mal wiederholen
- $c \in N = \{1, 2, ...\}$

für n=3:

$$P((s, s, w)) = \frac{s \cdot (s + c) \cdot w}{(s + w)(s + w + c)(s + w + 2c)}$$

allgemein:

Es kommt nur auf die Anzahl der schwarzen Kugeln an (bzw. weißen).

Wir betrachten zunächst einen Pfad:

$$\omega = (\omega_1, ..., \omega_n)$$

$$P(\omega) = \frac{s(s+c)(s+2c)\cdot ... \cdot(s+(k-1)c)\cdot w(w+c)\cdot ... \cdot(w+(n-k-1)c)}{(s+w)(s+w+c)\cdot ... \cdot(s+w+(n-1)c)}$$

mehrere Pfade:
P(unter den n Kugeln sind k schwarze)=$\binom{n}{k} P(\omega)$

Wiederholung: Definition 1.4: bedingte Wahrscheinlichkeit

Definition 1.5: vollständiges Ereignisstem

Eine Menge von Ereignissen (B_k) mit $B_k \subset \Omega$ heißt vollständiges Ereignissystem (oder Zerlegung) von Ω, falls:

1) $P(B_k) > 0$ für $k = 1, 2, ...$
2) B_k sind paarweise unvereinbar
3) $\Omega = \dot{\bigcup} B_k$.

Satz 1.6: volle/ totale Wahrscheinlichkeit

Es sei (B_k) eine Zerlegung von Ω und A ein Ereignis, dann gilt:

$$P(A) = \sum_k P(A|B_k)P(B_k).$$

Bemerkung:

(B_k) Zerlegung

A ist eingetreten $\rightarrow P(B_k|A) = ?$

$P(B_k)$: a-priori-Wahrscheinlichkeiten (vor dem Versuch)

$P(B_k|A)$: a-posteriori-Wahrscheinlichkeiten (nach dem Versuch)

Satz 1.7: Bayesssche Formel

Ist $P(A) > 0$ und (B_k) eine Zerlegung von Ω, dann gilt:

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum\limits_i P(A|B_i)P(B_i)}.$$

Beispiele:

1. HIV-Suchtest

   - sucht nach Antikörpern gegen das Virus

   - bei einer HIV- infizierten Person positives Testergebnis mit einer Wkt. von 99,8%

   - bei einer nicht HIV- infizierten Person negatives Testergebnis mit einer Wkt. von 99%

   - Gesamtbevölkerung 2004: 0,01% HIV-Infizierte

   V: zufällig ausgewählte Person ist HIV-infiziert

   T_+: positives Testergebnis

   T_-: negatives Testergebnis

$$\begin{aligned}
P(T_+|V) &= 0,998 \\
P(T_{|}\bar{V}) &= 0,99 \\
P(V|T_+) &= \frac{P(V \cap T_+)}{P(T_+)} \\
&= \frac{P(T_+|V)P(V)}{P(T_+|V)P(V) + P(T_+|\bar{V})P(\bar{V})} \\
&= \frac{0,998 \cdot 0,001}{0,998 \cdot 0,001 + 0,01 \cdot 0,999} \approx 0,091
\end{aligned}$$

2. 3-Türen-Problem

   A_i: Auto hinter Tür i

   K_i: Kandidat wählt Tür i

   M_i: Moderator wählt Tür i

$$P(A_3|K_2 \cap M_1) = \frac{P(A_3 \cap K_2 \cap M_1)}{P(K_2 \cap M_1)} = \frac{2}{3}$$

   Wechsler gewinnt genau dann, wenn er am Anfang nicht das Auto getippt hat. $P = \frac{2}{3}$

Nicht- Wechsler gewinnt genau dann, wenn er von Anfang an das Auto getipppt hat.

$P = \frac{1}{3}$

3. Das Simpson- Paradoxon

	Bewerber m	**zugelassen m**	in %	**Bewerber w**	**zugelassen w**	in %
F_1	900	720	80	200	180	90
F_2	100	20	20	800	320	40
	1000	740	74	1000	500	50

Quote/ gewichtetes Mittel: $0,74 = 0,8 \cdot 0,9 + 0,2 \cdot 0,1$

Z: Zulassung

F_1, F_2: Fächer

M,W: Geschecht

$$\begin{aligned}
P(Z|W \cap F_i) &> P(Z|M \cap F_i) \\
P(Z|M) &= \frac{P(Z \cap M)}{P(M)} \\
&= \frac{P(Z \cap M \cap F_1) + P(Z \cap M \cap F_2)}{P(M)} \\
&= \frac{P(Z \cap M \cap F_1)}{P(M \cap F_1)} \cdot \frac{P(M \cap F_1)}{P(M)} + \frac{P(Z \cap M \cap F_2)}{P(M \cap F_2)} \cdot \frac{P(M \cap F_2)}{P(M)} \\
&= P(Z|M \cap F_1) \cdot \underbrace{P(F_1|M)}_{1-\alpha} + P(Z|M \cap F_2) \cdot \underbrace{P(F_2|M)}_{\alpha}
\end{aligned}$$

Unabhängigkeit von Ereignissen

Information über das Eintreten von B, ändert nichts an den Chancen von A. Es soll symmetrisch gelten:

$$P(A|B) = P(A)$$

Definition 1.6: Unabhängigkeit

Zwei Ereignisse heißen unabhängig (voneinander), falls die Produktformel gilt:

$$P(A \cap B) = P(A) \cdot P(B).$$

Bemerkungen:

1. Die Unabhängigkeit ist symmetrisch bezüglich A und B.
2. Die Definition ist auch anwendbar, wenn $P(A) = 0$ und/ oder $P(B) = 0$.
3. Unabhängigkeit ist mehr als fehlender Kausalzusammenhang.
4. Unabhängikeit ist oft eine Modellannahme. z.B. Mendel'sche Gesetze

Vierfeldertafeln

Beispiel:

	M	W	M+W
"normal"	0,459	0,498	0,957
rot- grün- blind	0,040	0,002	0,042
Summe	0,499	0,500	1

allgemein:

	A	$\overline{A}$	
B	$P(A \cap B)$	$P(\overline{A} \cap B)$	$P(B)$
$\overline{B}$	$P(A \cap \overline{B})$	$P(\overline{A} \cap \overline{B})$	
	$P(A)$		

Satz 1.8:

Sind die Ereignisse A und B unabhängig, dann sind es auch die Ereignisse A und $\bar{B}$, $\bar{A}$ und B sowie $\overline{A}$ und $\overline{B}$.

Definition 1.7: Unabhängigkeit einer Familie von Ereignissen

Eine Familie (A_k) von Ereignissen heißt unabhängig, falls für jede endliche Teilfamilie die Produktformel:

$$P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{in}) = P(A_{i1}) \cdot P(A_{i2}) \cdot \ldots \cdot P(A_{in})$$

gilt.

D.h. n Ereignisse sind unabhängig

$\Leftrightarrow$ für jede Teilmenge von diesen Ereignissen die Produktformel gilt.

Operationen mit unabhängigen Ereignissen

Satz 1.9:

Sind $A_1, ..., A_n$ unabhängige Ereignisse und ist für k das Ereignis B_k gleich A_k oder gleich $\bar{A}_k$, so sind auch die Ereignisse $B_1, ..., B_k$ unabhängig.

Satz 1.10:

Ist $(A_k)_{k \in I}$ eine Familie von unabhängigen Ereignissen und $I_1 \dot{\cup} \ldots \dot{\cup} I_m = I$ eine Zerlegung von I und wird für jedes k das Ereignis B_k durch Mengenoperation $\cup$, $\cap$ oder $\bar{\ }$ aus Ereignissen A_i mit $i \in I_k$ gebildet, so ist auch die Familie (B_k) eine Familie von unabhängigen Ereignissen.

1.7 Produktexperimente

- Gesamtvorgang bestehend aus n- "unabhängigen" Teilvorgängen
- i-te Teilvorgang (Ω_i, P_i)
- Ergebnismenge Ω für Gesamtvorgang: $\Omega = \Omega_1 \times \Omega_2 \times ... \times \Omega_n$

$$\omega = (\omega_1, \omega_2, ..., \omega_n) \in \Omega$$

$$P(\omega) = P_1(\omega_1) \cdot P_2(\omega_2) \cdot ... \cdot P_n(\omega_n)$$

Bernoulli Ketten und Binomialverteilung

Einzelvorgang: Bernoulli-Experiment

$\Omega_i = \{0, 1\}$

$P_i(1) = p$ $\quad 0 \leq p \leq 1$ $\quad$ - Erfolgswahrscheinlichkeit

$P_i(0) = 1 - p$ $\quad$ - Misserfolgswahrscheinlichkeit

Bernoulli Kette (BK) der Länge n:

- n Bernoulli-Experimente werden unabhängig durchgeführt

$$\Omega = \{(\omega_1, \omega_2, ..., \omega_n) : \omega_i \in \{0, 1\}\}$$

$$P((\omega_1, ..., \omega_n)) = p^{\sum\limits_{i=1}^{n} \omega_i} \cdot (1-p)^{n - \sum\limits_{i=1}^{n} \omega_i}$$

P hängt von der Anzahl der Einsen ab.

Die Anzahl der Erfolge in einer Bernoulli Kette ergibt sich aus:

$$X(\omega) = \sum_{i=1}^{n} \omega_i$$

Für X sind die Werte 0,1,...,n möglich.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Definition 1.8: Binomialverteilung

Die Verteilung der Anzahl der Erfolge in einer BK der Länge n mit der Erfolgswahrscheinlichkeit p heißt Binomialverteilung mit den Parametern n und p.

$X \sim B(n, p)$

Folien 3)-6)

Eigenschaften der Binomialverteilung:

Rekursionsformel: $p < 1$

$p_0 = (1-p)^n$

$p_{k+1} = \frac{n-k}{k+1} \cdot \frac{p}{1-p} \cdot p_k$ $\quad k = 0, 1, ..., n-1$

Satz 1.11:

Für den/ die wahrscheinlichsten Wert/-e k^* einer binomialverteilten ZG gilt:

$$np - (1-p) \leq k^* \leq np + p.$$

Etwas ungenauer: np - Erfolge.

Folie: Galtonbrett

Bsp.: Blumensträuße für Geburtstagskinder

$n = 400$ Personen

$p = \frac{1}{365}$ (Annahme)

$np = 400 \cdot \frac{1}{365} \approx 1$

X - Anzahl der Geburtstagskinder

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - \left[\binom{400}{2} \left(\frac{1}{365}\right)^2 \left(\frac{364}{365}\right)^{398} + \binom{400}{1} \left(\frac{1}{365}\right) \left(\frac{364}{365}\right)^{399} + \left(\frac{364}{365}\right)^{400} \right] \\ &\approx 0,10 \end{aligned}$$

Approximation der Binomialverteilung durch die Poissonverteilung

Satz 1.12: Satz von Poisson

Die Erfolgswahrscheinlichkeit sei von n abhängig: $p(n)$,

es gelte:

$$\lim_{n\to\infty} n \cdot p(n) = \lambda > 0,$$

dann gilt:

$$\lim_{n\to\infty} \binom{n}{k} (p(n))^k (1-p(n))^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda} \qquad [k = 0, 1, \ldots].$$

Definition 1.9: Poisson-Verteilung

Eine ZG X heißt poissonverteilt $X \sim \Pi(\lambda)$, wenn für ihre Wahrscheinlichkeiten gilt:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$k = 0, 1, ... \qquad \lambda > 0.$$

Poissonverteilungen werden auch als Verteilungen der seltenen Ereigisse bezeichnet.

Warum?

Wegen der langen BK mit kleiner Erfolgswahrscheinlichkeit.

Bsp.: zum Blumenstraußproblem

- ersetze n-tes Folgeglied durch Grenzwert

$n = 400 \qquad\qquad p = \frac{1}{365}$

$np = 1,096 =: \lambda_n$

$P(X \leq 2) \approx \left(1 + 1,096 + \frac{1,096^2}{2!}\right) e^{-1,096} = 0,90$

Wie genau ist dieser Wert?

Satz 1.13:

Es sei: $np(n) = \lambda > 0$, dann gilt:

$$\sum_{k=0}^{\infty} \left| \binom{n}{k} (p(n))^k (1 - p(n))^{n-k} - \frac{\lambda^k}{k!} e^{-\lambda} \right| \leq \frac{2\lambda}{n} \cdot \min(\lambda, 2).$$

Hierbei ist $\binom{n}{k} = 0$ für $k \geq n$.

Bsp.: Fehlerabschätzung zum Blumenstraußproblem

$|P(X \leq 2) - \Pi_\lambda(2)|$

$= |[P(X = 0) - \Pi_\lambda(0)] + [P(X = 1) - \Pi_\lambda(1)] + [P(X = 2) - \Pi_\lambda(2)]|$

$$= \left| \left[\binom{400}{0} p^0 (1-p)^{400} - \frac{\lambda^0}{0!} e^{-\lambda}\right] + \left[\binom{400}{1} p^1 (1-p)^{399} - \frac{\lambda^1}{1!} e^{-\lambda}\right] + \left[\binom{400}{2} p^2 (1-p)^{398} - \frac{\lambda^2}{2!} e^{-\lambda}\right] \right|$$

$$\underbrace{\leq}_{1} \left| \left[\binom{400}{0} p^0 (1-p)^{400} - \frac{\lambda^0}{0!} e^{-\lambda}\right] \right| + |[\ldots]| + \left| \left[\binom{400}{2} p^2 (1-p)^{398} - \frac{\lambda^2}{2!} e^{-\lambda}\right] \right|$$

$$\leq \sum_{k=0}^{\infty} \left| \binom{400}{k} p^k (1-p)^{400-k} - \frac{\lambda^k}{k!} e^{-\lambda} \right|$$

$$\underbrace{\leq}_{2} \frac{2 \cdot 1,096}{400} \cdot 1,096 = 0,006$$

1 Dreiecksungleichung

2 Satz 1.13)

Zusammenhang zwischen hypergeometrischer Verteilung und Binomialverteilung:

$N = R + \underbrace{(N-R)}_{nicht-Rote-Kugeln}$

n- Stichprobenumfang

X- Anzahl der roten Kugeln in der Stichrobe

mit Zurücklegen	ohne Zurücklegen
$p = \frac{R}{N}$	p ist abhängig davon, wie viele
	rote Kugeln schon gezogen wurden
(klassisches Bernoulliexperiment)	
$P(X=k) = \binom{n}{k}\left(\frac{R}{N}\right)^k\left(1-\frac{R}{N}\right)^{n-k}$	$P(X=k) = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}$
$k = 0, 1, ..., n$	$k = 0, 1, ..., \min(k, R)$

Eine Approximation ist bei großem N und kleinem n möglich.

Satz 1.14:

Es gelte:

$$\lim_{N,R\to\infty} \frac{R}{N} = p \in [0,1]\,,$$

dann gilt für jedes k:

$$\lim_{N,R\to\infty} \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}} = \binom{n}{k}(p)^k(1-p)^{n-k}\,.$$

1.8 Erwartungswert und Varianz von ZG

Erwartunswert und Varianz sind Kenngrößen der Lage und der Streuung der Verteilung einer ZG.

Verteilung der ZG X: $\dfrac{x_1 \quad x_2 \quad ... \quad x_k \quad ...}{p_1 \quad p_2 \quad ... \quad p_k \quad ...}$

Definition 1.10: Erwartungswert

Ist (Ω, P) ein diskreter Wraum und $X : \Omega \to R$ eine ZG, für die die Reihe

$$\sum_{\omega\in\Omega} |X(\omega)|\, P(\omega)$$

konvergiert, dann heißt:

$$EX = \sum_{\omega\in\Omega} X(\omega)P(\omega)$$

der Erwartungswert von X.

Bemerkung:

Der Wert $X(\omega)$ wird gewichtet mit seiner Wahrscheinlichkeit, d.h. EX ist ein gewichtetes Mittel.

Satz 1.15:

Falls EX existiert, gilt:

$$\sum_k x_k \cdot P(X = x_k) = \textstyle\sum_k x_k \cdot p_k.$$

Interpretation der Erwartungswerte:

1. Wir beobachten X n-mal.

   Häufigkeitsverteilung: $\dfrac{x_1 \quad x_2 \quad ... \quad x_r}{h_n(x_1) \quad h_n(x_2) \quad ... \quad h_n(x_r)}$

   Der Wert $h_n(x_k)$ ist die relative Häufigkeit von x_k bei n Beobachtungen.

   arithmetisches Mittel:

   $$\bar{x} = x_1h_n(x_1) + x_2h_n(x_2) + ... + x_rh_n(x_r)$$

   angenommen n ist sehr groß:

   $$\bar{x} \hookrightarrow x_1p_1 + ... + x_rp_r = EX$$

Beobachtungsebene	Modellebene
$h_n(A)$	$P(A)$
$\bar{x}$	EX

2. Wo ist das Gleichgewicht?

- EX markiert den Schwerpunkt der Verteilung

- EX ist empfindlich gegen Ausreißer

Bsp.: Warten auf den ersten Erfolg in einer BK

X- Anzahl der Versuche bis zum 1.Erfolg einschließlich

$P(X = k) = p(1-p)^{k-1} \qquad\qquad k = 1, 2, \dots$

$$\begin{aligned} EX &= \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \cdot p \\ &= p \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \\ &= p \frac{1}{(1-(1-p))^2} \\ &= \frac{1}{p} \end{aligned}$$

Nebenrechnung: $z = 1 - p$

$$\begin{aligned} \sum_{k=1}^{\infty} z^k &= \frac{z}{1-z} \\ \sum_{k=1}^{\infty} k z^{k-1} &= \frac{1-z+z}{(1-z)^2} \\ &= \frac{1}{(1-z)^2} \end{aligned}$$

Definition 1.11: faires Spiel

Ein 2-Personenspiel heißt fair, falls der Erwartungswert des Gewinns eines Spielers gleich seinem Einsatz ist.

Beispiele:

1. Roulette

   - 37 Felder

   - die Null ist ein besonders ausgezeichnetes Feld

   - man setzt den Einsatz e auf die 13

   G (Gewinn): $\begin{array}{cc} 0 & 36e \\ \hline \frac{36}{37} & \frac{1}{37} \end{array}$

   $EG = 0 \cdot \frac{36}{37} + 36e \cdot \frac{1}{37} = \frac{36}{37}e < e$

NG (Nettogewinn): $\frac{\text{-e} \quad 35\text{e}}{\frac{36}{37} \quad \frac{1}{37}}$

$E(NG) = -e \cdot \frac{36}{37} + \frac{35}{37}e = -\frac{1}{37}e < 0$

2. Versicherung

   Bruttoprämie = Nettoprämie + Verwaltungskosten

   Die Nettoprämie entspricht dem Risikoanteil.

   Berechnung der Nettoprämie:

   Versichert wird das Ereignis A mit einer Versicherungssumme C.

   X- Leistung der Versicherung

   X: $\frac{0 \qquad\qquad C}{1 - P(A) \quad P(A)}$

   $EX = C \cdot P(A)$ (Versicherungsmathematisches Äquivalenzprinzip)

   Nettoprämie $= C \cdot P(A)$

3. Wer wird Millionär

   Der Kandidat hat 16.000 Euro sicher, hat die 64.000 Euro Frage bereits richtig beantwortet und steht nun vor der 125.000 Euro Frage.

   Stellt der Erwartungswert hier eine adäquate Entscheidungshilfe?

   G: $\frac{16.000 \quad 125.000}{1 - p \qquad\quad p}$

   $EG = 16.000(1 - p) + 125.000p = 16.000 + 109.000p$

4. Indikatorvariable

   $$I_A(\omega) = \begin{cases} 1 & : \quad \omega \in A \\ 0 & : \quad \omega \notin A \end{cases}$$

   $E(I_A) = P(A)$

Eigenschaften des Erwartungswertes

Satz 1.16:

Seien X und Y Zufallsgrößen mit existierenden Erwartungswerten und $\lambda \in \mathcal{R}$.

Dann gilt:

1) $E(\lambda X) = \lambda EX$
2) $E(X+Y) = EX + EY$
3) Wenn f eine reelle Funktion $f : \mathcal{R} \to \mathcal{R}$ ist, dann gilt:
   $E(f \circ X) = E(f(X)) = \sum f(X(\omega))P(\omega) = \sum f(x_k)p_k.$

Bemerkung:

- E ist eine lineare Abbildung aus dem Vektorraum der ZG (mit existierenden Erwartungswerten) in die Menge der reellen Zahlen
- Analogie zu Integralen
- 2) lässt sich verallgemeinern:
  $E(X_1 + \cdots + X_n) = EX_1 + \cdots + EX_n$

Erwartungswert einer binomialvereilten ZG X

$X \sim B(n,p)$

$\Omega = \{(\omega_1, \ldots, \omega_n) : \omega_1 \in \{0,1\}\}$

$X(\omega) = \omega_1 + \cdots + \omega_n$

$A_i = \{\omega \in \Omega : \omega_1 = 1\} \qquad ; (i = 1, \ldots, n)$

$X = I_{A1} + I_{A2} + \cdots + I_{An}$

$$\begin{aligned} EX &= E(I_{A1}) + E(I_{A2}) + \cdots + E(I_{An}) \\ &= P(A_1) + P(A_2) + \cdots + P(A_n) \\ &= \underbrace{p + p + \cdots + p}_{n-mal} \\ &= np \end{aligned}$$

alternativ:

$$EX = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

Bemerkungen:

- EX ist nahe dem wahrscheinlichsten Wert (bei binomialverteilter ZG)
- im allgemeinen ist das *nicht* so
- im allgemeinen kommt EX unter den Werten der ZG gar nicht vor

Bsp.: Sammelbilderproblem

- N Sammelbilder $1, 2, \ldots, N$

- Ziel: $r \leq N$ verschiedene Bilder sammeln

Wie lange dauert das im Durchschnitt?

$Y_1 = 1$

Y_{k+1} - Wartezeit vom k-ten Bild bis zum (k+1)-ten verschiedenen Bild

$$\underbrace{1}_{Y_1}\underbrace{001}_{Y_2}\ldots\underbrace{00001}_{Y_{k+1}}$$

Y_{k+1} - ist geometrisch vereilt mit Erfolgswahrscheinlichkeit p Annahmen:

1) alle Bilder kommen gleich oft vor

2) gut gemischt

$$p = \frac{N-k}{N}$$

$$EY_{k+1} = \frac{1}{p} = \frac{N}{N-k}$$

$$EY_k = \frac{N}{N-(k-1)}$$

T_r - Zeit bis zum r-ten Sammelbild

$$\begin{aligned}
T_r &= Y_1 + Y_2 + \ldots + Y_r \\
ET_r &= EY_1 + EY_2 + \ldots + EY_r \\
&= 1 + \frac{N}{N-1} + \frac{N}{N-2} + \ldots + \frac{N}{N-(r-1)} \\
&= N(\frac{1}{N} + \frac{1}{N-1} + \ldots + \frac{1}{N-(r-1)} & |(r = N) \\
ET_n &= N \qquad \underbrace{\left(\frac{1}{N} + \frac{1}{N-1} + \ldots + \frac{1}{1}\right)}_{N-te-Partialsumme-der-harmonischen-Reihe}
\end{aligned}$$

Näherungsformel von Euler:

$$\lim_{n\to\infty} = (\underbrace{1+\frac{1}{2}+\cdots+\frac{1}{n}}_{s_n} - \ln c) = c$$

$c \approx 0,577$ Euler'sche Konstante

Also:

$$\begin{aligned} s_n &\ \widetilde{as.}\ \ln n + c, \qquad\qquad d.h. \lim_{n\to} \frac{s_n}{\ln n + c} = 1 \\ ns_n &\ \widetilde{as.}\ n(\ln n + c) \\ ET_N &\ \widetilde{as.}\ N(\ln N + c) \end{aligned}$$

Beispiel:

N	$ET_{\frac{N}{2}}$	ET_N	$N(\ln N + 0,577)$
30	20	120	119
50	34	225	224
100	71	519	518

Ein anderer Mittelwert- der Median

Beispiele:

1. Im Jahr 2003 betrug das durchschnittliche Nettovermögen pro Haushalt 133.000 EUR.
   Median: 50.000 EUR
   Der Median bedeutet hier, dass ca. die Hälfte der Leute mehr bzw. weniger als 50.000 EUR hatten.

2. X: Anzahl der Würfe bis zur ersten Sechs
   $EX = 6$
   Median: 4
   Die Wahrscheinlichkeit, dass die Anzahl der Würfe bis zur ersten Sechs ≤ 4 ist, beträgt ca. 0,53.
   Die Wahrscheinlichkeit, dass die Anzahl der Würfe bis zur ersten Sechs ≥ 4 ist, beträgt ca. 0,57.

Definition 1.12: Median

Ein Median einer ZG X ist eine Zahl m, für die gilt:

$$P(X \leq m) \geq 0,5 \qquad \text{und} \qquad P(X \geq M) \geq 0,5.$$

D.h. mit einer Wahrscheinlichkeit von mind. 50 % sind die Werte der ZG kleiner oder gleich dem Median, und ebenso für $\geq m$. Der Median ist unempfindlich gegen Außreißer.

Schätzwert für den Median:

Wir haben $x_1, x_2, \ldots, x_n$ unabhängige Beobachtungen der ZG X.

Diese werden in der Ordnungsstatistik der Größe nach geordnet:

$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$

Definition 1.13: Zentralwert

Der empirische Median (Zentralwert) $\tilde{x}$ ist definiert als:

$$\tilde{x} = \begin{cases} \frac{x_{(m)}+x_{(m+1)}}{} & : \quad n = 2m \\ x_{(m+1)} & : \quad n = 2m+1 \end{cases}$$

Mindestens die Hälfte der x_i ist $\leq \tilde{x}$ und mindestens die Hälfte der x_i ist $\geq \tilde{x}$.

Beispiel:

12,8,4,4,3,10,3,5,4,4,3

- ordnen der Größe nach

3,3,3,4,4,4,5,8,10,12

$n = 1$

$x_{(n+1)} = 4$

Beobachtungsebene	Modellebene
$h_n(A)$	$P(A)$
$\bar{x}$	EX
$\tilde{x}$	Median(X)
s^2	Var(X)
s	$\sigma(X)$

Streuung einer ZG X um EX

Beispiel:

X: $\begin{array}{ccc} 2 & 4 & 6 \\ \hline \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}$

Y: $\begin{array}{ccc} 2 & 4 & 6 \\ \hline \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \end{array}$

Z: $\begin{array}{ccc} 2 & 4 & 6 \\ \hline \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \end{array}$

$EX = EY = EZ = 4$

Vermutung: Z streut am wenigsten um den Erwartungswert und Y am meisten.

Wie findet man ein geeignetes Maß für die Streuung?

$$X - EX \to (X - EX)^2 \to E(X - EX)^2$$

Definition 1.14: Varianz

Sei X eine ZG, für die EX^2 existiert.

Dann heißt:

$$VarX = E(X - EX)^2$$

die Varianz von X und:

$$\sigma(X) = \sqrt{VarX}$$

die Standardabweichung von X.

weiter mit unserem Beispiel:

$$\begin{aligned} VarX &= (2-4)^2 \cdot \frac{1}{3} + (4-4)^2 \cdot \frac{1}{3} + (6-4)^2 \cdot \frac{1}{3} = \frac{1}{8} \\ VarY &= (2-4)^2 \cdot \frac{3}{8} + (4-4)^2 \cdot \frac{1}{4} + (6-4)^2 \cdot \frac{3}{8} = \frac{24}{8} = 3 \\ VarZ &= (2-4)^2 \cdot \frac{1}{8} + (4-4)^2 \cdot \frac{3}{4} + (6-4)^2 \cdot \frac{1}{8} = 1 \end{aligned}$$

Berechnung der Varianz:

$Ef(x)$ mit $f(x) = (x - EX)^2$

$VarX = \sum_k (x_k - EX)^2 \cdot p_k$

Definition 1.15: empirische Streuung

Die empirische Streuung der Beobachtungswerte $x_1, \ldots, x_n$ ist definiert als:

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Dann heißt:

$$s = \sqrt{s^2}$$

empirische Standardabweichung.

Eine andere Beziehung zwischen Varianz und EW

X: $\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \\ \hline p_1 & p_2 & \cdots & p_n \end{array}$

Vorhersage: v

Verlustfunktion: $(X - v)^2$

Welche Vorhersage minimiert den EW des Verlustes? $(EX = v)$

(Eine andere Beziehung: Verlustfunktion $|X - v| \to v =$ Median (X))

$$\begin{aligned} f(v) &= E(X-v)^2 \\ &= E(X^2 - 2vX + v^2) \\ &= EX - 2vEX + v^2 \\ &= (v - EX)^2 + EX^2 - (EX)^2 \end{aligned}$$

- f ist minimal für $v = EX$ und es gilt:

$$E(X-EX)^2 = EX^2 - (EX)^2$$

Beispiel: Varianz der Binomialverteilung

$X \sim B(n,p)$

$X = I_{A_1} + I_{A_2} + \ldots + I_{A_n}, \qquad A_k = \{\omega \in \Omega : \omega_1 = 1\}$

$EX = n \cdot p$

$$\begin{aligned}
EX^2 &= E(I_{A_1} + I_{A_2} + \ldots + I_{A_n})^2 \\
&= E(I_{A_1}^2 + I_{A_2}^2 + \ldots + I_{A_n}^2 + \sum_{i \neq j} I_{A_i} I_{A_j}) \\
&= E(I_{A_1} + I_{A_2} + \ldots + I_{A_n} + \sum_{i \neq j} I_{A_i \cap A_j}) \\
&= EI_{A_1} + EI_{A_2} + \ldots + EI_{A_n} + \sum_{i \neq j} EI_{A_i \cap A_j} \\
&= P(A_1) + P(A_2) + \ldots + P(A_n) + \sum_{i \neq j} P(A_i \cap A_j) \\
&= n \cdot p + \sum_{i \neq j} p^2 \\
EX^2 &= np + (n^2 - n)p^2
\end{aligned}$$

$$\begin{aligned}
VarX &= EX^2 - (EX)^2 \\
&= np + (n^2 - n)p^2 - (np)^2 \\
&= np + n^2p^2 - np^2 - n^2p^2 \\
VarX &= np - np^2 \\
&= np(1-p)
\end{aligned}$$

Die Varianz ist ein Maß für die Streuung um den EW einer ZG.

Wie verhält sich die Varianz einer binomialverteilten ZG?

$X \sim B(n,p)$

$EX = n \cdot p$

$VarX = n \cdot p \cdot (1-p)$

1. n fest, p variabel

   $p = 0 \quad\quad \Rightarrow VarX = 0$

   $p = 1 \quad\quad \Rightarrow VarX = 0$

   $p = \frac{1}{2} \quad\quad \Rightarrow VarX = \frac{n}{4}$

   In $p = \frac{1}{2}$ nimmt die Varianz ihr Maximum an, da hier die größte Unsicherheit im Einzelversuch vorliegt.

2. p fest z.B. $p = \frac{1}{4}$, n variabel

   $VarX = n\frac{1}{4}\frac{3}{4} \stackrel{n\to\infty}{\longrightarrow} \infty$ Fazit: Je mehr Versuche man macht, mit desto größeren Abweichungen vom $EW = np$ muss man rechnen.

Satz 1.17: weitere Eigenschaften der Varianz

Sind X und Y ZG, für die EX^2 und EY^2 existieren,

so gilt:

1. für beliebige $a, b \in R$

   $Var(aX + b) = a^2 VarX$

2. $Var(X + Y) = VarX + VarY + 2E\left[(X - EX)(Y - EY)\right]$

Warum ist es plausibel, dass b in 1. verschwindet?

Man kann die ZG verschieben, dabei bleibt die Streuung erhalten.

Bemerkungen:

1. Die Varianz ist keine lineare Abbildung.

2. Aus der Existenz von EX^2, folgt die Existenz von EX.

3. Die Darstellung $VarX = EX^2 - (EX)^2$ ist manchmal günstiger.

Standardisieren einer ZG:

Sei X eine ZG für die EX und EX^2 existieren.

$$X^* = \frac{X - EX}{\sqrt{VarX}}$$

$$EX^* = E\left(\frac{X - EX}{\sqrt{Var}}\right)$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{VarX}} \cdot E\,(X - EX) \\
&= \frac{1}{\sqrt{VarX}} \cdot EX - E(EX) = 0 \\
VarX^* &= Var\left(\frac{X - EX}{\sqrt{Var}}\right) \\
&= \frac{1}{VarX} \cdot Var(X - EX) \\
&= \frac{1}{VarX} \cdot VarX = 1
\end{aligned}
$$

X^* heißt die standardisierte ZG zu X.

1.9 Unabhaengigkeit von ZG, Kovarianz und Korrelation

Seien X und Y zwei ZG mit den folgenden Verteilungen:

X: $\dfrac{x_1 \quad x_2 \quad ,\ldots, \quad x_r \quad ,\ldots}{p_1 \quad p_2 \quad ,\ldots, \quad p_r \quad ,\ldots}$

Y: $\dfrac{y_1 \quad y_2 \quad ,\ldots, \quad y_s \quad ,\ldots}{q_1 \quad q_2 \quad ,\ldots, \quad q_s \quad ,\ldots}$

Vorüberlegungen:

Wann sind zwei ZG unabhängig?

$$
\begin{aligned}
P(Y = y_i | X = x_1) &= P(Y = y_i) \\
\frac{P(Y = y_i, X = x_1)}{P(X = x_1)} &= P(Y = y_i) \\
P(Y = y_i, X = x_1) &= P(X = x_1)P(Y = y_i)
\end{aligned}
$$

Um die Unabhängigkeit zu gewährleisten muss die obige Beziehung nicht nur für ein bestimmtes x_1 gelten sondern für alle möglichen x_j.

$$
\begin{aligned}
P(Y = y_i | X = x_j) &= P(Y = y_i) \\
\frac{P(Y = y_i, X = x_j)}{P(X = x_j)} &= P(Y = y_i) \\
P(Y = y_i, X = x_j) &= P(X = x_j)P(Y = y_i)
\end{aligned}
$$

Definition 1.16: Unabhängigkeit von zwei ZG

Zwei ZG X, Y heißen unabhängig, wenn für alle x_j und alle y_i gilt:

$$P(X = x_j, Y = y_i) = P(X = x_j)P(Y = y_i).$$

Gemeinsame Verteilung von X und Y:

X Y	y_1	y_2	$\cdots$	y_s	$\cdots$	Randverteilung von X	
x_1	p_{11}	p_{12}	$\cdots$	p_{1s}	$\cdots$	$p_{1\bullet}$	
x_2	p_{21}	p_{22}	$\cdots$	p_{2s}	$\cdots$	$p_{2\bullet}$	
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	
x_r	p_{r1}	p_{r2}	$\cdots$	p_{rs}	$\cdots$	$p_{r\bullet}$	$= P(X = x_r)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	
Randverteilung von Y	$p_{\bullet 1}$	$p_{\bullet 2}$	$\cdots$	$p_{\bullet s}$	$\cdots$		

Es gilt:

$$p_{rs} = P(X = x_r, Y = y_s)$$

Bemerkungen:

Im allgemeinen ist die gemeinsame Verteilung durch die Randverteilung nicht eindeutig bestimmt.

Bei Unabhängigkeit ergibt sich die gemeinsame Verteilung aus den Produkten der Randverteilung.

Beispiele:

1. Für welches c liegt Unabhängigkeit vor?

X Y	0	1	
0	$\frac{1}{2} - c$	c	$\frac{1}{2}$
1	c	$\frac{1}{2} - c$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

   Nur bei $c = \frac{1}{4}$ liegt Unabhängigkeit vor.

2. BK, $n = 3$

   X- Anzahl der Erfolge

$$Y = \begin{cases} Nr.\ des\ 1.Erfolges & : \quad falls\ \ X > 0 \\ 0 & : \quad falls\ \ X = 0 \end{cases}$$

$q =: 1 - p$

Y X	0	1	2	3	
0	q^3	0	0	0	q^3
1	0	pq^2	pq^2	pq^2	$3pq^2$
2	0	$2p^2q$	p^2q	0	$3p^2q$
3	0	p^3	0	0	p^3
	q^3	p	pq	pq^2	1

Für n beliebig gilt:

$$P(X = k, Y = l) = \binom{n-l}{k-1} p^k (1-p)^{n-k}$$

Für $0 < p < 1$ sind X und Y nicht unabhängig.

Bemerkung:

Unabhängigkeit von ZG ist oft eine Modellannahme.

Satz 1.18:

Sind X und Y unabhängige ZG und A, B $\subseteq R$, so gilt:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

A, B repräsentieren beliebige Aussagen über die Werte von X bzw. Y.

Definition 1.17: Unabhängigkeit von mehreren ZG

Die ZG $X_1, X_2, \ldots, X_n$ heißen unabhängig, falls für beliebige $A_1, A_2, \ldots, A_n \subseteq R$ die Produktformel gilt:

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdot \ldots \cdot P(X_n \in A_n).$$

Bemerkung:

1. Aus der Unabhängigkeit von n ZG folgt die Unabhängigkeit jeder endlichen Teilfamilie von ZG.

2. Aus der paarweisen Unabhängigkeit folgt nicht die Unabhängigkeit von n ZG.

Beispiel:

Paul (25) und Paula (20) heiraten heute.

Mit welcher Wkt. könnten sie ihre Silberhochzeit erleben?

X: restliches zufälliges Lebensalter von Paula

Y: restliches zufälliges Lebensalter von Paul

Annahme: X und Y sind unabhängig

Sterbetafel DAV 1994T

100.000 Neugeborene

Alter	Sterbewkt. m	Sterbewkt. w
0	100.000	100.000
$\vdots$		
20		98.366
$\vdots$		
25	96.956	
$\vdots$		
45		95.971
$\vdots$		
50	90.758	

$$\begin{aligned} P(X \geq 25, Y \geq 25) &= P(X \geq 25)P(Y \geq 25) \\ &= \frac{95.971}{98.366} \cdot \frac{90.758}{96.956} \\ &\approx 0,975 \cdot 0,936 \\ &\approx 0,91 \end{aligned}$$

Kenngrößen unabhängiger ZG:

Satz 1.19:

Seien $X_1, X_2, \dots, X_n$ unabhängige ZG, deren Varianzen existieren.

Dann gilt:

1. $E(X_1 \cdot X_2 \cdot \dots \cdot X_n) = EX_1 \cdot EX_2 \cdot \dots \cdot EX_n$
2. $Var(X_1 + X_2 + \dots + X_n) = VarX_1 + VarX_2 + \dots + VarX_n$
3. $cov(X_i, X_j) = 0, \forall i, j$

Wo addieren sich Varianzen?

Warum mittelt man beim Messen einer physikalischen Größe?

$X_1, X_2, \dots, X_n$

Wir nehmen n unabhängige Messungen unter den gleichen Bedingungen vor, d.h. $X_1, X_2, \dots, X_n$ sind unabhängige ZG mit derselben Verteilung.

Insbesondere gilt:

$$EX_i = EX_1$$

und

$$VarX_i = VarX_1$$

$$\begin{aligned}
\bar{X} &= \frac{1}{n}(X_1 + X_2 + \dots + X_n) \\
E\bar{X} &= E(\frac{1}{n}(X_1 + X_2 + \dots + X_n)) \\
&= \frac{1}{n}(EX_1 + \dots + EX_n) \\
&= \frac{1}{n} \cdot n \cdot EX_1 \\
&= EX_1 \\
Var\bar{X} &= Var(\frac{1}{n}(X_1 + X_2 + \dots + X_n)) \\
&= \frac{1}{n^2}(VarX_1 + VarX_2 + \dots + VarX_n) \\
&= \frac{1}{n^2} \cdot n \cdot VarX_1 \\
&= \frac{1}{n} VarX_1 \overset{n \to \infty}{\longrightarrow} 0
\end{aligned}$$

Je öfter ich messe, desto geringer ist die Abweichung vom EW.

$Var\bar{X} = E(\bar{X} - \underbrace{E\bar{X}}_{w-wahrer-Wert})^2$

$EX_1 = w$ bei Mesungen ohne systematischen Fehler

Lineare Regression, Kovarianz und Korrelation

Seien X und Y ZG, für die die Varianzen existieren.

Optimierungsprolem:

$$f(a,b) = E(Y-(a+bX))^2 \to \min$$

nach der Methode der kleinsten Quadrate (Gauß)

$a + bX$ - Regression von Y bezüglich X

$$\begin{aligned}
E(\underbrace{Y-bX}_{=:Z}-a)^2 &= E(Z-EZ+EZ-a)^2 \\
&= E((Z-EZ)^2 + 2(Z-EZ)(EZ-a) + (EZ-a)^2) \\
&= \underbrace{E(Z-EZ)^2}_{unabh. von a} + (EZ-a)^2 \to \min_{a,b} \\
\Rightarrow a_{min} &= EZ \\
f(b) &= E(Z-EZ)^2 \\
&= E(Y-bX-E(Y-bX))^2 \\
&= E(Y-EY-b(X-EX))^2 \\
&= E(Y-EY)^2 - 2bE((X-EX)(Y-EY)) + b^2E(X-EX)^2 \\
&= VarY - 2bCov(X,Y) + b^2 VarX \\
\Rightarrow b_{min} &= \frac{cov(X,Y)}{VarX} \\
\Rightarrow a_{min} &= EZ = EY - \frac{cov(X,Y)}{VarX}EX \\
f(a_{min}, b_{min}) &= VarY - 2\frac{(cov(X,Y))^2}{VarX} + \frac{(cov(X,Y))^2}{(VarX)^2}VarX \\
&= VarY - \frac{(cov(X,Y))^2}{VarX} \\
&= VarY\left(1 - \left(\frac{cov(X,Y)}{\sqrt{VarX}\sqrt{VarY}}\right)^2\right)
\end{aligned}$$

Definition 1.18: Korrelationskoeffizient

Seien X und Y ZG mit $0 < VarX, VarY < \infty$. Dann heißt:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{VarX}\sqrt{VarY}}$$

der Korrelationskoeffizient von X und Y.

Satz 1.20: Lineare Regression

Das Optimierungsproblem:

$$E(Y-(a+bX))^2 \to \min$$

besitzt die Lösung:

$$a_{min} = EY - \frac{cov(X,Y)}{VarX}EX$$
$$b_{min} = \frac{cov(X,Y)}{VarX}.$$

Der minimale Wert beträgt:

$$VarY(1-(\rho(X,Y))^2).$$
$$\hat{Y} = EY + \frac{cov(X,Y)}{VarX}(X-EX)$$

heißt Regression von Y bezüglich X.

Folgerungen:

1. Cauchy-Schwarzesche Ungleichung
   $(cov(X,Y))^2 \leq VarX \cdot VarY$
2. $|\rho(X,Y)| \leq 1$
3. $|\rho(X,Y)| = 1 \Leftrightarrow \exists a,b \in \Re$ mit $P(Y = aX + b) = 1$
4. $|\rho(X,Y)| \nearrow 1 \Rightarrow$ die Punkte liegen näher an der Regressionsgeraden
5. Bei der Regression handelt es sich um eine lineare Vorhersage, aber ein kausaler Zusammenhang liegt i.a. nicht vor.

Was sind die "Partner" von $cov(X,Y)$ und $\rho(X,Y)$ in der Beobachtungsebene?

Beobachtungsebene	Modellebene
$h_n(A)$	$P(A)$
$\bar{x}$	EX
$\tilde{x}$	Median(X)
s^2	Var(X)
s	$\sigma(X)$
$s_x^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$	$VarX$
$s_y^2 = \frac{1}{n}\sum(y_i - \bar{y})^2$	$VarY$
$s_{xy}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$	$cov(X, Y)$
$r = \frac{s_{xy}^2}{s_x s_y}$	$\rho(X, Y)$

Regressionsgerade:

$$
\begin{aligned}
g(x) &= \bar{y} + \frac{s_{xy}^2}{s_x s_y}(x - \bar{x}) \\
&= \bar{y} + \frac{r s_x s_y}{s_x^2}(x - \bar{x}) \\
&= \bar{y} + r\frac{s_y}{s_x}(x - \bar{x})
\end{aligned}
$$

$$f(a, b) = E(Y - (a + bX))^2 \to \min_{a,b}$$

$$f(a_{min}, b_{min}) = VarY\left(1 - (\rho(X, Y))^2\right)$$

Wir stellen damit nur ein lineares Verhältnis dar, d.h. nicht X beeinflusst Y.

Definition 1.19: unkorrellierte ZG

Die ZG X und Y mit $\rho(X, Y) = 0$ heißen unkorrelliert.

Satz 1.21:

Unabhängige ZG X und Y sind unkorrelliert.

Die Umkehrung gilt im allgemeinen nicht.

1.10 Tschebyschewsche Ungleichung und Gesetz der grossen Zahlen

Satz 1.22: Tschebyschewsche Ungleichung (1867)
Sei X eine ZG mit $EX^2 < \infty$ und $\epsilon > 0$ eine beliebige reelle Zahl.
Dann gilt:

$$P(|X - EX| \geq \epsilon) \leq \tfrac{VarX}{\epsilon^2}.$$

Folgerungen:

$$\begin{aligned}
\epsilon &:= k\sqrt{VarX} \\
P(|X - EX| \geq k\sqrt{VarX}) &\leq \frac{VarX}{k^2 VarX} = \frac{1}{k^2} \\
P(|X - EX| < k\sqrt{VarX}) &\geq 1 - \frac{1}{k^2} \\
k &= 1: \\
P(|X - EX| < k\sqrt{VarX}) &\geq 0 \\
k &= 2: \\
P(|X - EX| < k\sqrt{VarX}) &\geq 0,75 \\
k &= 3: \\
P(|X - EX| < k\sqrt{VarX}) &\geq 0,89
\end{aligned}$$

Beispiele:

1. $X \sim B(10; 0,25)$
   $EX = 2,5 \qquad \sqrt{VarX} = \sigma \approx 1,37$

   $k = 1 \quad P(1,13 < X < 3,87) = P(2 \leq X \leq 3) = 0,53 > 0$ (diskrete ZG)
   $k = 2 \quad P(0 \leq X \leq 5) = 0,98 > 0,75$
   $k = 3 \quad P(0 \leq X \leq 6) = 0,997 > 0,89$

   Die Tschebyschewsche Ungleichung unterschätzt hier die tatsächlichen Wahrscheinlichkeiten. Die Abweichungen resultieren aus der Allgemeinheit der Ungleichung.

2. Ist die Tschebyschewsche Ungleichung scharf?

X: $\begin{array}{ccc} -2 & 0 & 2 \\ \hline \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \end{array}$

$EX = 0$

$VarX = EX^2 = 1$

$k = 2 \qquad P(-2 < X < 2) = P(X = 0) = 0,75$

Die Ungleichung ist nicht zu verbessern, d.h. sie ist eine scharfe Ungleichung.

Wir betrachten nun:

$X_1, X_2, \ldots$ unabhängige ZG'en.

Dann ist auch: $\frac{X_1+X_2+\ldots+X_n}{n}$ eine ZG.

Für: $\omega \in \Omega$ ist:

$\frac{X_1(\omega)+X_2(\omega)+\ldots+X_n(\omega)}{n}$ eine Zahl, die die Realisierung der ZG angibt.

Satz 1.23: Das schwache Gesetz der großen Zahlen

Es seinen $X_1, X_2, \ldots$ unabhängige ZG mit gleichem EW EX_1 und $VarX_i \leq M < \infty$.

Dann gilt $\forall \epsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+\ldots+X_n}{n} - EX_1\right| \geq \epsilon\right) = 0$$

Durch das schwache Gesetz der großen Zahlen wird in der Modellebene der Sachverhalt des Stabilwerdens in der Beobachtungsebene wiedergespiegelt.

Satz 1.24: Das starke Gesetz der großen Zahlen

Es seinen $X_1, X_2, \ldots$ unabhängige ZG mit $EX_i = EX$ und $VarX_i = \sigma^2 < \infty$.

Dann gilt:

$$P(\{\omega \in \Omega : \lim_{n\to\infty} \frac{X_1+\ldots+X_n}{n} = EX_1\}) = 1$$

Folie 7: Gesetz der großen Zahlen

Beobachtungsebene	Modellebene
$h_n(A)$	$P(A)$
$\bar{x}$	EX
$\tilde{x}$	Median(X)
s^2	Var(X)
s	$\sigma(X)$
$s_x^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$	$VarX$
$s_y^2 = \frac{1}{n}\sum(y_i - \bar{y})^2$	$VarY$
$s_{xy}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$	$cov(X,Y)$
$r = \frac{s_{xy}^2}{s_x s_y}$	$\rho(X,Y)$
Stabilwerden des arithmetischen Mittels	Gesetz der großen Zahlen
Stabilwerden der relativen Häufigkeit	Berloullische Gesetz der großen Zahlen

Folgerung: Bernoullisches Gesetz der großen Zahlen

Seien $X_1, X_2 \ldots$ unabhängige ZG mit:

$P(X_i = 0) = 1 - p$

$P(X_i = 1) = p$

Dann gilt:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - np\right| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{1}{4n\epsilon^2} = 0.$$

Folie 8: Jakob Bernoulli (1654-1705) zum Gesetz der großen Zahlen

Die relative Häufigkeit als ZG:

$$\begin{aligned}
\frac{S_n}{n} &:= \frac{X_1 + X_2 + \ldots + X_n}{n} \\
E\left(\frac{S_n}{n}\right) &= p \\
Var\left(\frac{S_n}{n}\right) &= \frac{1}{n}p\,(1-p) \\
P\left(\frac{S_n}{n} = \frac{k}{n}\right) &= P(S_n = k) = \binom{n}{k} p^k\,(1-p)^{n-k}
\end{aligned}$$

Folie 9: Verteilung der relativen Häufigkeiten bei $p = 0,3$ in Abhängigkeit von n

1.11 Das Schaetzen einer unbekannten Wahrscheinlichkeit

Beispiel:

Der Anteil der A-Wähler in einer großen Stadt soll geschätzt werden.

Fragen:

Wie organisiert man die Stichprobe?

Wie groß muss n sein?

A: ein zufällig ausgewählter Wähler ist A-Wähler

$P(A) = p$: Anteil der A-Wähler

$X_1, X_2, \ldots, X_n$: mathematische Stichprobe

Modellannahmen:

- X_i sind unabhängig
- $P(X_i = 0) = 1 - p$
- $P(X_i = 1) = p$
- $X_1 + X_2 + \ldots + X_n := S_n$
- Schätzwert für p
  $\hat{p}(X_1, \ldots, X_n) = \frac{S_n}{n}$
  S_n ist eine ZG, sie gibt die relative Häufigkeit der Erfolge in der Stichprobe n an.
  $S_n \sim B(p, n)$

konkrete Stichprobe:

$n = 2000$

beobachtet 400 A-Wähler

$\hat{p}(x_1, \ldots, x_n) = \frac{400}{2000} = 0,2$

Wie genau ist dieser Schätzwert?

Güte der Schätzung:

$$P\left(\left|\frac{S_n}{n}-p\right|<\epsilon\right)\geq 1-\alpha$$

Dabei ist:

p- der wahre Wert

ϵ- die Genauigkeit

$1-\alpha$- die Sicherheit der Genauigkeit

Bernoullische Gesetz der großen Zahlen:

$$P\left(\left|\frac{S_n}{n}-p\right|\geq\epsilon\right)\leq\frac{1}{4n\epsilon^2}$$

für die Güte soll gelten:

$$P\left(\left|\frac{S_n}{n}-p\right|\geq\epsilon\right)\leq\alpha$$

es reicht zu fordern:

$$\frac{1}{4n\epsilon^2}\leq\alpha$$

$\Leftrightarrow$

$$n\geq\frac{1}{4\epsilon^2\alpha}$$

Satz 1.25: Stichprobenumfang

Damit die Schätzung $\hat{p}(X_1,\ldots,X_n)=\frac{S_n}{n}$ die Güte

$$P\left(\left|\frac{S_n}{n}-p\right|<\epsilon\right)\geq 1-\alpha \qquad \forall\alpha\in(0,1)$$

besitzt, reicht es, eine Stichprobe vom Umfang n mit

$n\geq\frac{1}{4\epsilon^2\alpha}$ zu erheben.

Beispiel:

$\epsilon=0,05$ $\qquad 1-\alpha=0,95\Rightarrow n\geq 2000$

Ein Stichprobenumfang $n\geq 2000$ garantiert mit 95% eine Abweichung der geschätzten Wahrscheinlichkeit vom wahren Wert von weniger als 0,05.

Bemerkung:

Es ist nicht das beste (kleinste) n wegen der Tschebyschewschen Ungleichung (im Bsp. $n = 2000$ und $\alpha = 0,05 \Rightarrow \epsilon \approx 0,02$).

Konfidenz-/Vertrauensintervalle für p

Start:$\forall p$

$$\begin{aligned}
P\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) &\geq 1-\alpha \\
P\left(-\epsilon < p - \frac{S_n}{n} < \epsilon\right) &\geq 1-\alpha \\
P\left(\underbrace{\frac{S_n}{n} - \epsilon}_{U(X_1,\ldots,X_n)} < p < \underbrace{\frac{S_n}{n} + \epsilon}_{O(X_1,\ldots,X_n)}\right) &\geq 1-\alpha \\
P\left(p \in (U(X_1,\ldots,X_n), O(X_1,\ldots,X_n)\right) &\geq 1-\alpha
\end{aligned}$$

D.h. per Konstruktion überdeckt das zufällige Intervall mit einer Wahrscheilichkeit $\geq 1-\alpha$ den wahren Wert p.
$U(X_1,\ldots,X_n)$ - untere Schranke
$O(X_1,\ldots,X_n)$ - obere Schranke

Beispiel:
$\hat{p}(x_1, x_2, \ldots, x_n) = 0,2$
$\epsilon = 0,05$
konkretes Konfidenzintervall:
$(0,2 - 0,05; 0,2 + 0,05) = (0,15; 0,25)$

Bemerkung:

1. Die Aussage: Das Intervall $(0,15; 0,25)$ enthält p mit einer Wahrscheinlichkeit von $0,95$ ist sinnlos.

2. Das Konfidenzintervall ist nur eine andere Formulierung der Güte.

1.12 Testen von Hypothesen ueber eine unbekannte Wahrscheinlichkeit p

Was ist eine Hypothese?

Hypothese H: Meinung über p als Modellparameter

Alternative A: konkurrierende Meinung über p als Modellparameter

Didaktischer Fall:

- einfache Hypothese gegen einfache Alternative

H: $p = 0,5$

A: $p = 0,75$

- Urne mit 50 weißen und 50 schwarzen Kugeln oder mit 75 weißen und 25 schwarzen Kugeln
- wir nehmen eine Stichprobe $n = 20$
- Ziehen mit Zurücklegen
- beobachte die ZG S_{20} - Anzahl der weißen Kugeln

unter H: (d.h. H gilt als Modell) erwarten wir $10 \pm a$ Erfolge

plausibel bei Alternative A:

lehne H ab, wenn $S_{20} \geq c$

Kritischer Bereich: $K = \{S_{20} \geq c\}$

Hintergrund: viele Erfolge sprechen gegen H und für A.

Fehlermöglicheiten:

	H "wahr"	H "falsch"
$x \in K$	Fehler 1.Art	-
$x \notin K$	-	Fehler 2.Art

Entscheidungsregel zu K:

beobachteter Wert in K $\Rightarrow$ lehne H ab

beobachteter Wert nicht in K $\Rightarrow$ behalte H bei

Definition 1.20: Signifikanztest zum Signifikanzniveau α

Ein Signifikanztest zum Signifikanzniveau α ist eine Entscheidungsregel mit dem kritischen Bereich K, für den gilt:

$$P_{(H)}(K) \leq \alpha \qquad \forall \text{ Parameter aus H.}$$

D.h. die Wahrscheinlichkeit von K in jedem durch H gegebenen Modell ist kleiner oder gleich α.

Es besteht eine Asymetrie in der Behandlung von H und A.

Für H wird das konservative Modell gewählt, gegen das überzeugende Argumente (Stichprobenergebnisse) vorgebracht werden sollen.

Standardwerte für α: 0,01; 0,05; 0,1

Wenn ein Beobachtungswert x signifikant ist (d.h. $x \in K$) bei $\alpha = 0,05$ dann ist er auch signifikant für ein größeres α, aber nicht unbedingt für ein kleineres α.

aus Bsp.:

$$P_{(H)}(S_{20} \geq 14) = \binom{20}{14} p^{20} + \cdots + \binom{20}{20} p^{20} = 0,06$$

im Bsp.:

$$\alpha = 0,05 \Rightarrow c \geq 15$$

mit Blick auf den Fehler 2. Art wähle $c = 15$

Entscheidungsregel:

$K = \{S_{20} \geq 15\}$

$x \in K \Rightarrow$ lehne H ab

$x \notin K \Rightarrow$ behalte H bei

Philosophie des Signifikanztests:

$x \in K$ bedeutet:

Unter der Hypothese H ist ein extrem vom erwarteten Wert abweichendes Ergebnis eingetreten. Derartige Abweichungen haben per Testkonstruktion unter H eine sehr kleine Wahrscheinlichkeit. Solche treten im Einzelfall nicht auf (so die Überzeugung), daher lehnen wir H ab. Die Alternative bietet eine bessere Erklärung für das beobachtete Ergebnis, unter der Prämisse, dass der Fehler 1. Art durch α begrenzt ist.

Einfache Hypothese gegen zusammengesetze Alternative:

Bsp.:

H: $p = 0,5$

A: $p > 0,5$

konkret: Unter 100 Neugeboren waren 60 Jungen, ist das ein Grund H abzulehnen?

$K = \{S_{100} \geq c\}$

Wie wählt man c?

> **Definition 1.21: beobachtetes Signifikanzniveau**
>
> Das beobachtete Signifikanzniveau ist
>
> $$P_{(H)}(S_n \geq x),$$
>
> wenn x der beobachtete Wert von S_n ist.
>
> Wie ist die Wahrscheinlichkeit des beobachteten oder eines noch extremeren Wertes (bzgl. H).

Bsp.:

$P_{(H)}(S_n \geq 60) = 0,03$

Bei einem Sibnifikanzniveau $\alpha = 0,03$ wären 60 Jungen eine signifikante Abweichung.

$\Rightarrow \alpha \geq 0,03$

Bei vorgegebenem Signifikanzniveau $\alpha = 0,05$:

$\Rightarrow$ Ablehnungsbereich $K = \{S_{100} \geq 59\}$

Fehler 2. Art:

bei $p = 0,75$

$P_{(0,75)}(S_{100} \leq 58) = 0,001$

Folie 10: Gütefunktion

Definition 1.22: Gütefunktion

Die Gütefunktion eines Tests mit dem kritischen Bereich K ist die Funktion β mit:

$$\beta(p) = P_{(p)}(S_n \in K),\ p \in [0,1].$$

D.h. $\beta(p)$ ist die Wahrscheinlichkeit die Hypothese abzulehnen, wenn p der Modellparameter ist.

Sei $p \in A$ dann ist $1 - \beta(p)$ die Wahrscheinlichkeit des Fehlers 2. Art.

Die Gütefunktion bleibt gleich, auch wenn wir aus der einfachen Hypothese H eine zusammengesetze machen $H :\leq 0,5$.

Chapter 2

Allgemeine (re) Wahrscheinlichkeitsraume

2.1 Sigma- Algebren und allgemeine Wahrscheinlichkeitsverteilungen

Beispiele:

- Kugel anstatt Würfel
- Lebensdauer (von Glühlampen)
  Wunsch: $\Omega = [0, \infty)$
- Auswahl auf gut Glück eines Punktes aus $[0, 1]$
  naheliegend: $P((0,2; 0,4]) = \frac{1}{5}$
  allgemein: $P((a, b]) = b - a$

Betrachtet man alle Teilmengen von $[0, 1]$: also $2^{[0,1]}$, gibt es keine eindeutige Fortsetzung von P mit:
$P((a, b]) = b - a$ auf die Menge aller Teilmengen von $[0, 1]$.
Lösung: wir müssen uns bei den Teilmengen einschränken.

Definition 2.1: σ-Algebra

Sei $\Omega \neq \oslash$.

Eine Familie $\mathcal{A}$ von Teilmengen von Ω heißt σ-Algebra, falls:

1. $\Omega \in \mathcal{A}$
2. wenn $A \in \mathcal{A}$, dann auch $\bar{A} \in \mathcal{A}$
3. wenn $A_1, A_2, \ldots \in \mathcal{A}$, dann auch $\bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{A}$.

Bsp.:

$\Omega = \Re$

$\mathcal{M}= \{(a,b] : a < b; a, b \in \Re\}$ ist keine σ-Algebra

Satz 2.1: Eigenschaften σ-Algebra

Ist $\mathcal{A}$ eine σ-Algebra von Teilmengen von Ω, so gilt:

1. $\oslash \in \Omega$
2. $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$ und $A \cap B \in \mathcal{A}$
3. $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcap\limits_{k=1}^{\infty} A_k \in \mathcal{A}$.

Folgerung:

$A, B \in \mathcal{A} \Rightarrow A \backslash B \in \mathcal{A}$

Bsp.:

$\Omega = \Re$

$\mathcal{M}= \{(a,b] : a < b; a, b \in \Re\}$

$\mathcal{M}$ ist keine σ-Algebra

Sei $\mathcal{B}$ die kleinste σ-Algebra die $\mathcal{M}$ enthält.

- Die Menge aller Teilmengen von $\Re$ ist eine σ-Algebra
- Seien $\mathcal{A}_1, \mathcal{A}_2$ σ-Algebren $\rightarrow \mathcal{A}_1 \cap \mathcal{A}_2 = \{A_1 \cap A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$ ist auch eine σ-Algebra.
- $\bigcap\limits_{\mathcal{M} \subset \mathcal{A}} \mathcal{A} = \mathcal{B}$ $\qquad$ $\mathcal{A}$ - σ-Algebra; $\mathcal{B}$-Borelsche σ-Algebra

$\Omega = [0,1] \qquad\qquad \mathcal{B}_{[0,1]} = \mathcal{B} \cap [0,1]$

$B \in \mathcal{B}$ heißt Borelmenge.

Definition 2.2:

Sei $\Omega \neq \oslash$ und $\mathcal{A}$ eine σ-Algebra von Teilmengen von Ω.

Dann heißt $(\Omega, \mathcal{A})$ ein messbarer Raum.

Eine Wahrscheinlichkeitsverteilung P auf $\mathcal{A}$ ist eine Funktion $P : \mathcal{A} \rightarrow \Re$ mit den folgenden Eigenschaften:

1. $P(A) \geq 0 \qquad \forall A \in \mathcal{A}$ (Nichtnegativität)
2. $P(\Omega) = 1$ (Normiertheit)
3. Für jede Folge $A_1, A_2, \ldots$ paarweise disjunkte Mengen aus $\mathcal{A}$ gilt:
   $P\left(\dot{\bigcup\limits_{k=1}^{\infty}} A_k\right) = \sum\limits_{k=1}^{\infty} P(A_k)$ (σ-Additivität).

Das Tripel $(\Omega, \mathcal{A}, P)$ heißt Wahrscheinlichkeitsraum.

Die Mengen $A \in \mathcal{A}$ heißen Ereignisse.

Künftig meist: $(\Re, \mathcal{B}, P)$ oder $([a,b], \mathcal{B}_{[a,b]}, P)$

Rückblick: Ω diskret

$\omega \mapsto P(\omega) \Rightarrow P(A) = \sum\limits_{\omega \in A} P(\omega) \qquad\qquad (*)$

Dieses gemäß $(*)$ definierte P hat die Eigenschaften 1.-3.

Wir haben daraus alle weiteren Eigenschaften abgeleitet.

Damit übertragen sich auch alle weiteren Eigenschaften einer Wahrscheinlichkeitsverteilung P auch in den allgemeinen WRaum.

Unterschied:

$\mathcal{A}$ ist keine beliebige Teilmenge von Ω, sondern von $\mathcal{A} \in \Omega$.

Bsp.:

$P\left(\bigcup\limits_{k=1}^{\infty} A_k\right)$ hat nur Sinn, wenn $\bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{A}$

ebenso $P(\bar{A})$ mit $\bar{A} \in \mathcal{A}$

meist $(\Re, \mathcal{B}, P)$ bzw. Teilmenge von $\Re$

Definition 2.3: Verteilungsfunktion

Eine Funktion $F : \Re \to \Re$ heißt Verteilungsfunktion (VF) einer Wahrscheinlichkeitsverteilung P auf $(\Re, \mathcal{B})$ wenn $\forall x \in \Re$ gilt:

$$F(x) = P((-\infty, x])$$

$$(-\infty, x] = \bigcup_{n}^{\infty}(-n, x] \text{ (Borelmenge)}$$

Satz 2.2: Eigenschaften der VF

Wenn F die VF einer Verteilung P ist, dann gilt:

i) F ist monoton wachsend

ii) F ist rechtsseitig stetig

iii) $\lim\limits_{x \to -\infty} F(x) = 0$ und $\lim\limits_{x \to +\infty} F(x) = 1$

Satz 2.3:

Sei F eine Funktion mit den Eigenshaften i)-iii) aus Satz 2.3) einer VF. Dann ist durch:

P((a,b])=F(b)-F(a), $\quad a < b, a, b \in \Re$

eindeutig eine Wahrscheinlichkeitsverteilung auf $(\Re, \mathcal{B})$ festgelegt.

2.2 Wahrscheinlichkeitsverteilungen mit Dichten

Definition 2.4: Dichte

Eine Dichte f ist eine Funktion $f : \Re \to \Re$ mit:

i) $f(x) \geq 0 \qquad \forall x \in \Re$

ii) $\int\limits_{-\infty}^{\infty} f(x)dx = 1,$

wobei das Integral als Riemann Integral verstanden wird.

Satz 2.4:

Sei f eine Dichte auf $\Re$. Dann ist durch:

$$F(x) = \int\limits_{-\infty}^{x} f(u)du$$

eine VF definiert. Diese legt durch:

$$P((a,b]) = F(b) - F(a)$$

eindeutig eine Wahrscheinlichkeitsverteilung auf $(\Re, \mathcal{B})$ fest.

Bemerkung:
$P((a,b]) = \int\limits_{a}^{b} f(u)du$
Die Dichte f ist stetig bzw. hat höchsten endlich viele Stetigkeitsstellen.

Satz 2.5:

Sei P eine Wahrscheinlichkeitsverteilung auf $(\Re, \mathcal{B})$ mit Dichte f und $c \in \Re$.

Dann gilt:

$$P(\{c\}) = 0.$$

Lemma:

Sei $(\Omega, \mathcal{A}$,P) ein Wahrscheinlichkeitsraum.

1. Für jede aufsteigende Folge von Ereignissen $A_i \in \mathcal{A}$,

   d.h. $A_1 \subset A_2 \subset \ldots$, gilt: $P(\bigcup A_i) = \lim P(A_i)$.

2. Für jede absteigende Folge von Ereignissen $A_i \in \mathcal{A}$,

   d.h. $A_1 \supset A_2 \supset \ldots$, gilt: $P(\bigcap A_i) = \lim P(A_i)$.

Standardmodelle mit Dichten

1. Gleichverteilung auf $[a, b]$

$$f(x) = \begin{cases} \frac{1}{b-a} & ; x \in [a,b] \\ 0 & ; sonst \end{cases}$$

   Merkmale:

   - Dichte ist nicht negativ

- Fläche unter dem Graphen $= 1$, da $(b-a)\frac{1}{b-a}$
- Intervalle gleicher Länge in $[a, b]$ sind gleichwahrscheinlich

Vorgang:
zufällige Auswahl eines Punktes aus dem Intervall $[a, b]$.
$(c, d) \subset [a, b]$

$$P(\omega \in (c,d)) = \int\limits_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}$$

2. Exponentialverteilung mit Parameter $\lambda > 0$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; x \geq 0 \\ 0 & ; x < 0 \end{cases}$$

Verteilungsfunktion:

$$\begin{array}{lrcl} & F(x) & = & P((-\infty, x]) = \int\limits_{-\infty}^{x} f(u)du \\ x \leq 0 & F(x) & = & 0 \\ x > 0 & F(x) & = & \int\limits_0^x f(u)du \\ & & = & \int\limits_0^x \lambda e^{-\lambda u} du \\ & & = & \left[-e^{-\lambda u}\right]_0^x \\ & & = & 1 - e^{-\lambda x} \end{array}$$

Charakteristisch für die Exponentialverteilung (unter den stetigen Verteilungen) ist die Vergessenseigenschaft.
Es gilt:

$$P((s+t, \infty)|(t, \infty)) = P((s, \infty))$$

Die Wahrscheinlichkeit, dass das Ereignis nach s+t eintritt, wenn bekannt ist das Ereignis bis t nicht eingetreten ist.
Beweis:

$$P((s+t, \infty)|(t, \infty)) = \frac{P((s+t, \infty))}{P((t, \infty))}$$

$$\begin{aligned} &= \frac{1-F(s+t)}{1-F(t)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} \\ &= 1-F(s) \\ &= P((s,\infty)) \end{aligned}$$

$1-F(x) = G(x)$
$\frac{G(s+t)}{G(t)} = G(s)$

$$\left\{ \begin{array}{ll} G(s+t) = G(s)G(t) & ;\forall s,t \geq 0 \\ G(0) = 1 & \\ Gstetig & \end{array} \right\} \Rightarrow G = a^x$$

Anwendungen:

- Lebensdauer bei nicht alternden Objekten (Glühlampen)

- Wartezeit auf Eintreffen eines Ereignisses (die sich unabhängig voneinander mit konstanter Intensität ereignen)

3. Normalverteilung mit den Parametern μ und σ^2, $\mu \in \Re$, $\sigma > 0$

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Symbol: $N(\mu, \sigma^2)$

Merkmale:

- die Funktion hat ihr Maximum bei $\mu = 0$

- je größer σ ist, desto flacher ist die Kurve

- μ bestimmt die Lage der Kurve

- zwei Wendepunkte bei $\mu \pm \sigma$

Anwendungen:

- Messgrößen

- näherungsweise Verteilung von Summen unabhängiger ZG

- Rendite von Aktien als grobes Modell

Standardnormalverteilung/ Gaußsche Glockenkurve

Spezialfall der Normalverteilung mit: $\mu = 0$ und $\sigma = 1$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \qquad x \in \Re$$

Verteilungsfunktion:

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

Symmetrieeigenschaft:

$$\Phi(x) = 1 - \Phi(-x)$$

Beispiel:

$a < 0$

$$\begin{aligned} P((a,b]) &= \Phi(b) - \Phi(a) \\ &= \Phi(b) - (1 - \Phi(-a)) \end{aligned}$$

weitere Eigenschaften:

$\Phi(0) = \frac{1}{2}$

$\Phi(\infty) = 1$

$\Phi(-\infty) = 0$

Zurückführen von $N(\mu, \sigma^2)$ auf $N(0,1)$

$$\begin{aligned} P_{\mu,\sigma^2}((a,b)) &= \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx \\ Substitution &: z = \frac{x-\mu}{\sigma} \qquad \sigma dz = dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{1}{2}z^2} \sigma dz \\ &= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \phi(z) dz \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

Beispiel:

$N(3,4) \qquad\qquad (a,b) = (2,6)$

$$\begin{aligned}
P_{3,4}((2,6)) &= \Phi\left(\frac{6-3}{2}\right) - \Phi\left(\frac{2-3}{2}\right) \\
&= \Phi(1,5) - \Phi(-0,5) \\
&= \Phi(1,5) - (1 - \Phi(0,5)) \\
&= 0,9332 + 0,6915 - 1 \approx 0,625 \\
P_{3,4}((-\infty,6)) &= \Phi(1,5) - 0 \\
P_{3,4}((2,\infty)) &= 1 - \Phi(-0,5) = \Phi(0,5)
\end{aligned}$$

$k\sigma$- Intervalle der Normalverteilung:

$$\begin{aligned}
P((\mu - k\sigma, \mu + k\sigma)) &= \Phi\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) \\
&= \Phi(k) - \Phi(-k) \\
&= 2\Phi(k) - 1 \\
&= \left\{\begin{array}{ll} 0,6826 & ;k=1 \\ 0,9544 & ;k=2 \\ 0,9974 & ;k=3 \end{array}\right\}
\end{aligned}$$

Die Wahrscheinlichkeiten der $k\sigma$- Intervalle sind fest, die Breite und Lage hängt von σ und μ ab.

Dichten im $\Re^2$

1. Gleichverteilung auf einem Gebiet, z.B. einer Kreisscheibe
   $f(x,y) = \left\{\begin{array}{ll} \frac{1}{\pi} & ;(x,y) \in K(0,1) \\ 0 & ;(x,y) \notin K(0,1) \end{array}\right\}$
   Bsp.:

$$\begin{aligned}
P(K(0,\frac{1}{2})) &= \int\int_{K(0,\frac{1}{2})} \frac{1}{\pi} dxdy \\
&= \frac{\left|K(0,\frac{1}{2})\right|}{\pi} \\
&= \frac{\pi\frac{1}{4}}{\pi} = \frac{1}{4}
\end{aligned}$$

2. Zweidimensionale Normalverteilung

$$f_{\mu,\sigma^2,\rho}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_1} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}}$$

Folie: Dreidimensionale Darstellung

2.3 ZG und ihre Verteilung im allgemeinen Fall

$(\Omega, \mathcal{A}, P)$ - WRaum

$X : \Omega \to \Re$

Gesucht wird z.B.: $P(X < 7) = P\underbrace{(\{\omega \in \Omega : X(\omega) < 7\})}_{? \in \mathcal{A}}$

Das $(? \in \mathcal{A})$ muss gesichert werden für genügend viele Aussagen über Werte von X.

Definition 2.5: ZG

Sei $(\Omega, \mathcal{A}, P)$ ein WRaum.

Eine Funktion $X : \Omega \to \Re$ heißt ZG, falls für jede Borelmenge $B \in \mathcal{B}$ gilt:

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A},$$

d.h. ist ein Ereignis im WRaum $(\Omega, \mathcal{A}, P)$

$\{\omega \in \Omega : X(\omega) \in B\} = X^{-1}(B)$

Urbild:

$f : X \to Y$

$f^{-1}(B) \qquad B \subset Y$

$f^{-1}(B) = \{x \in X : f(x) \in B\}$

Definition 2.6: Verteilung der ZG X

Sei X eine ZG auf Ω. Die durch

$$P_X(B) = P(X^{-1}(B)) \text{ ,} B \in \mathcal{B}$$

definierte Wahrscheinlichkeitsverteilung heißt Verteilung der ZG X.

Bemerkung:

1. X "transponiert" die Verteilung P nach $(\Re, \mathcal{B})$. Ergebnis ist P_X.

$$\begin{aligned} P_X(B) &= P(X \in B) = P(\{\omega \in \Omega : X(\omega) \in B\}) \\ P_X((a,b)) &= P(a < X < b) \\ P_X((-\infty, c]) &= P(X \leq c) \end{aligned}$$

2. Ist P_X wirklich eine Wahrscheinlichkeitsverteilung?

   (a) $P_X(B) \geq 0$

   (b) $P_X(\Re)$ $\qquad$ $\Re$ sicheres Ereignis bezüglich P_X

$$\begin{aligned} P_X(\Re) &= P(X^{-1}(\Re)) \\ &= P(\{\omega \in \Omega : X(\omega) \in \Re\}) \\ &= P(\Omega) = 1 \end{aligned}$$

   (c)

$$\begin{aligned} P_X(\dot{\cup} B_k) &= P(X^{-1}(\dot{\cup} B_k)) \\ &= P(\dot{\cup} X^{-1}(B_k)) \\ &= \sum P(X^{-1}(B_k)) \\ &= \sum P_X(B_k) \end{aligned}$$

3. Schreibweise: $X \sim P_X$

   z.B. $X \sim N(\mu, \sigma^2)$

   Hat P_X eine Dichte f, so schreibt man auch $X \sim f$.

4. Oft taucht $(\Omega, \mathcal{A}, P)$ nicht mehr auf. Es heißt einfach $X \sim P_X$.

5. Betrachte $(\Omega, \mathcal{A}, P) = (\Re, \mathcal{B}, P)$

   $X : \Re \to \Re$

   Alle stetigen / monotonen Funktionen und Funktionen mit abzählbar vielen Unstetigkeitsstellen sind ZG.

Standardisieren einer normalverteilten ZG

Satz 2.6:

1) Sei $X \sim N(\mu, \sigma^2)$, dann gilt: $X^* = \frac{X-\mu}{\sigma} \sim N(0,1)$. X^* heißt die standardisierte ZG zu X.

2) Sei $Y \sim N(0,1)$. Dann ist: $X = \sigma Y + \mu \sim N(\mu, \sigma^2)$.

Fazit:

Affine Transformationen von Normalverteilungen sind wieder Normalverteilungen.

Satz 2.7:

Sei X gleichverteilt auf $(0,1)$ und $\lambda > 0$. Dann ist die ZG:

$$Y = -\frac{1}{\lambda} \ln(1-X)$$

exponentialverteilt mit Parameter λ.

Unabhängigkeit von ZG

Definition 2.7:

Seien $X_1, X_2, \ldots, X_n$ ZG auf einem WRaum $(\Omega, \mathcal{A}, P)$.

Dann heißen $X_1, X_2, \ldots, X_n$ unabhängig, falls für beliebige $B_1, B_2, \ldots, B_n \in \mathcal{B}$ die Ereignisse $\{X_1 \in B_1\}, \{X_2 \in B_2\}, \ldots, \{X_n \in B_n\}$ unabhängig sind.

D.h. $P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = P(X_1 \in B_1) \cdot P(X_2 \in B_2) \cdot \ldots \cdot P(X_n \in B_n)$

Wir betrachten ZG mit Dichten.

$X = (X_1, X_2, \ldots, X_n)$ ist ein zufälliger Vektor.

$X \sim f(x_1, \ldots, x_n)$

Satz 2.8:

Seien X_i ZG mit den Dichten f_i. $X_1, X_2, \ldots, X_n$ sind unabhängig $\Leftrightarrow$

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdot \ldots \cdot f_n(x_n).$$

Faltungen

Problem:

$X_1 \sim f_1$, $X_2 \sim f_2$ unabhängig.

$S = X_1 + X_1$

Was ist die Verteilung von der Summe?

$$\begin{aligned}
P(S \leq s) &= P(X_1 + X_2 \leq s) \\
&= \int\limits_{\{(x_1,x_2)\in\Re^2 : x_1+x_2 \leq s\}} \int f_1(x_1) f_2(x_2) dx_2 dx_1 \\
&: \quad u := x_1 + x_2 \qquad v := x_2 \\
&= \int\limits_{-\infty}^{s} \int\limits_{-\infty}^{\infty} f_1(u-v) f_2(v) dv du \\
&= \int\limits_{-\infty}^{s} \underbrace{\left[\int\limits_{-\infty}^{\infty} f_1(u-v) f_2(v) dv \right]}_{f_1 * f_2(u)} du \\
F(s) &= \int\limits_{-\infty}^{s} f_1 * f_2(u) du
\end{aligned}$$

D.h. $f_1 * f_2$ ist die Dichte der Summe S.

Definition 2.8:

$f(u) = \int\limits_{-\infty}^{\infty} f_1(u-v) f_2(v) dv$ heißt Faltung der Dichten f_1 und f_2.

Es ist die Dichte der Summe unabhängiger ZG mit den Dichten f_1 und f_2.

Analogie zum diskreten Fall:

$X_1 \sim B(n,p)$, $X_2 \sim B(m,p)$

$S = X_1 + X_2 \sim B(n+m, p)$

$$\begin{aligned}
\underbrace{P(S=s)}_{\hat{=} f(s)} &= \sum_{v=0}^{s} P(X_1 = s-v, X_2 = v) \\
&= \underbrace{\sum_{v=0}^{s} P(X_1 = s-v) P(X_2 = v)}_{\hat{=} \int\limits_{-\infty}^{\infty} f_1(s-v) f_2(v) dv}
\end{aligned}$$

Beispiel:

1. X_1, X_2 gleichverteilt auf $[0,1]$, unabhängig, $S = X_1 + X_2$

$$\begin{aligned}
f(u) &= \int\limits_{-\infty}^{\infty} f(u-v) f(v) dv \\
&= \int\limits_{0}^{1} f(u-v) 1 dv \\
&= \left\{ \begin{array}{ll} 0 & ; u < 0, u \geq 2 \\ \int\limits_0^u 1 dv = u & ; 0 \leq u \leq 1 \\ \int\limits_{u-1}^{1} 1 dv = 2-u & ; 1 < u \leq 2 \end{array} \right\}
\end{aligned}$$

   für 12 gleichverteilte ZG gilt:

   - die Ecken glätten sich

   - bei 12 ZG ähnelt die Verteilung schon der Normalverteilung

   $S = X_1 + X_2 + \ldots + X_{12} \approx N(0, 6)$

2. Summe zweier unabhängiger normalverteilter ZG

   $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$

   $\Rightarrow X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

2.4 Kenngroessen einer ZG mit Dichte

Definition 2.9: Erwartungswert

Sei X eine ZG mit Dichte f. Falls $\int\limits_{-\infty}^{\infty} |x|\, f(x)dx$ existiert, so heißt:

$$EX = \int\limits_{-\infty}^{\infty} xf(x)dx$$

Erwartungswert der ZG X.

Interpretation:

EX ist der stabile Wert des arithmetischen Mittels aus vielen Beobachtungen.

EX ist auch hier ein gewichtetes Mittel.

diskret:

$\sum x_k \cdot \underbrace{p_k}_{Gewicht}$

stetig:

$\int x \underbrace{f(x)dx}_{Gewicht}$

Satz 2.9:

Ist X eine ZG mit Dichte f und g eine stetige Funktion $g : \Re \to \Re$, so existiert $Eg(X)$, falls

$$\int |g(x)|\, f(x)dx < \infty \text{ ist}$$

und es gilt:

$$Eg(X) = \int\limits_{-\infty}^{\infty} g(x)f(x)dx.$$

Interpetation:

Wo ist der Schwerpunkt?

Für den Schwerkunkt gilt:

für beliebige f:

$x_s = \frac{\int\limits_{-\infty}^{\infty} xf(x)dx}{\int\limits_{-\infty}^{\infty} f(x)dx}$,

für Dichte f:

$x_s = EX$.

Alle Eigenschaften des EW, die im diskreten Fall gelten, gelten fort.

Beispiele:

1. X gleichverteilt auf $[a,b]$
   $EX = \int\limits_{-\infty}^{\infty} xf(x)dx = \int\limits_{a}^{b} x\frac{1}{b-a}dx = \ldots = \frac{a+b}{2}$

2. $X \sim N(\mu, \sigma^2)$
   $EX = \int\limits_{-\infty}^{\infty} x \underbrace{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}_{:=f_{\mu,\sigma^2}(x)} dx$

   1.Fall $\mu = 0$

$$\begin{aligned} EX &= \int\limits_{-\infty}^{\infty} xf_{\mu,\sigma^2}(x)dx \\ &= \int\limits_{-\infty}^{0} xf_{\mu,\sigma^2}(x)dx + \int\limits_{0}^{\infty} xf_{\mu,\sigma^2}(x)dx \\ u := -x & \\ &= \int\limits_{\infty}^{0} -uf_{\mu,\sigma^2}(-u)(-du) + \int\limits_{0}^{\infty} xf_{\mu,\sigma^2}(x)dx \\ &= -\int\limits_{0}^{\infty} uf_{\mu,\sigma^2}(u)du + \int\limits_{0}^{\infty} xf_{\mu,\sigma^2}(x)dx = 0 \end{aligned}$$

   d.h. $EX = 0$ für $\mu = 0$

   2.Fall: $Y = X - \mu$

   aus 1. folgt: $EY = 0$

   aus den Eigenschaften des Erwartungswertes folgt:

   $EY = EX - \mu$

   $EX = \mu$

Varianz

Die Varianz im Diskreten: $E(X-EX)^2$, bleibt im Stetigen erhalten.

$$VarX = E(X-EX)^2 = \int\limits_{-\infty}^{\infty} (X-EX)^2 f(x)dx$$

Alle Eigenschaften gelten fort.

Beispiele

1. X gleichverteilt auf $[a,b]$
   $VarX = \int\limits_a^b (X - \frac{a+b}{2})^2 \cdot \frac{1}{b-a} dx = \ldots = \frac{(b-a)^2}{12}$

2. $X \sim N(\mu, \sigma^2)$
   $X - \mu \sim N(0, \sigma^2)$
   $VarX = Var(X-\mu)$
   o.B.d.A. $\mu = 0$

$$\begin{aligned} VarX &= \int\limits_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} x \cdot x e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \left[\left| x e^{-\frac{x^2}{2\sigma^2}} (-\sigma^2) \right|_{-\infty}^{\infty} - \int\limits_{-\infty}^{\infty} -\sigma^2 e^{-\frac{x^2}{2\sigma^2}} dx \right] \\ &= \sigma^2 \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \sigma e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \sigma^2 \cdot 1 \end{aligned}$$

Warum mitteln wir beim Messen einer physikalischen Größe?

- häufige Annahme: Messgröße $X \sim (\mu, \sigma^2)$
- Zusatzannahme: $\mu = w$ (wahrer Wert)
- $X_1, \ldots, X_n$ unabhängige Messungen mit $N(w, \sigma^2)$

$\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ arithmetisches Mittel
$\bar{X} \sim N(w, \frac{1}{n}\sigma^2)$

Drei Standardprobleme:

1. geg.: n, ϵ ges.: $1-\alpha$

$$
\begin{aligned}
P(\left|\bar{X}_n - w\right| < \epsilon) &= P\left(\underbrace{\left|\frac{\bar{X}_n - w}{\frac{\sigma}{\sqrt{n}}}\right|}_{\sim N(0,1)} < \frac{\epsilon}{\frac{\sigma}{\sqrt{n}}}\right) \\
&= \Phi\left(\frac{\epsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\epsilon\sqrt{n}}{\sigma}\right) \\
&= 2\Phi\left(\frac{\epsilon\sqrt{n}}{\sigma}\right) - 1 \\
&= 1-\alpha
\end{aligned}
$$

Bsp.: $n = 100$ $\epsilon = 0,1 \cdot \sigma$

$P(\left|\bar{X}_{100} - w\right| < 0,1\sigma) = 2\Phi\left(\frac{10\cdot 0,1\sigma}{\sigma}\right) - 1 = 0,68$

2. geg.: $\epsilon, 1-\alpha$ ges.: n

$$
\begin{aligned}
P(\left|\bar{X}_n - w\right| < \epsilon) &= \geq 1-\alpha \\
\text{aus 1)} & \\
P(\left|\bar{X}_n - w\right| < \epsilon) &= 2\Phi\left(\frac{\epsilon\sqrt{n}}{\sigma}\right) - 1 \\
\text{es soll gelten :} & \\
2\Phi\left(\frac{\epsilon\sqrt{n}}{\sigma}\right) - 1 &\geq 1-\alpha \\
\Phi\left(\frac{\epsilon\sqrt{n}}{\sigma}\right) &\geq 1-\frac{\alpha}{2} \\
\Phi(x_{1-\frac{\alpha}{2}}) &= 1-\frac{\alpha}{2} \\
\frac{\epsilon\sqrt{n}}{\sigma} &\geq x_{1-\frac{\alpha}{2}} \\
\sqrt{n} &\geq \frac{\sigma}{\epsilon} x_{1-\frac{\alpha}{2}} \\
n &\geq \frac{\sigma^2}{\epsilon^2}(x_{1-\frac{\alpha}{2}})^2
\end{aligned}
$$

$$n = \frac{\sigma^2}{\epsilon^2}(x_{1-\frac{\alpha}{2}})^2$$

3. geg.: $n, 1-\alpha$ ges.: ϵ

$$\begin{aligned} 2\Phi\left(\frac{\epsilon\sqrt{n}}{\sigma}\right) - 1 &\geq 1-\alpha \\ \frac{\epsilon\sqrt{n}}{\sigma} &\geq x_{1-\frac{\alpha}{2}} \\ \epsilon &= \frac{\sigma}{\sqrt{n}} x_{1-\frac{\alpha}{2}} \end{aligned}$$

Chapter 3

Grenzwertsaetze

3.1 Grenzwertsatz von de Moivre- Laplace

Praktisches Problem:

$X \sim B(1000, \frac{1}{3})$

ges.:

$P(270 \leq X \leq 350) = \sum_{k=270}^{350} \binom{1000}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{1000-k}$

Wie kann man dafür eine Näherung finden?

$S_n \sim B(n,p)$ $\qquad$ n groß

$S_n = X_1 + X_2 + \ldots + X_n$

$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} = \frac{S_n - np}{\sigma}$

Werte von S_n^*:

$\frac{0-np}{\sigma}, \frac{1-np}{\sigma}, \ldots, \frac{n-np}{\sigma}$

Abstand zwischen zwei benachbarten Werten:

$\frac{k+1-np}{\sigma} - \frac{k-np}{\sigma} = \frac{1}{\sigma} = \underbrace{\frac{1}{\sqrt{np(1-p)}}}_{\stackrel{n\to\infty}{\longrightarrow}0}$

$ES_n^* = 0$

$VarS_n^* = 1$

$p_k = \binom{n}{k} p^k (1-p)^{n-k}$

Satz 3.1: Grenzwertsatz von de Moivre- Laplace

Sei $0 < p < 1$ und $S_n \sim B(n,p)$. Dann gilt für beliebige $a < b$:

$$\lim_{n\to\infty} P(a \leq S_n^* \leq b) = \int\limits_a^b \phi(x)dx = \Phi(b) - \Phi(a).$$

Praktische Anwendung:

$P(S_n = k)$ $\quad\quad$ n groß

Faustregel:

$np(1-p) > 9$

$$\begin{aligned} P(S_n = k) &= P\left(\frac{S_n - np}{\sqrt{np(1-p)}} = \frac{k - np}{\sqrt{np(1-p)}}\right) \\ &\sim \frac{1}{\sqrt{np(1-p)}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{np(1-p)}}\right)^2} \end{aligned}$$

Beispiel:

600 mal Würfeln

S_{600}: Anzahl der Sechsen

$$\begin{aligned} P(S_{600} = 101) &= P\left(S_n^* = \frac{1}{\sqrt{5/6 \cdot 100}}\right) \\ &\sim \frac{1}{\sqrt{5/6 \cdot 100}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{1}{\sqrt{5/6 \cdot 100}}\right)^2} \\ &\approx 0,043 \end{aligned}$$

exakter Wert: 0,043232

Folgerung:

Sei $0 < p < 1$ und $S_n \sim B(n,p)$.

Dann gilt für $a, b \in \Re$:

1. $\lim\limits_{n\to\infty} P(S_n^* \le b) = \int\limits_{-\infty}^{b} \phi(s)ds = \Phi(b)$

2. $\lim\limits_{n\to\infty} P(S_n^* \ge a) = \int\limits_{a}^{\infty} \phi(s)ds = 1 - \Phi(a)$

D.h. im Grenzwertsatz von de Moivre- Laplace darf man $a = -\infty$ und $b = +\infty$ setzen.

Praktische Anwendung:

ges.:

$P(c \le S_n \le d)$ n groß genug

$$\begin{aligned} P(c \le S_n \le d) &= P\left(\frac{c-np}{\sqrt{np(1-p)}} \le S_n^* \le \frac{d-np}{\sqrt{np(1-p)}}\right) \\ &\approx \Phi\left(\frac{d-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{c-np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

Beispiel:

$n = 600$- mal Würfeln, S_n: Anzahl der Sechsen

$$\begin{aligned} P(90 \le S_n \le 110) &= P\left(\frac{90-100}{\sqrt{\frac{500}{6}}} \le S_n^* \le \frac{110-100}{\sqrt{\frac{500}{6}}}\right) \\ &\approx \Phi\left(\frac{10}{\sqrt{\frac{500}{6}}}\right) - \Phi\left(-\frac{10}{\sqrt{\frac{500}{6}}}\right) \\ &= 0,73 \end{aligned}$$

Wie genau ist dieser Wert?

exakter Wert:

$P(90 \le S_n \le 110) = 0,75$

Fehlerabschätzung im Grenzwertsatz:

Satz 3.2: Berry-Esseen

Ist $S_n \sim B(n,p)$ und $0 < p < 1$, so gilt:

$$\sup_{x \in \Re} |P(S_n^* \leq x) - \Phi(x)| \leq \frac{1}{\sqrt{2\pi n}} \cdot \frac{(1-p)^2+p^2}{\sqrt{p(1-p)}} \leq \frac{1}{\sqrt{2\pi np(1-p)}}$$

für $p = 0,5$: $\leq \frac{1}{\sqrt{2\pi n}}$

Stetigkeitskorrektur:

Wir hatten bisher:

$P(a \leq S_n \leq b) \approx \Phi\left(\frac{b-np}{\sigma}\right) - \Phi\left(\frac{a-np}{\sigma}\right)$

mit Stetigkeitskorrektur:

$P(a \leq S_n \leq b) \approx \Phi\left(\frac{b+\frac{1}{2}-np}{\sigma}\right) - \Phi\left(\frac{a-\frac{1}{2}-np}{\sigma}\right)$

Beispiel: (Würfelbeispiel)

$P(90 \leq S_n \leq 110) \approx \Phi\left(\frac{110+\frac{1}{2}-100}{\sigma}\right) - \Phi\left(\frac{90-\frac{1}{2}-100}{\sigma}\right) = 0,75$

Bei einem sehr großen σ macht die Stetigkeitskorrektur keinen Sinn, aber bei kleinem σ.

Beispiel: Die Macht einer resoluten Minderheit

1. Ausschuss aus 5 Personen

   Zwei Personen sind für das Projekt A.

   Die drei restlichen Personen sind unentschlossen.

   A: Projekt A wird beschlossen

   X_3: Anzahl der unentschlossenen Personen die für A sind

   Annahme:

   $X_3 \sim B\left(3, \frac{1}{2}\right)$

   unabhängig, keine Präferenzen

$$\begin{aligned} P(A) &= P(X_3 \geq 1) \\ &= 1 - P(X_3 = 0) \\ &= 1 - 0,5^3 \\ &= 0,875 \end{aligned}$$

2. Abstimmung mit 1.000.000 Personen

   2.000 Personen sind für A.

   Die drei restlichen Personen sind unentschlossen.

   $n = 998.000$

   X_n: Anzahl der unentschlossenen Personen die für A sind

   Annahmen: $X_n \sim B\left(n, \frac{1}{2}\right)$

   $np = 499.000$

   $np(1-p) = 249.500 >> 9$

$$\begin{aligned} P(A) &= P(X_n + 2.000 \geq 500.001) \\ &= P(X_n \geq 498.001) \\ &\approx 1 - \Phi\left(\frac{498.001 - 499.000}{\sqrt{249.500}}\right) \\ &= 1 - \Phi(-2) \\ &= \Phi(2) \\ &= 0,98 \end{aligned}$$

   Fehler:

   $\leq \frac{1}{\sqrt{2\pi n}} = 0,0003$

$k\sigma$- Intervalle der Binomialverteilung für große n

$$\begin{aligned} P(np - k\sigma \leq S_n \leq np + k\sigma) &= P(-k \leq \frac{S_n - np}{\sigma} \leq k) \\ &\approx \Phi(k) - \Phi(-k) \\ &= 2\Phi(k) - 1 \\ &= \left\{ \begin{array}{ll} 0,68 & ; k = 1 \\ 0,95 & ; k = 2 \\ 0,997 & ; k = 3 \end{array} \right\} \end{aligned}$$

Folie: Lotto am Sonnabend

Beispiel: Lotto 6 aus 49

A: die 13 ist unter den Glückszahlen

$P(A) = \frac{6}{49}$

$n = 2.099$ Ziehungen

X: Anzahl der Ziehungen mit der 13

Annahme:

$X \sim B(2.099, \frac{6}{49})$

$EX = 2.099 \cdot \frac{6}{49} \approx 257$

$VarX = 225,55$

$\sigma \approx 15$

Überprüfung der Faustregel:

$15 > 3$

$\Rightarrow$ approximative 2σ- und 3σ- Intervalle

2σ- Intervall: $[257 - 2 \cdot 15; 257 + 2 \cdot 15] = [227; 287]$

3σ- Intervall: $[257 - 3 \cdot 15; 257 + 3 \cdot 15] = [212; 307]$

Wenn man alle Zahlen betrachtet: $\mathcal{X}^2$- Anpassungstest

Ergebnis: kein Zweifel an der Gleichverteilung bei $\alpha = 0,05$

$\sqrt{n}$- Gesetz und $\frac{1}{\sqrt{n}}$- Gesetz

Sei n groß:

$P(np - 2\sqrt{np(1-p)} \leq S_n \leq np + 2\sqrt{np(1-p)}) \approx 0,95$

$p(1-p)$ ist maximal für $p = \frac{1}{2}$

d.h.

$\sqrt{p(1-p)} \leq \frac{1}{2}$

$$\begin{aligned} P(np - \sqrt{n} \leq S_n \leq np + \sqrt{n}) &\geq 0,95 \\ P(p - \frac{\sqrt{n}}{n} \leq \frac{S_n}{n} \leq p + \frac{\sqrt{n}}{n}) &\geq 0,95 \end{aligned}$$

$$P(p - \frac{1}{\sqrt{n}} \leq \frac{S_n}{n} \leq p + \frac{1}{\sqrt{n}}) \quad \geq \quad 0,95$$

$\sqrt{n}$- Gesetz:

Je mehr Versuche man macht, mit desto größeren Abweichungen der Anzahl der Erfolge vom erwarteten Wert np muss man rechnen.

Die Größenordnung der Abweichung ist $\sqrt{n}$.

$\frac{1}{\sqrt{n}}$- Gesetz:

Je mehr Versuche man macht, mit desto kleineren Abweichungen der relativen Häufigkeit der Erfolge vom erwartetetn Wert p muss man rechnen.

Die Größenordnung der Abweichung ist $\frac{1}{\sqrt{n}}$.

Schätzen und Testen einer unbekannten Wahrscheinlichkeit p bei großem n

Beispiel: Bundestagswahl 1972

Eine Woche vor der Wahl:

Stichprobe: $n = 2.000$ Wähler, davon $6,1\%$ für die FDP

Wie genau ist diese Schätzung bei einem Sicherheitsniveau von $1 - \alpha$?

$$\begin{aligned}
P\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) \quad &\geq \quad 1 - \alpha \\
P\left(|S_n - np| < n\epsilon\right) \quad &\geq \quad 1 - \alpha \\
P\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| < \frac{n\epsilon}{\sqrt{np(1-p)}}\right) \quad &\geq \quad 1 - \alpha \\
P\left(\underbrace{|S_n^*|}_{\sim N(0,1)} < \frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) \quad &\geq \quad 1 - \alpha \\
P\left(-\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}} < S_n^* < \frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) \quad &\geq \quad 1 - \alpha \\
\Phi\left(\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) \quad &>\approx \quad 1 - \alpha
\end{aligned}$$

$$
\begin{aligned}
2\Phi\left(\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right)-1 &\geq 1-\alpha \\
\Phi\left(\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) &\geq 1-\frac{\alpha}{2} \\
\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}} &\geq x_{1-\frac{\alpha}{2}} \\
\epsilon &\geq \frac{1}{\sqrt{n}} x_{1-\frac{\alpha}{2}} \sqrt{p(1-p)}
\end{aligned}
$$

günstigste Wahl im ungünstigsten Fall ($p = \frac{1}{2}$):

$$\epsilon = \frac{1}{2\sqrt{n}} x_{1-\frac{\alpha}{2}}$$

konkret:

$\alpha = 0,05 \Rightarrow X_{1-\frac{\alpha}{2}} = 1,96$

$$\epsilon = \frac{1}{2\sqrt{2.000}} 1,96 \approx 0,022$$

Mit einer Wahrscheinlichkeit von mindestens 0,95 weicht die relative Häufigkeit aus der Stichprobe vom Umfang 2.000 um höchstens $2,2\%$ vom wahren p ab.

Konkretes Konfidenzintervall $(0,039; 0,083)$

Eine Woche später hat die FDP $8,4\%$ der Stimmen erhalten.

Welcher Stichprobenumfang reicht aus, um eine vorgegebene Güte zu sichern?

$$
\begin{aligned}
P\left(\left|\frac{S_n}{n}-p\right|<\epsilon\right) &\geq 1-\alpha \\
&\vdots \\
\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}} &\geq x_{1-\frac{\alpha}{2}} \\
\sqrt{n} &\geq \frac{1}{\epsilon} x_{1-\frac{\alpha}{2}} \sqrt{p(1-p)} \\
n &\geq \frac{1}{\epsilon^2} x^2_{1-\frac{\alpha}{2}} \frac{1}{4} \\
n &= \frac{1}{4\epsilon^2} x^2_{1-\frac{\alpha}{2}}
\end{aligned}
$$

Beispiel:

$\epsilon = 0,01$

$1-\alpha = 0,95$

Zusatzinformation: $p \leq 0,2 \Rightarrow p(1-p) \leq 0,16$

$n = 3136$

Beispiel: Zwillingsgeburten:

1950- 1979: 32.486.960 Geburten, davon 349.983 Zwillingsgeburten

Mehrlingsgeburten sind eine Geburt.

$p_z = \frac{349.983}{32.486.960} = 0,0108$

Beobachtung: häufiger Mehrlingsgeburten (insbesondere Zwillingsgeburten)

H: $p_z = 0,0108$

A: $p_z > 0,0108$

Stichprobe: n Geburten, S_n Anzahl der Zwillingsgeburten

unter H: $S_n \sim B(n; 0,0108)$

$K = \{S_n \geq c\}$

$$c \geq \sigma x_{1-\alpha} + np_z$$

$c = 152.329 \Rightarrow$ H ablehnen

beobachtetes Signifikanzniveau:

Wie wahrscheinlich wäre das beobachtete oder ein noch extremeres Ereignis?

$P_H(S_n \geq 154.506)$

Bei welchem Signifikanzniveau wäre $c = 154.506$ möglich?

$P_H(S_n^* \geq \frac{154.506-151.689}{\sqrt{150.944}}) \approx 1 - \Phi(7,25) = 0$

3.2 Der zentrale GWS fuer unabhaengige und identisch verteilte ZG

nächste Verallgemeinerung zum GWS von de Moivre- Laplace:

$X_1, X_2, \ldots, X_n$- unabhängig

$S_n = X_1 + X_2 + \ldots + X_n$

$ES_n = E(X_1 + X_2 + \ldots + X_n) = EX_1 + \ldots + EX_n = nEX_1 = \mu$

(Vor.: Verteilung muss einen EW besitzen)

$VarS_n = Var(X_1 + X_2 + \ldots + X_n) = VarX_1 + \ldots + VarX_n = nVarX_1 = n\sigma^2$

(Vor.: $0 < \sigma^2 < \infty$)

Satz 3.3:

Es seinen $X_1, X_2, \ldots$ unabhängig, identisch verteilte ZG mit $0 < \sigma^2 < \infty$. Dann gilt für alle $x \in \Re$:

$$\lim_{n\to\infty} P(S_n^* \leq x) = \lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{n}\sigma} \leq x\right) = \Phi(x).$$

Bemerkungen:

1. $S_n = X_1 + \ldots + X_n$:

   Viele unabhängige, im Vergleich zum Ganzen kleine Effekte überlagern sich und liefern näherungsweise die Normalverteilung.

2. Es gibt zahlreiche Verallgemeinerungen dieses Satzes.

3. Mitteln von Messwerten $\bar{X}_n = \frac{S_n}{n}$

$$\begin{aligned} P(a \leq \bar{X}_n \leq b) &= P(an \leq S_n \leq bn) \\ &= P\left(\frac{(b-\mu)\sqrt{n}}{\sigma} \leq S_n^* \leq \frac{(a-\mu)\sqrt{n}}{\sigma}\right) \end{aligned}$$

   Die drei Probleme bei Messgrößen können näherungsweise gelöst werden.

   z.B.

   $P(\left|\bar{X}_n - \mu\right| \leq 3\frac{\sigma}{\sqrt{n}}) \approx 0,997$

   D.h. es gilt das $\frac{1}{\sqrt{n}}$- Gesetz.

4. Es existieren Aussagen über den Fehler der Approximation.