

Schließende Statistik

Vorlesung an der Universität des Saarlandes

Dr. Martin Becker

Wintersemester 2017/18



Organisatorisches I

- Vorlesung: Freitag, 12-14 Uhr, Gebäude B4 1, Audimax
- Übungen: siehe Homepage, Beginn: ab Montag (23.10.)
- Prüfung: 2-stündige Klausur nach Semesterende (1. Prüfungszeitraum)

Wichtig:

Anmeldung (ViPa) vom 13.-28. November (bis 15 Uhr) möglich

Abmeldung bis 19. Januar 2018 (12 Uhr) möglich

- Hilfsmittel für Klausur
 - ▶ „Moderat“ programmierbarer Taschenrechner, auch mit Grafikfähigkeit
 - ▶ 2 *beliebig gestaltete* DIN A 4-Blätter (bzw. 4, falls nur einseitig)
 - ▶ Benötigte Tabellen werden gestellt, aber **keine weitere Formelsammlung!**
- Durchgefallen — was dann?
 - ▶ „Wiederholungskurs“ im kommenden (Sommer-)Semester
 - ▶ „Nachprüfung“ (voraussichtlich) erst September/Oktober 2018 (2. Prüfungszeitraum)
 - ▶ „Reguläre“ Vorlesung/Übungen wieder im Wintersemester 2018/19

Organisatorisches II

- Kontakt: Dr. Martin Becker
Geb. C3 1, 2. OG, Zi. 2.17
e-Mail: martin.becker@mx.uni-saarland.de
- Sprechstunde nach Vereinbarung (Terminabstimmung per e-Mail)
- Informationen und Materialien auf Homepage:
<http://www.lehrstab-statistik.de>
- Material zu dieser Veranstaltung: Vorlesungsfolien *i.d.R. vor Vorlesung* zum Download (inklusive Drucker-freundlicher 2-auf-1 bzw. 4-auf-1 Versionen)
- Wie in „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“:
 - ▶ Neben theoretischer Einführung der Konzepte auch einige Beispiele auf Vorlesungsfolien
 - ▶ Einige wichtige Grundlagen werden gesondert als „Definition“, „Satz“ oder „Bemerkung“ hervorgehoben
 - ▶ **Aber:** Auch vieles, was nicht formal als „Definition“, „Satz“ oder „Bemerkung“ gekennzeichnet ist, ist wichtig!
- Übungsblätter *i.d.R.* nach Vorlesung zum Download
- Ergebnisse (*keine Musterlösungen!*) zu einigen Aufgaben im Netz
- Ausführlichere Lösungen zu den Übungsaufgaben (nur) in Übungsgruppen!

Benötigte Konzepte

aus den mathematischen Grundlagen

- Rechnen mit Potenzen

$$a^m \cdot b^m = (a \cdot b)^m \quad a^m \cdot a^n = a^{m+n} \quad \frac{a^m}{a^n} = a^{m-n} \quad (a^m)^n = a^{m \cdot n}$$

- Rechnen mit Logarithmen

$$\ln(a \cdot b) = \ln a + \ln b \quad \ln\left(\frac{a}{b}\right) = \ln a - \ln b \quad \ln(a^r) = r \cdot \ln a$$

- Rechenregeln auch mit Summen-/Produktzeichen, z.B.

$$\ln\left(\prod_{i=1}^n x_i^{r_i}\right) = \sum_{i=1}^n r_i \ln(x_i)$$

- Maximieren differenzierbarer Funktionen

- ▶ Funktionen (ggf. partiell) ableiten
- ▶ Nullsetzen von Funktionen (bzw. deren Ableitungen)

- „Unfallfreies“ Rechnen mit 4 Grundrechenarten und Brüchen...

Benötigte Konzepte

aus Veranstaltung „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“

- Diskrete und stetige Zufallsvariablen X , Verteilungsfunktionen, Wahrscheinlichkeitsverteilungen, ggf. Dichtefunktionen
- Momente (Erwartungswert $E(X)$, Varianz $\text{Var}(X)$, höhere Momente $E(X^k)$)
- „Einbettung“ der deskriptiven Statistik in die Wahrscheinlichkeitsrechnung
 - ▶ Ist Ω die (endliche) Menge von Merkmalsträgern einer deskriptiven statistischen Untersuchung, $\mathcal{F} = \mathcal{P}(\Omega)$ und P die Laplace-Wahrscheinlichkeit

$$P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}; B \mapsto \frac{\#B}{\#\Omega},$$

so kann jedes numerische Merkmal X als Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ verstanden werden.

- ▶ Der Träger von X entspricht dann dem Merkmalsraum $A = \{a_1, \dots, a_m\}$, die Punktwahrscheinlichkeiten den relativen Häufigkeiten, d.h. es gilt $p(a_j) = r(a_j)$ bzw. — äquivalent — $P_X(\{a_j\}) = r(a_j)$ für $j \in \{1, \dots, m\}$.
- Verteilung von $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ für unabhängig identisch verteilte X_i
 - ▶ falls X_i normalverteilt
 - ▶ falls $n \rightarrow \infty$ (Zentraler Grenzwertsatz!)

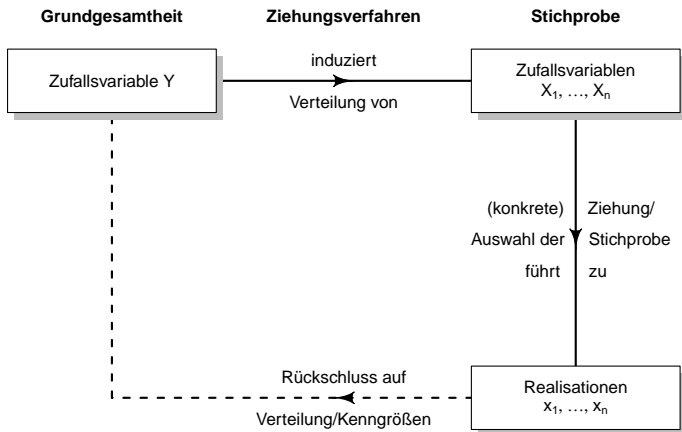
Grundidee der schließenden Statistik

- Ziel der schließenden Statistik/induktiven Statistik:

*Ziehen von Rückschlüssen auf die
Verteilung einer (größeren) Grundgesamtheit auf Grundlage der
Beobachtung einer (kleineren) Stichprobe.*

- Rückschlüsse auf die Verteilung können sich auch beschränken auf spezielle Eigenschaften/Kennzahlen der Verteilung, z.B. den Erwartungswert.
- „Fundament“: **Drei Grundannahmen**
 - 1 Der interessierende Umweltausschnitt kann durch eine (ein- oder mehrdimensionale) Zufallsvariable Y beschrieben werden.
 - 2 Man kann eine *Menge* W von Wahrscheinlichkeitsverteilungen angeben, zu der die *unbekannte* wahre Verteilung von Y gehört.
 - 3 Man beobachtet Realisationen x_1, \dots, x_n von (Stichproben-)Zufallsvariablen X_1, \dots, X_n , deren gemeinsame Verteilung *in vollständig bekannter Weise* von der Verteilung von Y abhängt.
- Ziel ist es also, aus der Beobachtung der n Werte x_1, \dots, x_n mit Hilfe des bekannten Zusammenhangs zwischen den Verteilungen von X_1, \dots, X_n und Y Aussagen über die Verteilung von Y zu treffen.

„Veranschaulichung“ der schließenden Statistik



Bemerkungen zu den 3 Grundannahmen

- Die 1. Grundannahme umfasst insbesondere die Situation, in der die Zufallsvariable Y einem (ein- oder mehrdimensionalen) Merkmal auf einer *endlichen* Menge von Merkmalsträgern entspricht, vgl. die Einbettung der deskriptiven Statistik in die Wahrscheinlichkeitsrechnung auf Folie 6. In diesem Fall interessiert man sich häufig für Kennzahlen von Y , z.B. den Erwartungswert von Y (als Mittelwert des Merkmals auf der Grundgesamtheit).
- Die Menge W von Verteilungen aus der 2. Grundannahme ist häufig eine *parametrische* Verteilungsfamilie, zum Beispiel die Menge aller Exponentialverteilungen oder die Menge aller Normalverteilungen mit Varianz $\sigma^2 = 2^2$. In diesem Fall ist die Menge der für die Verteilung von Y denkbaren Parameter interessant (später mehr!). Wir betrachten dann nur solche Verteilungsfamilien, in denen verschiedene Parameter auch zu verschiedenen Verteilungen führen („Parameter sind *identifizierbar*.“).
- Wir beschränken uns auf *sehr* einfache Zusammenhänge zwischen der Verteilung der interessierenden Zufallsvariablen Y und der Verteilung der Zufallsvariablen X_1, \dots, X_n .

Beispiel I

Stichprobe aus endlicher Grundgesamtheit Ω

- Grundgesamtheit: $N = 4$ Kinder (**A**nna, **B**eatrice, **C**hristian, **D**aniel) gleichen Alters, die in derselben Straße wohnen: $\Omega = \{A, B, C, D\}$
- Interessierender Umweltausschnitt: monatliches Taschengeld Y (in €) bzw. später spezieller: Mittelwert des monatlichen Taschengelds der 4 Kinder (entspricht $E(Y)$ bei Einbettung wie beschrieben)
- (Verteilungsannahme:) Verteilung von Y unbekannt, aber sicher in der Menge der diskreten Verteilungen mit maximal $N = 4$ (nichtnegativen) Trägerpunkten und Punktwahrscheinlichkeiten, die Vielfaches von $1/N = 1/4$ sind.

Im Beispiel nun: Zufallsvariable Y nehme Werte

ω	A	B	C	D
$Y(\omega)$	15	20	25	20

an, habe also folgende zugehörige Verteilung:

y_i	15	20	25	Σ
$p_Y(y_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Beispiel II

Stichprobe aus endlicher Grundgesamtheit Ω

- **Beachte:** Verteilung von Y nur im Beispiel bekannt, in der Praxis: Verteilung von Y natürlich unbekannt!
- Einfachste Möglichkeit, um Verteilung von Y bzw. deren Erwartungswert zu ermitteln: alle 4 Kinder nach Taschengeld befragen!
- Typische Situation in schließender Statistik: nicht alle Kinder können befragt werden, sondern nur eine kleinere Anzahl $n < N = 4$, beispielsweise $n = 2$. Erwartungswert von Y (mittleres Taschengeld aller 4 Kinder) kann dann nur noch **geschätzt** werden!
- Ziel: Rückschluss aus der Erhebung von $n = 2$ Taschengeldhöhen auf die größere Grundgesamtheit von $N = 4$ Kindern durch
 - ▶ Schätzung des mittleren Taschengeldes aller 4 Kinder
 - ▶ Beurteilung der Qualität der Schätzung (mit welchem „Fehler“ ist zu rechnen)
- (Qualität der) Schätzung hängt ganz entscheidend vom Ziehungs-/Auswahlverfahren ab!

Beispiel III

Stichprobe aus endlicher Grundgesamtheit Ω

- Erhebung von 2 Taschengeldhöhen führt zu Stichprobenezufallsvariablen X_1 und X_2 .
- X_1 bzw. X_2 entsprechen in diesem Fall dem Taschengeld des 1. bzw. 2. befragten Kindes
- **Sehr wichtig** für Verständnis:
 X_1 und X_2 sind Zufallsvariablen, da ihr Wert (Realisation) davon abhängt, **welche Kinder** man zufällig ausgewählt hat!
- Erst **nach Auswahl** der Kinder (also nach „Ziehung der Stichprobe“) steht der Wert (die Realisation) x_1 von X_1 bzw. x_2 von X_2 fest!

Variante A

- Naheliegendes Auswahlverfahren: nacheinander **rein zufällige** Auswahl von 2 der 4 Kinder, d.h. **zufälliges Ziehen ohne Zurücklegen mit Berücksichtigung der Reihenfolge**
- Alle $(4)_2 = 12$ Paare (A, B) ; (A, C) ; (A, D) ; (B, A) ; (B, C) ; (B, D) ; (C, A) ; (C, B) ; (C, D) ; (D, A) ; (D, B) ; (D, C) treten dann mit der gleichen Wahrscheinlichkeit $(1/12)$ auf und führen zu den folgenden „Stichprobenrealisationen“ (x_1, x_2) der Stichprobenvariablen (X_1, X_2) :

Beispiel IV

Stichprobe aus endlicher Grundgesamtheit Ω

- Realisationen (x_1, x_2) zur Auswahl von 1. Kind (Zeilen)/2. Kind (Spalten):

	A	B	C	D
A	unmöglich	(15,20)	(15,25)	(15,20)
B	(20,15)	unmöglich	(20,25)	(20,20)
C	(25,15)	(25,20)	unmöglich	(25,20)
D	(20,15)	(20,20)	(20,25)	unmöglich

- Resultierende gemeinsame Verteilung von (X_1, X_2) :

$x_1 \backslash x_2$	15	20	25	Σ
15	0	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{4}$
20	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
25	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
Σ	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

- Es fällt auf (**Variante A**):
 - X_1 und X_2 haben die gleiche Verteilung wie Y .
 - X_1 und X_2 sind **nicht** stochastisch unabhängig.

Beispiel V

Stichprobe aus endlicher Grundgesamtheit Ω

- Naheliegend: Schätzung des Erwartungswertes $E(Y)$, also des mittleren Taschengeldes aller 4 Kinder, durch den (arithmetischen) Mittelwert der erhaltenen Werte für die 2 befragten Kinder.
- **Wichtig: Nach** Auswahl der Kinder ist dieser Mittelwert eine Zahl, es ist aber sehr nützlich, den Mittelwert schon **vor** Auswahl der Kinder (dann) als Zufallsvariable (der Zufall kommt über die zufällige Auswahl der Kinder ins Spiel) zu betrachten!
- Interessant ist also die Verteilung der **Zufallsvariable** $\bar{X} := \frac{1}{2}(X_1 + X_2)$, also des Mittelwerts der Stichprobenzufallsvariablen X_1 und X_2 .
Die (hiervon zu unterscheidende!) **Realisation** $\bar{x} = \frac{1}{2}(x_1 + x_2)$ ergibt sich erst (als Zahlenwert) nach Auswahl der Kinder (wenn die Realisation (x_1, x_2) von (X_1, X_2) vorliegt)!
- Verteilung von \bar{X} hier (**Variante A**):

\bar{x}_i	17.5	20	22.5	Σ
$p_{\bar{X}}(\bar{x}_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Beispiel VI

Stichprobe aus endlicher Grundgesamtheit Ω

Variante B

- Weiteres mögliches Auswahlverfahren: 2-fache **rein zufällige** und **voneinander unabhängige** Auswahl eines der 4 Kinder, wobei erlaubt ist, dasselbe Kind mehrfach auszuwählen, d.h. **zufälliges Ziehen mit Zurücklegen und Berücksichtigung der Reihenfolge**
- Alle $4^2 = 16$ Paare (A, A) ; (A, B) ; (A, C) ; (A, D) ; (B, A) ; (B, B) ; (B, C) ; (B, D) ; (C, A) ; (C, B) ; (C, C) ; (C, D) ; (D, A) ; (D, B) ; (D, C) ; (D, D) treten dann mit der gleichen Wahrscheinlichkeit $(1/16)$ auf und führen zu den folgenden „Stichprobenrealisationen“ (x_1, x_2) der Stichprobenvariablen (X_1, X_2) (zur Auswahl von 1. Kind (Zeilen)/2. Kind (Spalten)):

	A	B	C	D
A	(15,15)	(15,20)	(15,25)	(15,20)
B	(20,15)	(20,20)	(20,25)	(20,20)
C	(25,15)	(25,20)	(25,25)	(25,20)
D	(20,15)	(20,20)	(20,25)	(20,20)

Beispiel VII

Stichprobe aus endlicher Grundgesamtheit Ω

- Resultierende gemeinsame Verteilung von (X_1, X_2) :

$x_1 \backslash x_2$	15	20	25	Σ
15	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{4}$
20	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
25	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{4}$
Σ	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

- Es fällt auf (**Variante B**):
 - X_1 und X_2 haben die gleiche Verteilung wie Y .
 - X_1 und X_2 **sind** stochastisch unabhängig.
- Verteilung von \bar{X} hier (**Variante B**):

\bar{x}_i	15	17.5	20	22.5	25	Σ
$p_{\bar{X}}(\bar{x}_i)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$	1

Zufallsstichprobe

- Beide Varianten zur Auswahl der Stichprobe führen dazu, dass alle Stichprobenezufallsvariablen X_i ($i = 1, 2$) **identisch** verteilt sind wie Y .
- Variante **B** führt außerdem dazu, dass die Stichprobenezufallsvariablen X_i ($i = 1, 2$) **stochastisch unabhängig** sind.

Definition 2.1 ((Einfache) Zufallsstichprobe)

Seien $n \in \mathbb{N}$ und X_1, \dots, X_n Zufallsvariablen einer Stichprobe vom Umfang n zu Y . Dann heißt (X_1, \dots, X_n)

- ▶ **Zufallsstichprobe** vom Umfang n zu Y , falls die Verteilungen von Y und X_i für alle $i \in \{1, \dots, n\}$ übereinstimmen, alle X_i also identisch verteilt sind wie Y ,
 - ▶ **einfache (Zufalls-)Stichprobe** vom Umfang n zu Y , falls die Verteilungen von Y und X_i für alle $i \in \{1, \dots, n\}$ übereinstimmen und X_1, \dots, X_n außerdem stochastisch unabhängig sind.
- (X_1, X_2) ist in Variante A des Beispiels also eine Zufallsstichprobe vom Umfang 2 zu Y , in Variante B sogar eine einfache (Zufalls-)Stichprobe vom Umfang 2 zu Y .

- X_1, \dots, X_n ist also nach Definition 2.1 auf Folie 17 genau dann eine **Zufallsstichprobe**, falls für die Verteilungsfunktionen zu Y, X_1, \dots, X_n

$$F_Y = F_{X_1} = \dots = F_{X_n}$$

gilt.

- Ist X_1, \dots, X_n eine **einfache Stichprobe** vom Umfang n zu Y , so gilt für die *gemeinsame* Verteilungsfunktion von (X_1, \dots, X_n) sogar

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

Ist Y diskrete Zufallsvariable gilt also insbesondere für die beteiligten Wahrscheinlichkeitsfunktionen

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_Y(x_1) \cdot \dots \cdot p_Y(x_n) = \prod_{i=1}^n p_Y(x_i) ,$$

ist Y stetige Zufallsvariable, so existieren Dichtefunktionen von Y bzw. (X_1, \dots, X_n) mit

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

Stichprobenrealisation/Stichprobenraum

Definition 2.2 (Stichprobenrealisation/Stichprobenraum)

Seien $n \in \mathbb{N}$ und X_1, \dots, X_n Zufallsvariablen einer Stichprobe vom Umfang n zu Y . Seien x_1, \dots, x_n die beobachteten Realisationen zu den Zufallsvariablen X_1, \dots, X_n . Dann heißt

- (x_1, \dots, x_n) **Stichprobenrealisation** und
 - die Menge \mathcal{X} aller möglichen Stichprobenrealisationen **Stichprobenraum**.
-
- Es gilt offensichtlich immer $\mathcal{X} \subseteq \mathbb{R}^n$.
 - „Alle möglichen Stichprobenrealisationen“ meint alle Stichprobenrealisationen, die für *irgendeine* der möglichen Verteilungen W von Y aus der Verteilungsannahme möglich sind.
 - Wenn man davon ausgeht, dass ein Kind „schlimmstenfalls“ $0 \in$ Taschengeld erhält, wäre im Beispiel also $\mathcal{X} = \mathbb{R}_+^2$ (Erinnerung: $\mathbb{R}_+ := \{x \in \mathbb{R} \mid x \geq 0\}$).
 - Meist wird die Information der Stichprobenezufallsvariablen bzw. der Stichprobenrealisation weiter mit sog. „Stichprobenfunktionen“ aggregiert, die oft (große) Ähnlichkeit mit Funktionen haben, die in der deskriptiven Statistik zur Aggregation von Urlisten eingesetzt werden.

Stichprobenfunktion/Statistik

Definition 2.3 (Stichprobenfunktion/Statistik)

Seien $n \in \mathbb{N}$ und X_1, \dots, X_n Zufallsvariablen einer Stichprobe vom Umfang n zu Y mit Stichprobenraum \mathcal{X} . Dann heißt eine Abbildung

$$T : \mathcal{X} \rightarrow \mathbb{R}; (x_1, \dots, x_n) \mapsto T(x_1, \dots, x_n)$$

Stichprobenfunktion oder Statistik.

- Stichprobenfunktionen sind also Abbildungen, deren Wert mit Hilfe der Stichprobenrealisation bestimmt werden kann.
- Stichprobenfunktionen müssen (geeignet, z.B. \mathcal{B}^n - \mathcal{B} -) messbare Abbildungen sein; diese Anforderung ist aber für alle hier interessierenden Funktionen erfüllt, Messbarkeitsüberlegungen bleiben also im weiteren Verlauf außen vor.
- Ebenfalls als Stichprobenfunktion bezeichnet wird die (als Hintereinanderausführung zu verstehende) Abbildung $T(X_1, \dots, X_n)$, wegen der Messbarkeitseigenschaft ist dies immer eine **Zufallsvariable**. Die Untersuchung der zugehörigen Verteilung ist für viele Anwendungen von **ganz wesentlicher** Bedeutung.

- Wenn man sowohl die Zufallsvariable $T(X_1, \dots, X_n)$ als auch den aus einer vorliegenden Stichprobenrealisation (x_1, \dots, x_n) resultierenden Wert $T(x_1, \dots, x_n)$ betrachtet, so bezeichnet man $T(x_1, \dots, x_n)$ oft auch als **Realisation** der Stichprobenfunktion.
- Im Taschengeld-Beispiel war die betrachtete Stichprobenfunktion das arithmetische Mittel, also konkreter

$$T : \mathbb{R}^2 \rightarrow \mathbb{R}; T(x_1, x_2) = \bar{x} := \frac{1}{2}(x_1 + x_2)$$

bzw. — als Zufallsvariable betrachtet —

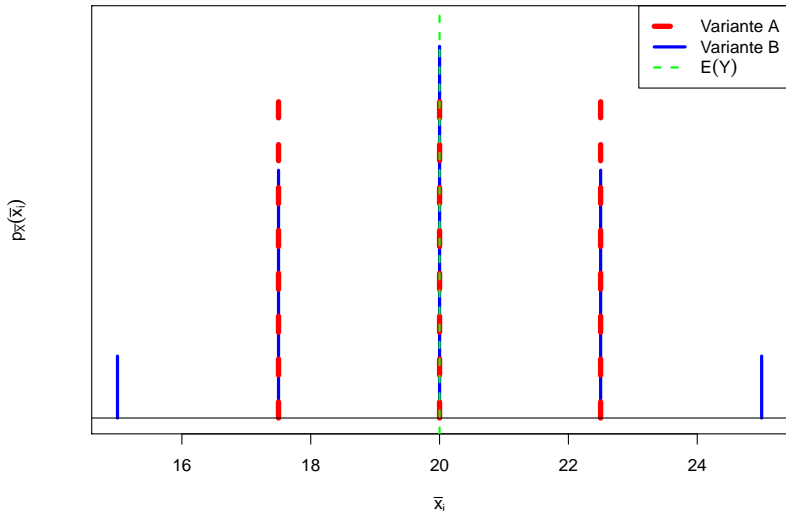
$$T(X_1, X_2) = \bar{X} := \frac{1}{2}(X_1 + X_2) .$$

- Je nach Anwendung erhalten Stichprobenfunktionen auch speziellere Bezeichnungen, z. B.
 - ▶ **Schätzfunktion** oder **Schätzer**, wenn die Stichprobenfunktion zur Schätzung eines Verteilungsparameters oder einer Verteilungskennzahl verwendet wird (wie im Beispiel!),
 - ▶ **Teststatistik**, wenn auf Grundlage der Stichprobenfunktion Entscheidungen über die Ablehnung oder Annahme von Hypothesen über die Verteilung von Y getroffen werden.

Beispiel VIII

Stichprobe aus endlicher Grundgesamtheit Ω

Vergleich der Verteilungen von \bar{X} in beiden Varianten:



Beispiel IX

Stichprobe aus endlicher Grundgesamtheit Ω

- Verteilung von Y

y_i	15	20	25	Σ
$p_Y(y_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

hat Erwartungswert $E(Y) = 20$ und Standardabweichung $Sd(Y) \approx 3.536$.

- Verteilung von \bar{X} (Variante **A**):

\bar{x}_i	17.5	20	22.5	Σ
$p_{\bar{X}}(\bar{x}_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

hat Erwartungswert $E(\bar{X}) = 20$ und Standardabweichung $Sd(\bar{X}) \approx 2.041$.

- Verteilung von \bar{X} (Variante **B**):

\bar{x}_i	15	17.5	20	22.5	25	Σ
$p_{\bar{X}}(\bar{x}_i)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$	1

hat Erwartungswert $E(\bar{X}) = 20$ und Standardabweichung $Sd(\bar{X}) = 2.5$.

Beispiel X

Stichprobe aus endlicher Grundgesamtheit Ω

- In beiden Varianten schätzt man das mittlere Taschengeld $E(Y) = 20$ also „im Mittel“ richtig, denn es gilt für beide Varianten $E(\bar{X}) = 20 = E(Y)$.
- Die Standardabweichung von \bar{X} ist in Variante A kleiner als in Variante B; zusammen mit der Erkenntnis, dass beide Varianten „im Mittel“ richtig liegen, schätzt also Variante A „genauer“.
- In beiden Varianten hängt es vom Zufall (genauer von der konkreten Auswahl der beiden Kinder — bzw. in Variante B möglicherweise zweimal desselben Kindes — ab), ob man *nach Durchführung der Stichprobenziehung* den tatsächlichen Mittelwert als Schätzwert erhält oder nicht.
- Obwohl \bar{X} in Variante A die kleinere Standardabweichung hat, erhält man in Variante B den tatsächlichen Mittelwert $E(Y) = 20$ mit einer größeren Wahrscheinlichkeit ($3/8$ in Variante B gegenüber $1/3$ in Variante A).

Parameterpunktschätzer

- Im Folgenden: Systematische Betrachtung der Schätzung von Verteilungsparametern, wenn die Menge W der (möglichen) Verteilungen von Y eine **parametrische** Verteilungsfamilie gemäß folgender Definition ist: (Z.T. Wdh. aus „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“)

Definition 3.1 (Parametrische Verteilungsfamilie, Parameterraum)

- 1 Eine Menge von Verteilungen W heißt **parametrische Verteilungsfamilie**, wenn jede Verteilung in W durch einen endlich-dimensionalen Parameter $\theta = (\theta_1, \dots, \theta_K) \in \Theta \subseteq \mathbb{R}^K$ charakterisiert wird.

Um die Abhängigkeit von θ auszudrücken, notiert man die Verteilungen, Verteilungsfunktionen sowie die Wahrscheinlichkeits- bzw. Dichtefunktionen häufig als

$$P(\cdot | \theta_1, \dots, \theta_K), F(\cdot | \theta_1, \dots, \theta_K) \text{ sowie } p(\cdot | \theta_1, \dots, \theta_K) \text{ bzw. } f(\cdot | \theta_1, \dots, \theta_K).$$

- 2 Ist W die Menge von Verteilungen aus der 2. Grundannahme („Verteilungsannahme“), so bezeichnet man W auch als **parametrische Verteilungsannahme**. Die Menge Θ heißt dann auch **Parameterraum**.

Bemerkungen

- Wir betrachten nur „identifizierbare“ parametrische Verteilungsfamilien, das heißt, unterschiedliche Parameter aus dem Parameterraum Θ müssen auch zu unterschiedlichen Verteilungen aus W führen.
- Die Bezeichnung θ dient lediglich zur vereinheitlichten Notation. In der Praxis behalten die Parameter meist ihre ursprüngliche Bezeichnung.
- In der Regel gehören alle Verteilungen in W zum gleichen Typ, zum Beispiel als
 - ▶ Bernouilliverteilung $B(1, p)$: Parameter $p \equiv \theta$, Parameterraum $\Theta = [0, 1]$
 - ▶ Poissonverteilung $\text{Pois}(\lambda)$: Parameter $\lambda \equiv \theta$, Parameterraum $\Theta = \mathbb{R}_{++}$
 - ▶ Exponentialverteilung $\text{Exp}(\lambda)$: Parameter $\lambda \equiv \theta$, Parameterraum $\Theta = \mathbb{R}_{++}$
 - ▶ Normalverteilung $N(\mu, \sigma^2)$: Parametervektor $(\mu, \sigma^2) \equiv (\theta_1, \theta_2)$,
Parameterraum $\mathbb{R} \times \mathbb{R}_{++}$
 (mit $\mathbb{R}_{++} := \{x \in \mathbb{R} \mid x > 0\}$).
- Suche nach **allgemein anwendbaren** Methoden zur Konstruktion von Schätzfunktionen für unbekannte Parameter θ aus parametrischen Verteilungsannahmen.
- Schätzfunktionen für einen Parameter(vektor) θ sowie deren Realisationen (!) werden üblicherweise mit $\hat{\theta}$, gelegentlich auch mit $\tilde{\theta}$ bezeichnet.
- Meist wird vom Vorliegen einer einfachen Stichprobe ausgegangen.

Methode der Momente (Momentenmethode)

- Im Taschengeldbeispiel: Schätzung des Erwartungswerts $E(Y)$ *naheliegenderweise* durch das arithmetische Mittel $\bar{X} = \frac{1}{2} (X_1 + X_2)$.
- Dies entspricht der Schätzung des 1. (theoretischen) Moments von Y durch das 1. empirische Moment der Stichprobenrealisation (aufgefasst als Urliste im Sinne der deskriptiven Statistik).
- Gleichsetzen von theoretischen und empirischen Momenten bzw. Ersetzen theoretischer durch empirische Momente führt zur gebräuchlichen **(Schätz-)Methode der Momente** für die Parameter von parametrischen Verteilungsfamilien.
- Grundlegende Idee: Schätze Parameter der Verteilung so, dass zugehörige theoretische Momente $E(Y)$, $E(Y^2)$, ... mit den entsprechenden empirischen Momenten \bar{X} , \bar{X}^2 , ... der Stichprobenzufallsvariablen X_1, \dots, X_n (bzw. deren Realisationen) übereinstimmen.
- Es werden dabei (beginnend mit dem ersten Moment) gerade so viele Momente einbezogen, dass das entstehende Gleichungssystem für die Parameter eine eindeutige Lösung hat.
Bei eindimensionalen Parameterräumen genügt *i.d.R.* das erste Moment.

Momente von Zufallsvariablen

- Bereits aus „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“ bekannt ist die folgende Definition für die (theoretischen) Momente von Zufallsvariablen:

Definition 3.2 (k -te Momente)

Es seien Y eine (eindimensionale) Zufallsvariable, $k \in \mathbb{N}$.

Man bezeichnet den Erwartungswert $E(Y^k)$ (falls er existiert) als das **(theoretische) Moment k -ter Ordnung** von Y , oder auch das **k -te (theoretische) Moment** von Y und schreibt auch kürzer

$$E Y^k := E(Y^k).$$

- Erinnerung (unter Auslassung der Existenzbetrachtung!):
Das k -te Moment von Y berechnet man für diskrete bzw. stetige Zufallsvariablen Y durch

$$E(Y^k) = \sum_{y_i} y_i^k \cdot p_Y(y_i) \quad \text{bzw.} \quad E(Y^k) = \int_{-\infty}^{\infty} y^k \cdot f_Y(y) dy ,$$

wobei y_i (im diskreten Fall) alle Trägerpunkte von Y durchläuft.

Empirische Momente von Stichproben

- Analog zu empirischen Momenten von Urlisten in der deskriptiven Statistik definiert man empirische Momente von Stichproben in der schließenden Statistik wie folgt:

Definition 3.3 (empirische Momente)

Ist (X_1, \dots, X_n) eine (einfache) Zufallsstichprobe zu einer Zufallsvariablen Y , so heißt

$$\overline{X^k} := \frac{1}{n} \sum_{i=1}^n X_i^k$$

das **empirische k -te Moment**, oder auch das **Stichprobenmoment der Ordnung k** . Zu einer Realisation (x_1, \dots, x_n) von (X_1, \dots, X_n) bezeichnet

$$\overline{x^k} := \frac{1}{n} \sum_{i=1}^n x_i^k$$

entsprechend die zugehörige **Realisation** des k -ten empirischen Moments.

Durchführung der Momentenmethode

- Zur Durchführung der Momentenmethode benötigte Anzahl von Momenten meist gleich der Anzahl der zu schätzenden Verteilungsparameter.
- Übliche Vorgehensweise:
 - ▶ Ausdrücken/Berechnen der theoretischen Momente in Abhängigkeit der Verteilungsparameter
 - ▶ Gleichsetzen der theoretischen Momente mit den entsprechenden empirischen Momenten und Auflösen der entstehenden Gleichungen nach den Verteilungsparametern.
- Alternativ, falls Verteilungsparameter Funktionen theoretischer Momente sind: Ersetzen der theoretischen Momente in diesen „Formeln“ für die Verteilungsparameter durch die entsprechenden empirischen Momente.
- Nützlich ist für die alternative Vorgehensweise gelegentlich der Varianzzerlegungssatz

$$\text{Var}(X) = E(X^2) - [E(X)]^2 .$$

Beispiele (Momentenmethode) I

1 Schätzung des Parameters p einer Alternativ-/Bernoulliverteilung:

- ▶ Verteilungsannahme: $W = \{B(1, p) \mid p \in \Theta = [0, 1]\}$
- ▶ Theoretisches 1. Moment: $E(Y) = p$ (bekannt aus W 'rechnung)
- ▶ Gleichsetzen (hier besonders einfach!) von $E(Y)$ mit 1. empirischen Moment \bar{X} liefert sofort Momentenmethodenschätzer (Methode 1) $\hat{p} = \bar{X}$.

Der Schätzer \hat{p} für die Erfolgswahrscheinlichkeit p nach der Methode der Momente entspricht also gerade dem Anteil der Erfolge in der Stichprobe.

2 Schätzung des Parameters λ einer Exponentialverteilung:

- ▶ Verteilungsannahme: $W = \{\text{Exp}(\lambda) \mid \lambda \in \Theta = \mathbb{R}_{++}\}$
- ▶ Theoretisches 1. Moment: $E(Y) = \frac{1}{\lambda}$ (bekannt aus W 'rechnung)
- ▶ Gleichsetzen von $E(Y)$ mit 1. empirischen Moment \bar{X} liefert (Methode 1)

$$\bar{X} \stackrel{!}{=} E(Y) = \frac{1}{\lambda} \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{\bar{X}}.$$

(Vorsicht bei Berechnung der Realisation: $\frac{1}{\bar{x}} \neq \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}$)

Beispiele (Momentenmethode) II

④ Schätzung der Parameter (μ, σ^2) einer Normalverteilung:

- ▶ Verteilungsannahme: $W = \{N(\mu, \sigma^2) \mid (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_{++}\}$
Hier bekannt: $E(Y) = \mu$ und $\text{Var}(Y) = \sigma^2$.
↪ Alternative Methode bietet sich an (mit Varianzzerlegungssatz):
- ▶ Verteilungsparameter $\mu = E(Y)$
Verteilungsparameter $\sigma^2 = E(Y^2) - [E(Y)]^2$
- ▶ Einsetzen der empirischen Momente anstelle der theoretischen Momente liefert $\hat{\mu} = \bar{X}$ sowie $\hat{\sigma}^2 = \overline{X^2} - \bar{X}^2$ als Schätzer nach der Momentenmethode.
- ▶ Am Beispiel der Realisation

8.75, 10.37, 8.33, 13.19, 10.66, 8.36, 10.97, 11.48, 11.15, 9.39

einer Stichprobe vom Umfang 10 erhält man mit

$$\bar{x} = 10.265 \quad \text{und} \quad \overline{x^2} = 107.562$$

die realisierten Schätzwerte

$$\hat{\mu} = 10.265 \quad \text{und} \quad \hat{\sigma}^2 = 107.562 - 10.265^2 = 2.192 .$$