

Vorlesung: Statistik II für Wirtschaftswissenschaft

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

Sommersemester 2017



- Einführung
- 1 Wahrscheinlichkeit: Definition und Interpretation
- 2 Elementare Wahrscheinlichkeitsrechnung
- 3 Zufallsgrößen
- 4 Spezielle Zufallsgrößen**
 - stetige Verteilungen
- 5 Mehrdimensionale Zufallsvariablen
- 6 Grenzwertsätze

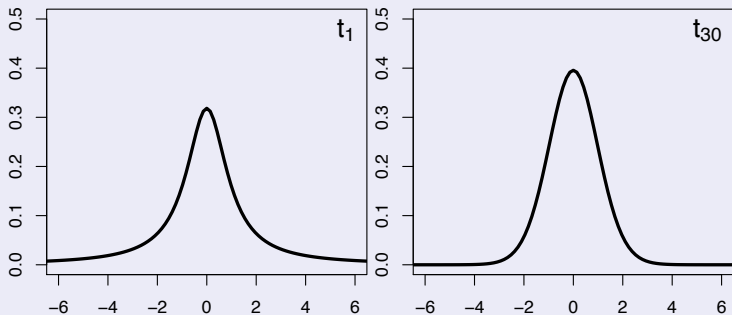
Definition t-Verteilung

Seien X und Y_1, \dots, Y_n unabhängige Zufallsvariablen mit $X \sim N(0, 1)$ und $Y_i \sim N(0, 1)$. Dann ist der Quotient

$$\frac{X}{\sqrt{\sum_{i=1}^n Y_i^2 / n}} \sim t_n$$

t -verteilt mit n Freiheitsgraden.

Beispiele der Dichtefunktion

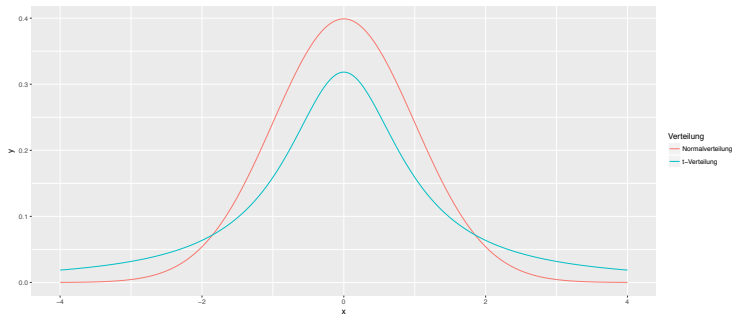


Anwendungen bei Finanzdaten

- Häufig wird die Normalverteilung für die Verteilung von Renditen genutzt
- Problematisch, da die Wahrscheinlichkeit von extremen Ausreißern (Crash) unterschätzt wird
- Abhilfe: Verwende Verteilungen mit **heavy tails** z.B. die t-Verteilung



Vergleich t-Verteilung und Normalverteilung





- Einführung
- ① Wahrscheinlichkeit: Definition und Interpretation
- ② Elementare Wahrscheinlichkeitsrechnung
- ③ Zufallsgrößen
- ④ Spezielle Zufallsgrößen
- ⑤ Mehrdimensionale Zufallsvariablen**
- ⑥ Grenzwertsätze

Mehrdimensionale Zufallsvariablen

Analog zu den Maßzahlen und Überlegungen aus der deskriptiven Statistik:

$$\begin{pmatrix} X \\ Y \end{pmatrix}$$

also z.B. $\omega \in \Omega$, zufällig gezogene Person und damit $X(\omega)$ und $Y(\omega)$
Auswertung der Merkmale jeweils an *derselben* Person.

⇒ zweidimensionale Zufallsvariable $\begin{pmatrix} X \\ Y \end{pmatrix}$ (wie bei Zusammenhangsanalyse in Statistik I)

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

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

Zweidimensionale Verteilungen

Betrachtet werden zwei eindimensionale diskrete Zufallselemente X und Y (zu demselben Zufallsexperiment). 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:

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

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

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

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

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

Seien X und Y zwei Zufallsvariablen. Dann heißt

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

Kovarianz von X und Y .

- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- 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

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

- Stochastisch unabhängige Zufallsvariablen sind unkorreliert. Die Umkehrung gilt jedoch im allgemeinen nicht.
- Vergleiche Statistik I: Kovarianz misst nur lineare Zusammenhänge.

Definition

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 $\text{Cov}(X, Y) = 0$.



- Einführung
- ① Wahrscheinlichkeit: Definition und Interpretation
- ② Elementare Wahrscheinlichkeitsrechnung
- ③ Zufallsgrößen
- ④ Spezielle Zufallsgrößen
- ⑤ Mehrdimensionale Zufallsvariablen
- ⑥ Grenzwertsätze**

Big Data: Beobachtung von *großen* Datensätzen

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

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 \tilde{X} bei einer einfachen Zufallsauswahl sind.

Beispiel:

\tilde{X} Einkommen, n Personen zufällig ausgewählt

X_1	Einkommen der	ersten	zufällig ausgewählten Person
X_2	Einkommen der	zweiten	zufällig ausgewählten Person
\vdots		\vdots	
X_n	Einkommen der	n -ten	zufällig ausgewählten Person

Stichprobenvariable

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

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

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

Erwartungswert und Varianz von \bar{X}_n

X_1, X_2, \dots, X_n seien unabhängig und identisch verteilt.

$$X_1, X_2, \dots, X_n \quad i.i.d.$$

Ist $\mathbb{E}(X_i) = \mu$ und $\text{Var}(X_i) = \sigma^2$, so gilt:

$$\begin{aligned}\mathbb{E}(X_1 + X_2 + \dots + X_n) &= n\mu \\ \text{Var}(X_1 + X_2 + \dots + X_n) &= n\sigma^2 \\ \mathbb{E}\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) &= \mu \\ \text{Var}\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) &= \frac{\sigma^2}{n}\end{aligned}$$

Diese Eigenschaften bilden die Grundlage für die folgenden Sätze.

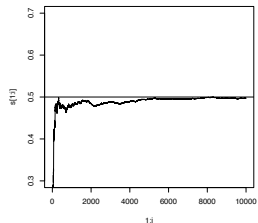
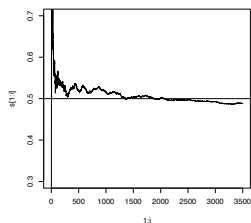
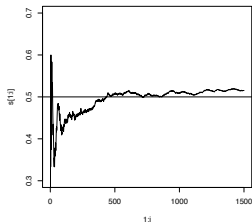
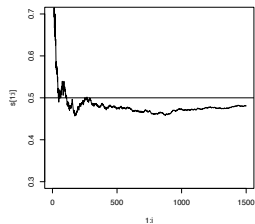
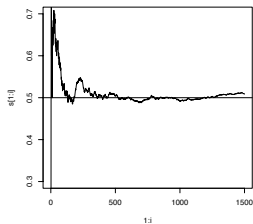
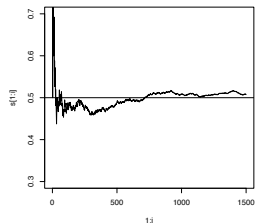
Das schwache Gesetz der großen Zahlen

Betrachte für wachsenden Stichprobenumfang n :

- X_1, \dots, X_n i.i.d.
- $X_i \in \{0, 1\}$ binäre Variablen mit $\pi = P(X_i = 1)$
Beispiele: Pro/Contra, Kopf/Zahl, A tritt ein/ A tritt nicht ein
- $H_n =$ die relative Häufigkeit der Einsen in den ersten n Versuchen.



Simulationen



Beobachtungen

- 1 Am Anfang sehr unterschiedlicher, unregelmäßiger Verlauf der Pfade.
- 2 Mit wachsendem n pendeln sich die Pfade immer stärker um π herum ein, d.h. mit wachsendem Stichprobenumfang konvergiert die relative Häufigkeiten eines Ereignisses gegen seine Wahrscheinlichkeit.
- 3 Formalisierung von 2.: Legt man sehr kleine Korridore/Intervalle um π , so ist bei sehr großem n der Wert von H_n fast sicher in diesem Korridor.

Das Ereignis „Die relative Häufigkeit H_n liegt im Intervall der Breite 2ε um π ,“ lässt sich schreiben als:

$$\begin{aligned}\pi - \varepsilon &\leq H_n \leq \pi + \varepsilon \\ -\varepsilon &\leq H_n - \pi \leq \varepsilon \\ |H_n - \pi| &\leq \varepsilon\end{aligned}$$



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:

$P(A)$, die Wahrscheinlichkeit eines Ereignisses A , kann man sich vorstellen als Grenzwert der relativen Häufigkeit des Eintretens von A in einer unendlichen Versuchsreihe identischer Wiederholungen eines Zufallsexperiments.

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

Sei z.B. π der (unbekannte) Anteil der SPD Wähler, so ist die relative Häufigkeit in der Stichprobe eine „gute Schätzung für π “. Je größer die Stichprobe ist, umso größer ist die Wahrscheinlichkeit, dass die relative Häufigkeit sehr nahe beim wahren Anteil π ist.

Gesetz der großen Zahl (allgemein)

Das Ergebnis lässt sich verallgemeinern auf Mittelwerte beliebiger Zufallsvariablen:

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“, „ X_n konvergiert in Wahrscheinlichkeit gegen μ “.)

- **Interpretation des Erwartungswerts:** μ kann in der Tat interpretiert werden als Durchschnittswert in einer unendlichen Folge von Wiederholungen des Zufallsexperiments.
- **Spiele.** Wenn ein Spiel mit negativem Erwartungswert häufig gespielt wird, verliert man mit sehr hoher Wahrscheinlichkeit (Grund für Rentabilität von Spielbanken und Wettbüros)

Die Verteilungsfunktion

Jetzt betrachten wir die empirische Verteilungsfunktion: In jedem Punkt x ist $F_n(x)$ vor der Stichprobe eine Zufallsvariable, also ist F_n eine zufällige Funktion

Wie vergleicht man die zufällige Funktion $F_n(x)$ mit der Funktion $F(x)$?
Der Abstand hängt ja von dem Punkt x ab, in dem gemessen wird!

Idee: Maximaler Abstand

$$\max_{x \in \mathbb{R}} |F_n^{X_1, \dots, X_n}(x) - F(x)|$$

Existiert nicht immer; formal muss man das sogenannte Supremum betrachten.



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

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

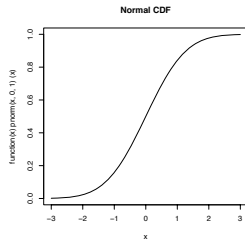
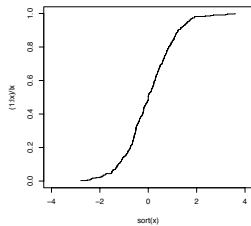
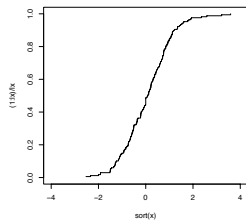
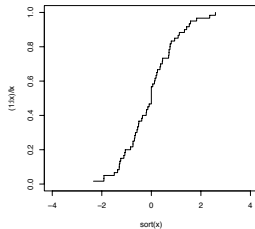
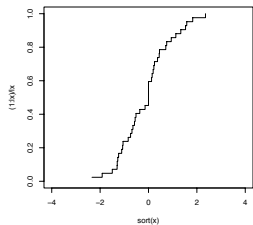
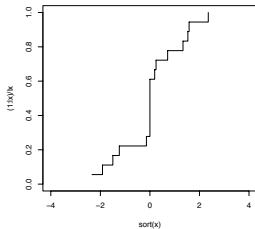
gilt für jedes $c > 0$

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



- „Erträglichkeitsschranke“ c vorgegeben. Wsk, dass maximaler Abstand größer c ist geht für hinreichend großes n gegen $0 \implies$ überall kleiner Abstand. Man kann $\{D_n > c\}$ interpretieren als „Die Stichprobe führt den Betrachter hinter das Licht.“. Dann ist also die Wahrscheinlichkeit mit hinreichend großem n praktisch null.
- Anschaulich: Praktisch sicher spiegelt die empirische Verteilungsfunktion einer unendlichen Stichprobe die wahre Verteilungsfunktion wider.
- Falls die Stichprobe groß genug ist, so wird letztendlich immer repräsentativ für die Grundgesamtheit, d.h. man kann Verteilungsgesetzmäßigkeiten durch Beobachtungen erlernen (grundlegend für die Statistik) \rightarrow „Hauptsatz“.

Beispiele



Der zentrale Grenzwertsatz I

- Gibt es für große Stichprobenumfänge Regelmäßigkeiten im Verteilungstyp?
- Gibt es eine Standardverteilung, mit der man oft bei großen empirischen Untersuchungen rechnen kann?



Der zentrale Grenzwertsatz II

Seien X_1, \dots, X_n i.i.d. mit $\mathbb{E}(X_i) = \mu$ und $\text{Var}(X_i) = \sigma^2 > 0$ sowie

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

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

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z).$$

Für die Eingangsfragen gilt also:

Ja, wenn man die Variablen geeignet mittelt und standardisiert, 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.

Standardisieren

Die Funktion kommt durch *Standardisieren* und durch *geeignetes mitteln* zustande.

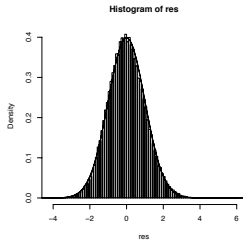
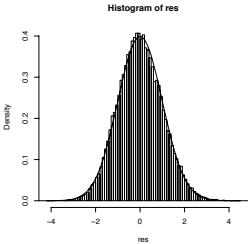
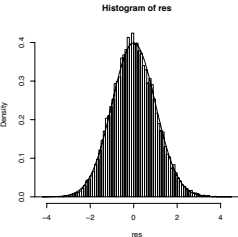
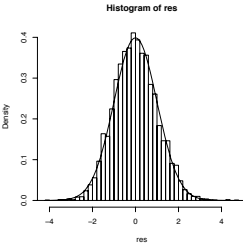
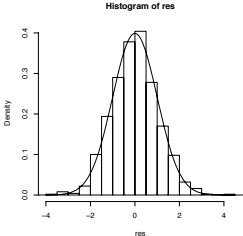
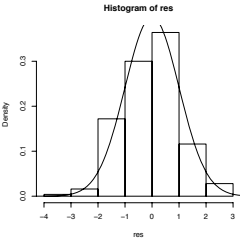
Dabei ist es wichtig, durch \sqrt{n} (und nicht durch n) zu teilen.

$$\sum X_i \quad \longrightarrow \text{verliert sich; } \text{Var}(\sum X_i) \rightarrow \infty$$

$$\frac{1}{n} \sum x_i \quad \longrightarrow \text{Var} \left(\frac{1}{n} \sum X_i \right) \rightarrow 0$$



Beispiele



Anwendung des zentralen Grenzwertsatz auf \bar{X} I

Gemäß dem Gesetz der großen Zahlen weiß man: $\bar{X}_n \rightarrow \mu$

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*)?

Anwendung des zentralen Grenzwertsatz auf \bar{X} II

Aus dem zentralen Grenzwertsatz folgt:

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

oder auch

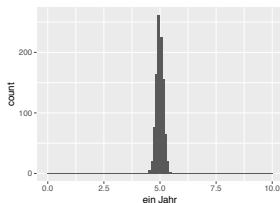
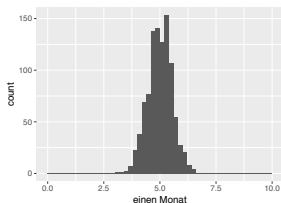
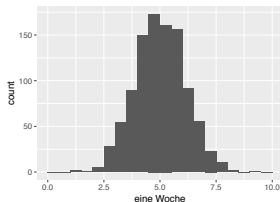
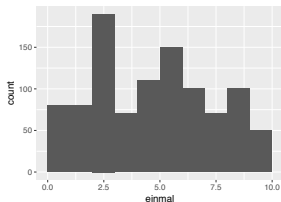
$$\bar{X}_n \stackrel{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right).$$

$\frac{\sigma^2}{n}$ wird mit wachsendem n immer kleiner

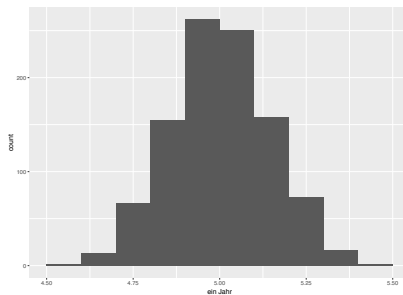
- * Schwankung im richtigen Wert (μ)
- * Ausschläge werden kleiner

Warten auf den Bus

Bestimme Wartezeit, Durchschnittliche Wartezeit in 1 Woche, 1 Monat, 1 Jahr



1 Jahr



Approximation der Binomialverteilung I

Sei $X \sim B(n, \pi)$. Kann man die Verteilung von X 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

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

Derselbe Trick wurde bei der Berechnung von Erwartungswerten angewendet.

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

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

Approximation der Binomialverteilung II

Damit lässt sich der zentrale Grenzwertsatz anwenden:

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

und damit

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

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.

Es gibt verschiedene Faustregeln, ab wann diese Approximation gut ist, z.B.

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

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

Wichtig: Ob die Approximation hinreichend genau ist, hängt insbesondere vom substanzwissenschaftlichen Kontext ab.