

1.7 Grenzwertsätze und Approximationen

Gerade in der empirischen Sozialforschung arbeitet man häufig mit *großen* Stichprobenumfängen.

- Was ist aus der Sicht der Wahrscheinlichkeitsrechnung das Besondere daran?
- Vereinfacht sich etwas und wenn ja was?
- Kann man „Wahrscheinlichkeitsgesetzmäßigkeiten“ durch Betrachten vielfacher Wiederholungen erkennen?

1.7.1 Das i.i.d.-Modell

Betrachtet werden diskrete oder stetige Zufallsvariablen X_1, \dots, X_n , die *i.i.d.* (independently, identically distributed) sind, d.h. die

- 1) unabhängig sind und
- 2) die gleiche Verteilung besitzen.

Ferner sollen der Erwartungswert μ und die Varianz σ^2 existieren. Die Verteilungsfunktion werde mit F bezeichnet.

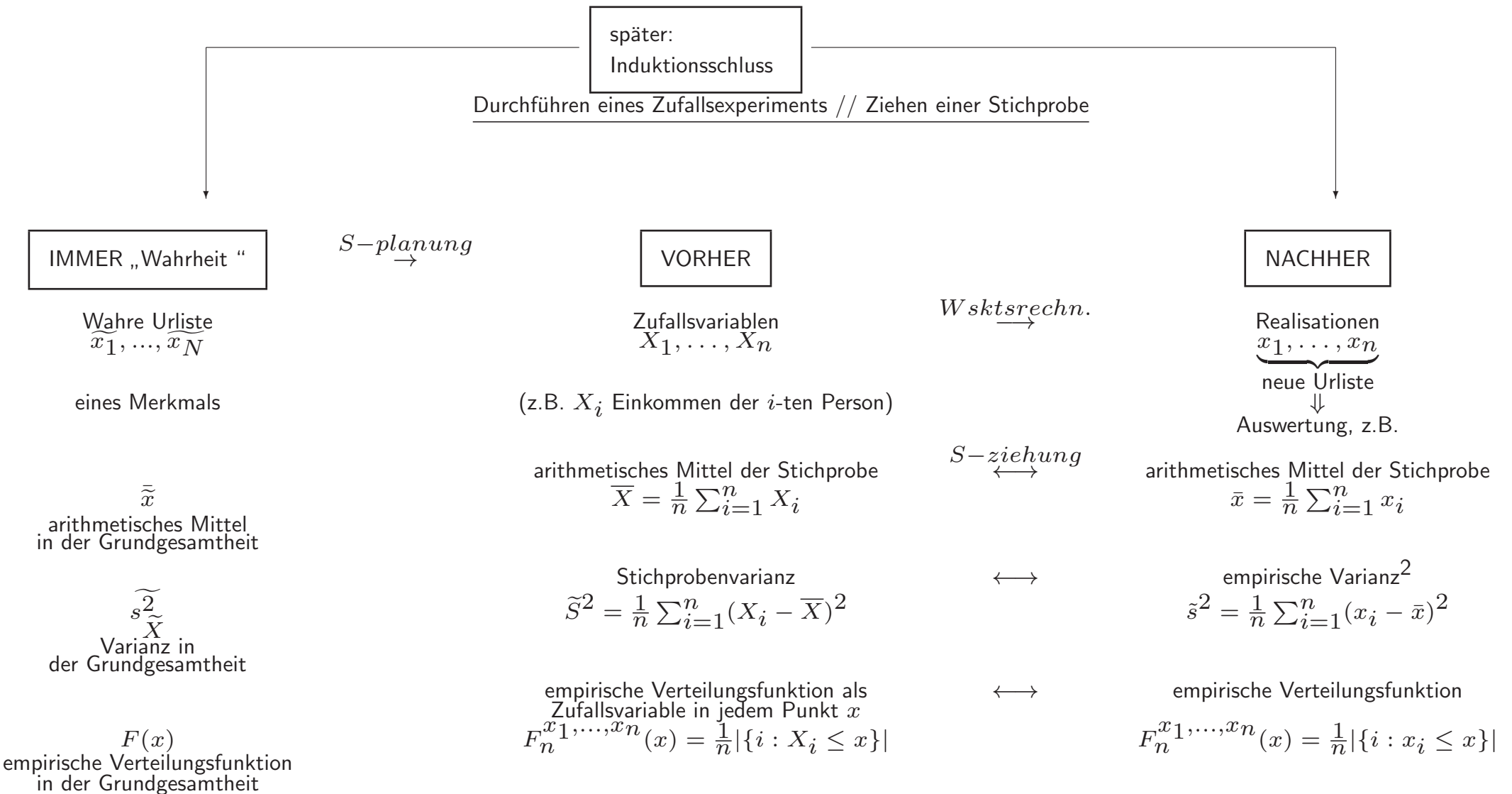
Dies bildet insbesondere die Situation ab, in der X_1, \dots, X_n eine Stichprobe eines Merkmals X bei reiner Zufallsauswahl bilden.

Jede Funktion von X_1, \dots, X_n ist wieder eine Zufallsvariable, z.B. das arithmetische Mittel oder die Stichprobenvarianz

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Vor dem Ziehen der Stichprobe: Wahrscheinlichkeitsaussagen möglich \implies Wahrscheinlichkeitsrechnung anwenden

- Gerade bei diesen Zufallsgrößen ist die Abhängigkeit von n oft wichtig, man schreibt dann \bar{X}_n, \tilde{S}_n^2
- Sind X_1, \dots, X_n jeweils $\{0, 1\}$ -Variablen, so ist \bar{X}_n gerade die empirische *relative Häufigkeit* von Einsen in der Stichprobe vom Umfang n . Notation: H_n

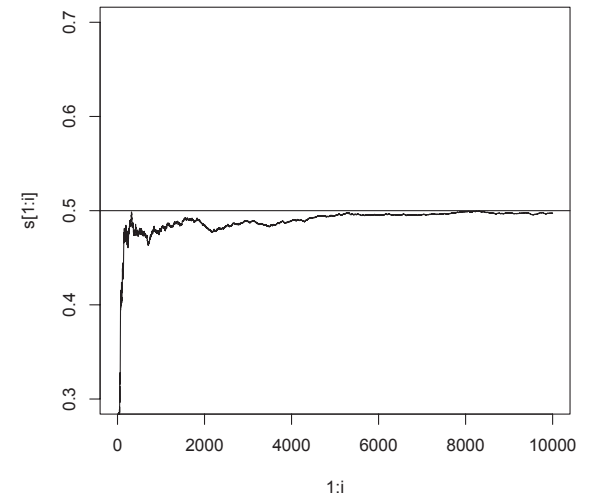
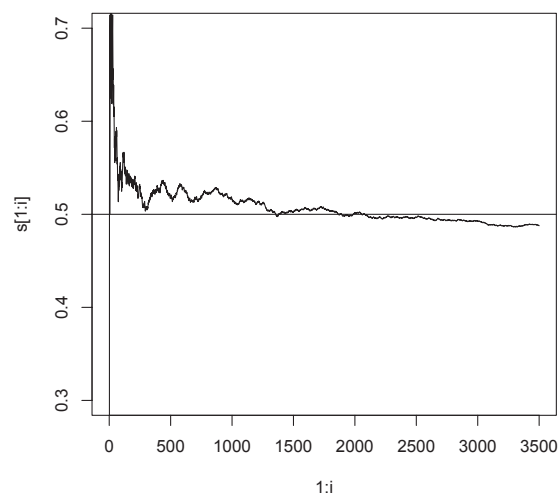
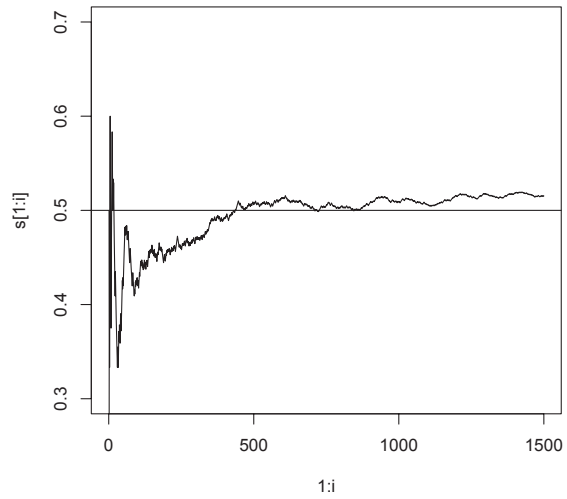
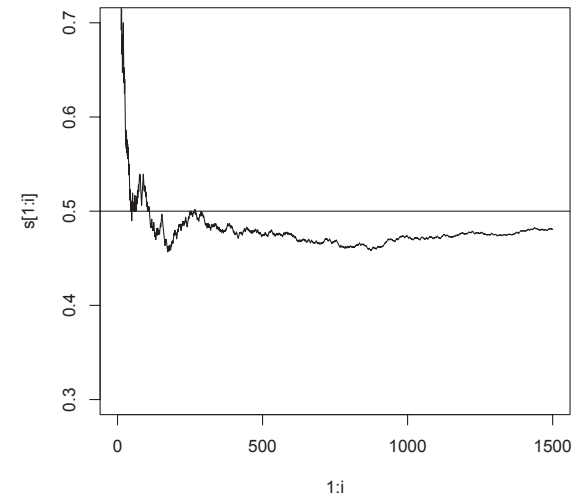
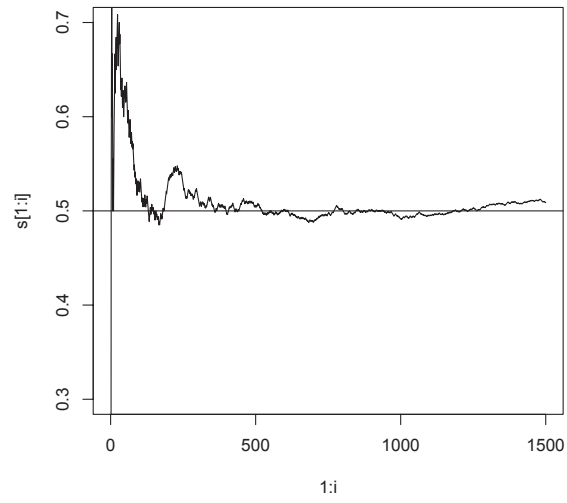
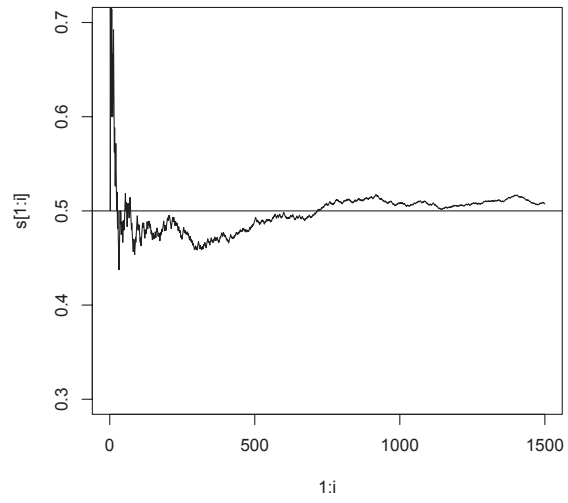


²Gehört nicht zur Grundgesamtheit; hier „ $\tilde{}$ “ für empirische Version

1.7.2 Das schwache Gesetz der großen Zahlen

Betrachte für wachsenden Stichprobenumfang n :

- X_1, \dots, X_n i.i.d.
- $X_i \in \{0, 1\}$ binäre Variablen mit $\pi = P(X_i = 1)$
- $H_n =$ die relative Häufigkeit der Einsen in den ersten n Versuchen.



Was fällt auf?

Theorem 1.81. [Theorem von Bernoulli]

Seien X_1, \dots, X_n , i.i.d. mit $X_i \in \{0, 1\}$ und $P(X_i = 1) = \pi$. Dann gilt für

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

(relative Häufigkeit der „Einsen“) und beliebig kleines $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|H_n - \pi| \leq \epsilon) = 1.$$

Anschauliche Interpretation: Die relative Häufigkeit eines Ereignisses nähert sich praktisch sicher mit wachsender Versuchszahl an die Wahrscheinlichkeit des Ereignisses an.

Zwei wichtige Konsequenzen:

- 1) Häufigkeitsinterpretation von Wahrscheinlichkeiten:
- 2) Induktion: Man kann dieses Ergebnis nutzen, um Information über eine unbekannte Wahrscheinlichkeit ($\pi \hat{=}$ Anteil in einer Grundgesamtheit) zu erhalten.

Das Ergebnis lässt sich verallgemeinern auf Mittelwerte beliebiger Zufallsvariablen:

Schwaches Gesetz der großen Zahl: Gegeben seien X_1, \dots, X_n i.i.d. Zufallsvariablen mit (existierendem) Erwartungswert μ und (existierender) Varianz σ^2 . Dann gilt für

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

und beliebiges $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq \epsilon) = 1$$

Schreibweise:

$$\bar{X}_n \xrightarrow{P} \mu$$

(„Stochastische Konvergenz“, „ \bar{X}_n konvergiert in Wahrscheinlichkeit gegen μ “.)

Konsequenz für die Interpretation des Erwartungswerts:

1.7.3 Der Hauptsatz der Statistik

Satz 1.82. [Hauptsatz der Statistik]

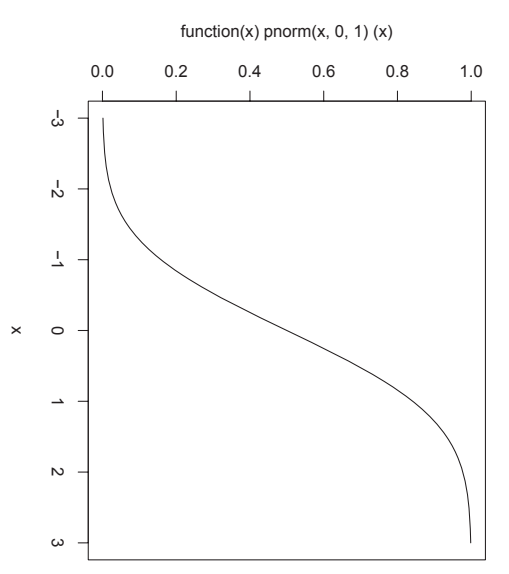
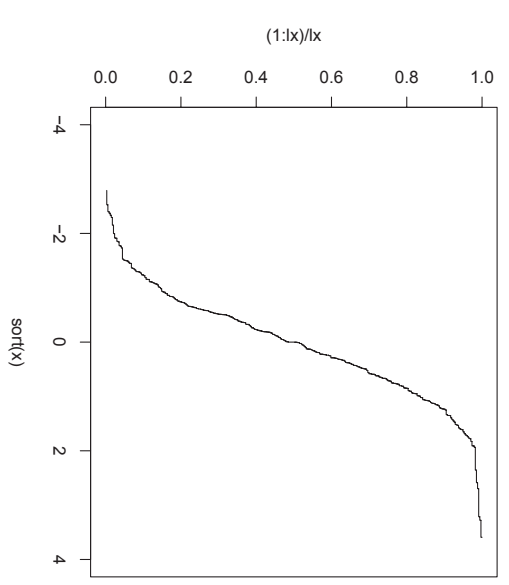
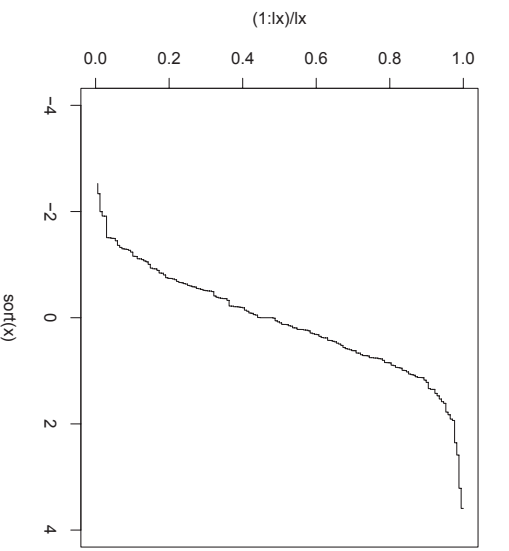
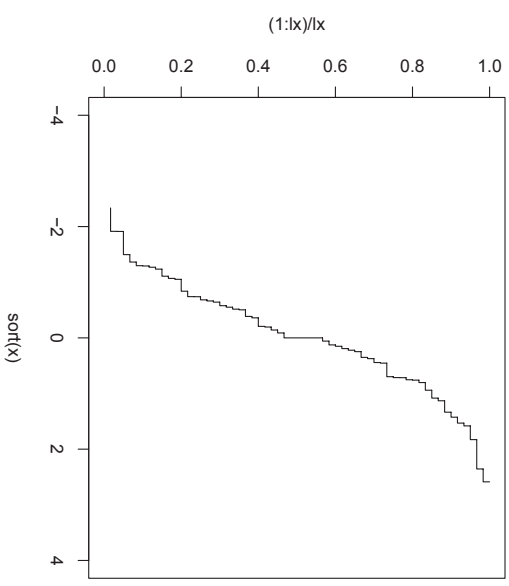
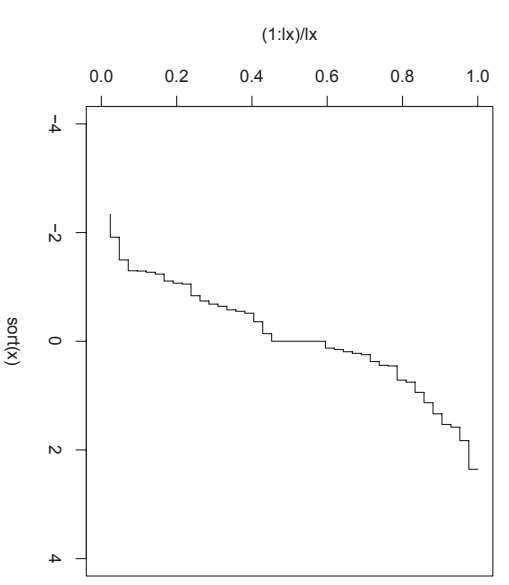
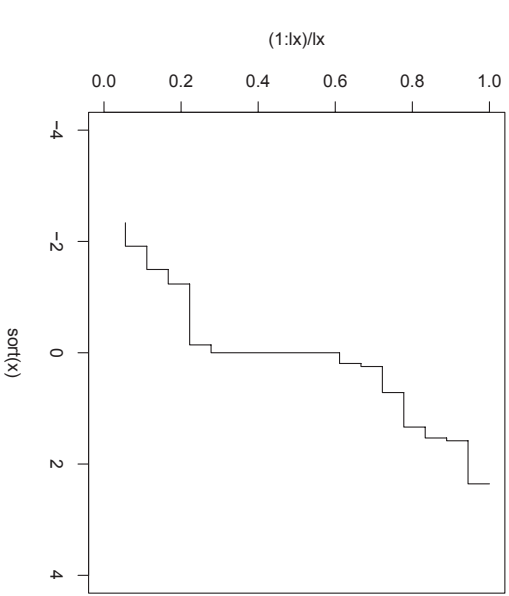
Seien X_1, \dots, X_n i.i.d. mit Verteilungsfunktion F und sei $F_n(x)$ die empirische Verteilungsfunktion der ersten n Beobachtungen. Mit

$$D_n := \sup_x |F_n(x) - F(x)|,$$

gilt für jedes $c > 0$

$$\lim_{n \rightarrow \infty} P(D_n > c) = 0.$$

Interpretation:



1.7.4 Der zentrale Grenzwertsatz

- Gibt es für große Stichprobenumfänge Regelmäßigkeiten im Verteilungstyp?
- Gibt es eine Standardverteilung, mit der man oft bei großen empirischen Untersuchungen rechnen kann? Damit kann man dann insbesondere Fehlermargen einheitlich behandeln.

Satz 1.83. [Zentraler Grenzwertsatz]

Seien X_1, \dots, X_n i.i.d. mit (existierendem) Erwartungswert $E(X_i)$ und (existierender) Varianz $\text{Var}(X_i) > 0$ sowie

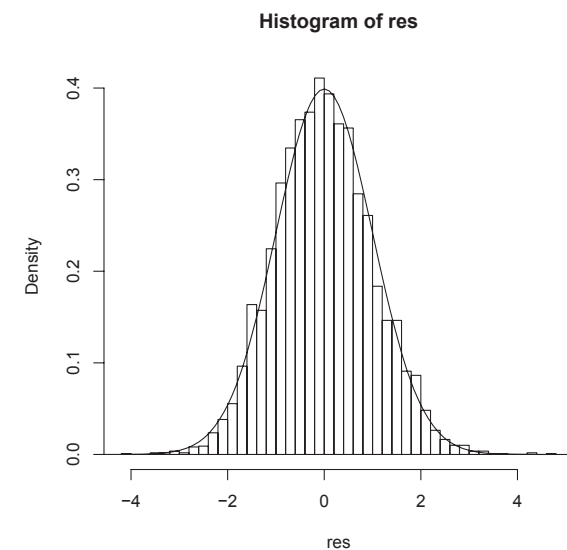
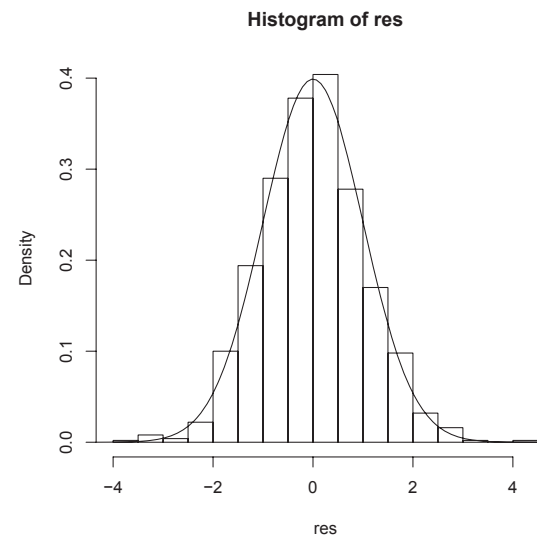
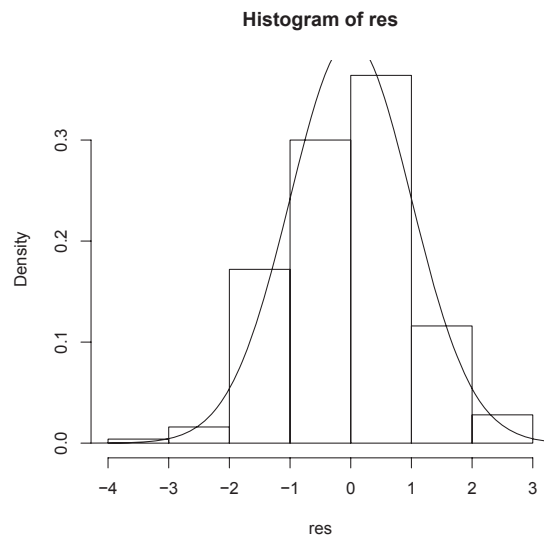
$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i - E(X_i)}{\sqrt{\text{Var}(X_i)}} \right).$$

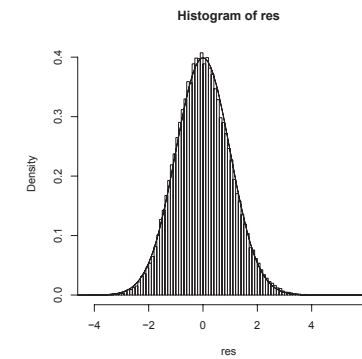
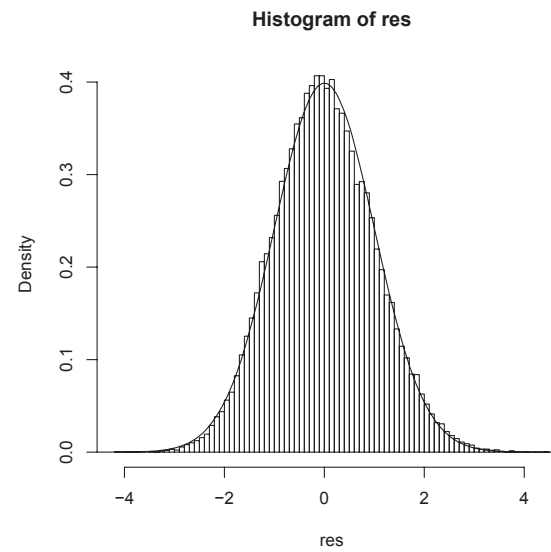
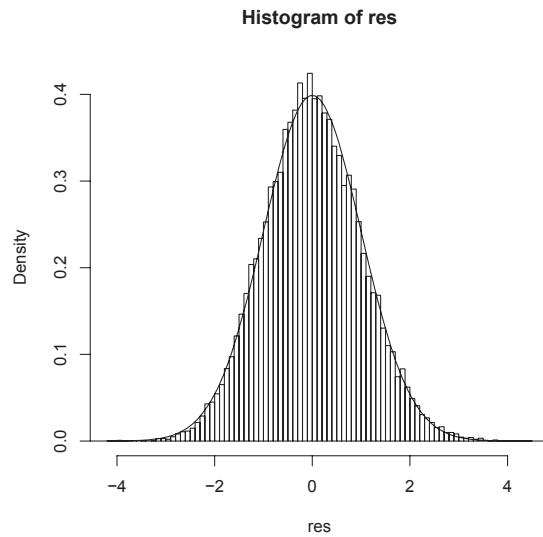
Dann gilt: Z_n ist *asymptotisch standardnormalverteilt*, in Zeichen: $Z_n \stackrel{a}{\sim} \mathcal{N}(0; 1)$, d.h. es gilt für jedes z

$$\lim_{n \rightarrow \infty} \underbrace{P(Z_n \leq z)}_{\substack{\text{Verteilungsfunktion} \\ \text{von } Z_n}} = \underbrace{\Phi(z)}_{\substack{\text{Verteilungsfunktion} \\ \text{der Standardnormalverteilung}}}$$

Für die Eingangsfragen gilt also:

- Ja, wenn man die Variablen geeignet mittelt, standardisiert und mit $\frac{1}{\sqrt{n}}$ reskaliert, dann kann man bei großem n näherungsweise mit der Normalverteilung rechnen. Dabei ist für festes n die Approximation umso besser, je „symmetrischer“ die ursprüngliche Verteilung ist.





Anwendung des zentralen Grenzwertsatz auf \bar{X} :

Gemäß dem Gesetz der großen Zahlen weiß man: $\bar{X}_n \longrightarrow E(X_i)$

Für die Praxis ist es aber zudem wichtig, die konkreten Abweichungen bei großem aber endlichem n zu quantifizieren, etwa zur Beantwortung folgender Fragen:

- Gegeben eine Fehlermarge ε und Stichprobenumfang n : Wie groß ist die Wahrscheinlichkeit, dass \bar{X} höchstens um ε von μ abweicht?
- Gegeben eine Fehlermarge ε und eine „Sicherheitswahrscheinlichkeit“ γ : Wie groß muss man n mindestens wählen, damit mit mindestens Wahrscheinlichkeit γ das Stichprobenmittel höchstens um ε von μ abweicht (*Stichprobenplanung*)?

Aus dem zentralen Grenzwertsatz folgt mit $\mu = E(X_i)$ und $\sigma^2 = \text{Var}(X_i)$:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right) &= \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n} \cdot \sigma} \\ &= \frac{n\bar{X}_n - n\mu}{\sqrt{n} \cdot \sigma} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\approx} \mathcal{N}(0, 1) \end{aligned}$$

oder auch („Standardisierung rüchwärts angewandt“)

$$\bar{X}_n \stackrel{a}{\approx} \mathcal{N} \left(\mu, \frac{\sigma^2}{n} \right).$$

Wichtige Anwendung: Approximation der Binomialverteilung

Sei $X \sim B(n, \pi)$. Kann man die Verteilung von X (jetzt ja nur eine Beobachtung!) approximieren?

Hier hat man zunächst nur ein X . Der zentrale Grenzwertsatz gilt aber für eine Summe vieler Glieder. Idee: Schreibe X als Summe von binären Zufallsvariablen.

X ist die Anzahl der Treffer in einer *i.i.d.* Folge Y_1, \dots, Y_n von Einzelversuchen, wobei

$$Y_i = \begin{cases} 1 & \text{Treffer} \\ 0 & \text{kein Treffer} \end{cases}$$

Die Y_i sind i.i.d. Zufallsvariablen mit $Y_i \sim \text{Bin}(1, \pi)$ und es gilt

$$X = \sum_{i=1}^n Y_i, \quad \mathbb{E}(Y_i) = \pi, \quad \text{Var}(Y_i) = \pi \cdot (1 - \pi).$$

Damit lässt sich der zentrale Grenzwertsatz anwenden:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i - \mathbf{E}(Y_i)}{\sqrt{\text{Var}(Y_i)}} \right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{Y_i - \pi}{\sqrt{\pi(1 - \pi)}} \right) = \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n Y_i - n \cdot \pi}{\sqrt{\pi(1 - \pi)}} \\ &= \frac{\sum_{i=1}^n Y_i - n \cdot \pi}{\sqrt{n \cdot \pi(1 - \pi)}} \stackrel{a}{\sim} \mathcal{N}(0, 1) \end{aligned}$$

und damit

$$\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}} \stackrel{a}{\approx} \mathcal{N}(0, 1)$$

so dass

$$P(X \leq x) \approx \Phi \left(\frac{x - n \cdot \pi}{\sqrt{n \cdot \pi(1 - \pi)}} \right)$$

falls n groß genug ist.

Man beachte: Ob eine Approximation gut genug ist, ist eine *inhaltliche* Entscheidung. Trotzdem werden oft Faustregeln angegeben, ab wann diese Approximation als gut gelte, z.B.

$$n \cdot \pi \geq 5 \quad \text{und} \quad n \cdot (1 - \pi) \geq 5$$

$$n \cdot \pi(1 - \pi) \geq 9$$

Stetigkeitskorrektur: Durch die Approximation der *diskreten* Binomialverteilung durch die *stetige* Normalverteilung geht der diskrete Charakter verloren. Man erhält als Approximation $P(X = x) \approx 0$ für jedes $x \in \mathbb{N}$, was gerade für mittleres n unerwünscht ist.

Benutze deshalb

$$P(X \leq x) = P(X \leq x + 0.5)$$

bei ganzzahligem $x \in \mathbb{N}$.

Man erhält als bessere Approximation

$$P(X \leq x) \approx \Phi \left(\frac{x + 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}} \right)$$

und damit

$$\begin{aligned} P(X = x) &= P(X \leq x + 0.5) - P(X \leq x - 0.5) \\ &\approx \Phi \left(\frac{x + 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}} \right) - \Phi \left(\frac{x - 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}} \right) \end{aligned}$$

Bsp. 1.84. *Fiktives Beispiel*

Ein Politiker ist von einer gewissen, umstrittenen Maßnahme in seiner Partei überzeugt und überlegt, ob es taktisch geschickt ist, zur Unterstützung der Argumentation eine Mitgliederbefragung zu dem Thema durchzuführen. Er wählt als "Probelauf" 200 Mitglieder zufällig aus und beschließt, eine Mitgliederbefragung zu „riskieren“, falls er in der Stichprobe mindestens 52% Zustimmung erhält.

Wie groß ist die Wahrscheinlichkeit, in der Stichprobe mindestens 52% Zustimmung zu erhalten, obwohl der wahre Anteil nur 48% beträgt?

1.8 Mehrdimensionale Zufallsvariablen

Im Folgenden Beschränkung auf den diskreten Fall und zweidimensionale Zufallsvariablen.

„Schnelldurchgang unter Bezug auf das Eindimensionale, Statistik I und auf die Tatsache, dass $X = x_i$ und $Y = y_j$ Ereignisse sind“

Das Hauptinteresse gilt (entsprechend der Kontingenztafel in Statistik I) der gemeinsamen Verteilung

$$P(\{X = x_i\} \cap \{Y = y_j\})$$

Definition 1.85.

Betrachtet werden zwei eindimensionale diskrete Zufallselemente X und Y (zu demselben Zufallsexperiment, also über denselben Grundraum). Die Wahrscheinlichkeit

$$P(X = x_i, Y = y_j) := P(\{X = x_i\} \cap \{Y = y_j\})$$

in Abhängigkeit von x_i und y_j heißt *gemeinsame Verteilung* der mehrdimensionalen Zufallsvariable $\begin{pmatrix} X \\ Y \end{pmatrix}$ bzw. der Variablen X und Y .

Randwahrscheinlichkeiten:

$$p_{i\bullet} = P(X = x_i) = \sum_{j=1}^m P(X = x_i, Y = y_j)$$

$$p_{\bullet j} = P(Y = y_j) = \sum_{i=1}^k P(X = x_i, Y = y_j)$$

Bedingte Verteilungen:

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}$$

Stetiger Fall (nicht klausurrelevant): Zufallsvariable mit zweidimensionaler Dichtefunktion $f(x, y)$:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \left(\int_c^d f(x, y) dy \right) dx$$

Definition 1.86.

Seien X und Y zwei Zufallsvariablen. Dann heißt

$$\sigma_{X,Y} := \text{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y)))$$

Kovarianz von X und Y .

Rechenregeln:

- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X) \cdot \mathbf{E}(Y)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- Mit $\tilde{X} = a_X X + b_X$ und $\tilde{Y} = a_Y Y + b_Y$ ist

$$\text{Cov}(\tilde{X}, \tilde{Y}) = a_X \cdot a_Y \cdot \text{Cov}(X, Y)$$

- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y)$

Definition 1.87.

Zwei Zufallsvariablen X und Y mit $\text{Cov}(X, Y) = 0$ heißen *unkorreliert*.

Satz 1.88.

Stochastisch unabhängige Zufallsvariablen sind unkorreliert. Die Umkehrung gilt jedoch im allgemeinen nicht.

Definition 1.89.

Gegeben seien zwei Zufallsvariablen X und Y . Dann heißt

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}$$

Korrelationskoeffizient von X und Y .

Eigenschaften des Korrelationskoeffizienten:

- Mit $\tilde{X} = a_X X + b_X$ und $\tilde{Y} = a_Y Y + b_Y$ ist

$$|\rho(\tilde{X}, \tilde{Y})| = |\rho(X, Y)|.$$

- $-1 \leq \rho(X, Y) \leq 1$.
- $|\rho(X, Y)| = 1 \iff Y = aX + b$
- Sind $\text{Var}(X) > 0$ und $\text{Var}(Y) > 0$, so gilt $\rho(X, Y) = 0$ genau dann, wenn $\text{Cov}(X, Y) = 0$.

Bsp. 1.90. [Chuckk-a-Luck:]

X_1 Gewinn, wenn beim ersten Wurf ein Einsatz auf 1 gesetzt wird.

X_6 Gewinn, wenn beim ersten Wurf ein Einsatz auf 6 gesetzt wird.

Kovarianz zwischen X_1 und X_6 :

$$\text{Cov}(X_1, X_6) = \mathbb{E}(X_1 \cdot X_6) - \mathbb{E}(X_1) \cdot \mathbb{E}(X_6)$$

Zur Berechnung von $\mathbb{E}(X_1 \cdot X_6)$ Hilfsvariable Z denken: $Z = X_1 \cdot X_6$;

dann: $\mathbb{E}(X_1 \cdot X_6) = \mathbb{E}(Z) = \sum_{z \in \mathcal{Z}} z \cdot P(Z = z)$

(x_1, x_6)	$P(X_1 = x_1, X_6 = x_6)$	$x_1 \cdot x_6$	(x_1, x_6)	$P(X_1 = x_1, X_6 = x_6)$	$x_1 \cdot x_6$
$(-1, -1)$	$\frac{64}{216}$	1	$(-1, 3)$	$\frac{1}{216}$	-3
$(-1, 1)$	$\frac{48}{216}$	-1	$(3, -1)$	$\frac{1}{216}$	-3
$(1, -1)$	$\frac{48}{216}$	-1	$(1, 1)$	$\frac{24}{216}$	1
$(-1, 2)$	$\frac{12}{216}$	-2	$(1, 2)$	$\frac{3}{216}$	2
$(2, -1)$	$\frac{12}{216}$	-2	$(1, 2)$	$\frac{3}{216}$	2

z	$P(Z = z) \cdot 216$
-3	$1 + 1 = 2$
-2	$12 + 12 = 24$
-1	$48 + 48 = 96$
1	$64 + 24 = 88$
2	$3 + 3 = 6$

$$E(Z) = \sum_{z \in \mathcal{Z}} z \cdot P(Z = z),$$

$$\begin{aligned} \text{also } E(X_1 \cdot X_6) &= ((-3) \cdot 2 + (-2) \cdot 24 + (-1) \cdot 96 + 1 \cdot 88 + 2 \cdot 6) \cdot \frac{1}{216} \\ &= \frac{-50}{216} = -0.23148 \end{aligned}$$

$$\begin{aligned} \Rightarrow \text{Cov}(X_1, X_6) &= E(X_1 \cdot X_6) - E(X_1) \cdot E(X_6) \\ &= -0.23148 - (-0.0787) \cdot (-0.0787) = -0.23768 \end{aligned}$$

X_1 und X_6 sind negativ korreliert.

2 Induktive Statistik

2.1 Grundprinzipien der induktiven Statistik

Ziel: Inferenzschluss, Repräsentationsschluss: Schluss von einer Stichprobe auf Eigenschaften der Grundgesamtheit, aus der sie stammt.

- Von Interesse sei ein Merkmal X in der Grundgesamtheit \mathcal{G} .
- Ziehe eine Stichprobe (g_1, \dots, g_n) von Elementen aus \mathcal{G} und werte X jeweils aus.
- Man erhält Werte x_1, \dots, x_n . Diese sind Realisationen der i.i.d Zufallsvariablen oder Zufallselemente X_1, \dots, X_n , wobei die Wahrscheinlichkeitsverteilung der X_1, \dots, X_n genau die Häufigkeitsverhältnisse in der Grundgesamtheit widerspiegelt (vgl. Bem. 1.47).

Die Frage lautet also: wie kommt man von Realisationen x_1, \dots, x_n von i.i.d. Zufallsvariablen X_1, \dots, X_n auf die Verteilung der X_i ?

- Dazu nimmt man häufig an, man kenne den Grundtyp der Verteilung der X_1, \dots, X_n . Unbekannt seien nur einzelne Parameter davon (vgl. Kap. 1.6).

Beispiel: X_i sei normalverteilt, unbekannt seien nur μ, σ^2 .

⇒ *parametrische Verteilungsannahme* (meist im Folgenden)

- Alternativ: Verteilungstyp nicht oder nur schwach festgelegt (z.B. symmetrische Verteilung)

⇒ *nichtparametrische Modelle* (“*verteilungsfreie Verfahren*“) (hier kaum behandelt)

- Klarerweise gilt im Allgemeinen (generelles Problem bei der Modellierung): Parametrische Modelle liefern schärfere Aussagen – wenn ihre Annahmen zutreffen. Wenn ihre Annahmen nicht zutreffen, dann existiert die große Gefahr von Fehlschlüssen.

Wichtige Fragestellungen der induktiven Statistik: Trennung mittels der Auswertung einer

- Zufallsstichprobe
- möglichst gute ★ ★ ★
- Aussagen ★★
- über bestimmte Charakteristika ★
- der Grundgesamtheit.

★ Welche Charakteristika sind für die Fragestellung relevant? Natürlich werden für die Inferenz bezüglich des Erwartungswerts andere Methoden als für Schlüsse über die Varianz benötigt.

★★ verschiedene Formen:

- Punktschätzung: z.B. wahrer Anteil 0.4751
- Intervallschätzung: z.B. wahrer Anteil liegt zwischen 0.46 und 0.48
- Hypothesentest: Die Annahme, der Anteil liegt höchstens bei 50% kann nicht aufrecht erhalten werden

★★★ Was heißt gut?

- Festlegung von „Gütekriterien“ (Genauigkeit? Wahrscheinlichkeit eines Fehlers gering?)
- Wie konstruiert man ein gutes/optimales Verfahren?
- Sicherheitsstellung der „Objektivität der statistischen Analyse“. Jeder wendet das beste Verfahren an \Rightarrow gleiche Auswertung

Wichtig: Festgelegt werden *vor* dem Ziehen der Stichprobe (bzw. vor dem Bekanntwerden der Daten) *Auswertungsverfahren*, also Zufallsvariablen. Ihre Eigenschaften werden mittels der Wahrscheinlichkeitsrechnung ermittelt.

Man beachte, dass die ganze Argumentation auf der Zufälligkeit der Stichprobenziehung aufbaut. **Methoden der statistischen Inferenz sind also nicht geeignet für nicht zufällige Auswahlen und auch nicht für Vollerhebungen.** Bei letzteren sind sie nicht notwendig, da man ja hier die Grundgesamtheit kennt. Bei nicht zufälligen Auswahlen greift die Grundidee, den nicht ausschließbaren Induktionsfehler durch die Wahrscheinlichkeitsrechnung zu kontrollieren, nicht. Die entsprechenden Schlüsse weisen also einen unkontrollierten Fehler auf. Nicht zufällige Auswahlen entstehen z.B. durch Auswahl auf das Gerätewohl (z.B. im Internet; Fenster poppt auf: „Haben Sie Zeit, uns ein paar Fragen zu beantworten?“). Problematisch in diesem Kontext sind auch Untersuchungen, bei denen die Teilnahmeverweigerung von Personen nicht zufällig ist, sondern mit inhaltlich interessierenden Merkmalen der Verweigerenden zusammenhängt. Hier sind mindestens aufwändige, modellbasierte Korrekturverfahren nötig; oft ist aber auch eine auf absolut präzise Ergebnisse zielende induktive Verallgemeinerung der Ergebnisse der Auswertung schlicht nicht zulässig.