Grenzwertsätze und Konvergenzbegriffe

Gegeben: Eine Folge von Zufallsvariablen:

$$X_1, X_2, X_3, \ldots$$

Formaler:

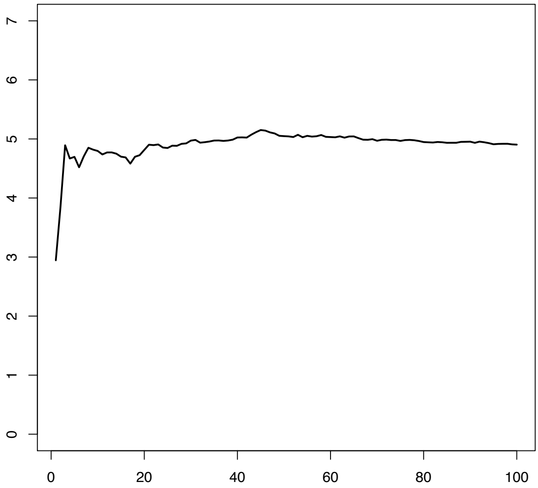
$$X_n, \qquad n = 1, 2, \ldots$$

Betrachte speziell die Folge der arithmetischen Mittelwerte:

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad n = 1, 2, \ldots$$

$$X_1, \quad \frac{X_1 + X_2}{2}, \quad \frac{X_1 + X_2 + X_3}{3}, \ldots$$

Computerexperiment I:



Gesetz der großen Zahl

Beobachtung:

- Die Folge der $\overline{X}_n$ nähert sich einem festen Wert an.
- Welcher Wert ist das?
- Wie kann man diese 'Konvergenz' beschreiben?

Fundamentales Resultat 1: Gesetz der großen Zahl

Computerexperiment II: Verteilung von $\overline{X}_n^*$

Simuliere auf dem Computer eine Zufallsstichprobe vom Umfang n

$$X_1, \ldots, X_n \sim F,$$

F: vorgegeben, und berechne $\overline{X}_n$. Beispiel in R:

```
x = rexp(100)
res = mean(x)
```

Wiederhole dies $S = 10000$-mal, um eine Stichprobe von $\overline{X}_n$-Werten zu erhalten. Führe dies für verschiedene Verteilungen durch.

Um Histogramme von verschiedenen Simulationen besser vergleichen zu können, standardisieren wir $\overline{X}_n$. D.h.: Berechne statt $\overline{X}_n$

$$\overline{X}_n^* = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

wobei μ der Erwartungswert und σ die Standardabweichung der gewählten Verteilung ist.

Beispielprogramme

R/Splus: Simuliere 10000 $\overline{X}_{100}$-Werte für $X_i \sim U[0,1]$.

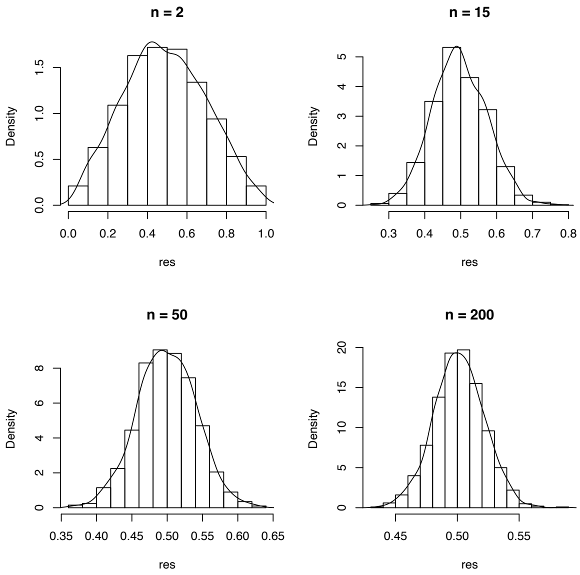
```
res = numeric(10000)
for ( i in (1:10000) ) {
 x = runif(100)
 res[i] = mean(x)
}
hist(res)
```

Matlab/Scilab:

```
res = zeros(10000,1);
for i = 1:10000, x = rand(10,1,'exp');
 res(i)= mean(x);
end;
histplot( 8, res );
```

Frage: Wie muss das Programm modifiziert werden, um Histogramme für die standardisierten Mittelwerte zu erhalten?

Computerexperiment 1: Beobachtungen normalverteilt



Computerexperiment 1: Beobachtungen normalverteilt

Erklärung:

Wir wissen, dass Linearkombinationen

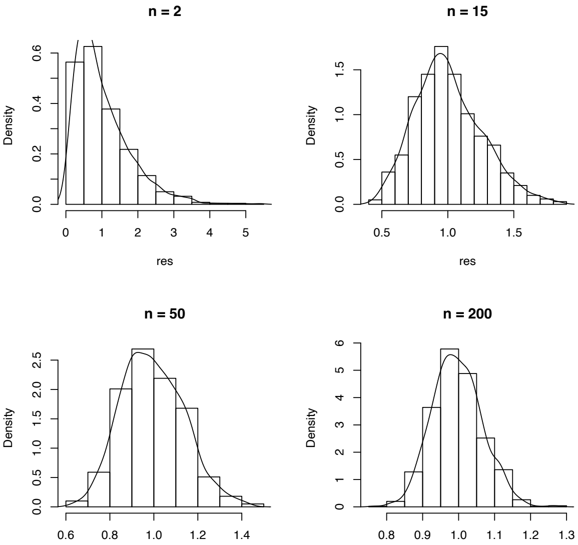
$$a_1 X_1 + \cdots + a_k X_k$$

von normalverteilten Zufallsvariablen wieder normalverteilt sind. Daher ist auch $\overline{X}_n$ normalverteilt.

Das Computerexperiment ist also im Einklang mit der Theorie, wenn die X_i normalverteilt sind.

Frage: Was passiert, wenn die X_i anderen Verteilungen folgen?

Computerexperiment 2: Beobachtungen nicht normalverteilt



Asymptotische Verteilung

Beobachtung:

- Für kleine n ist die Verteilung sehr schief und nicht durch eine Normalverteilungsdichte approximierbar.
- Für großes n scheint die Verteilung von $\overline{X}_n$ der Normalverteilung sehr ähnlich zu sein,
- und zwar auch dann, wenn die einzelnen X_i nicht normalverteilt sind!
- Gilt das tatsächlich?
- Wie kann man diese Form der 'Konvergenz' beschreiben?

Fundamentales Resultat 2: Zentraler Grenzwertsatz

Aber zunächst zum ersten fundamentalen Resultat...

Das Gesetz der großen Zahlen

$X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit

$$\mu = E(X_1), \qquad \sigma^2 = \mathrm{Var}(X_1)$$

Sei

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

Frage

Wie groß ist der Fehler, wenn man $\overline{X}_n$ statt μ verwendet?

Das Gesetz der großen Zahlen

Fehler:

$$F_n = |\overline{X}_n - \mu|$$

Toleranz

$$\varepsilon > 0$$

Ereignis:

$$\{F_n > \varepsilon\} = \{|\overline{X}_n - \mu| > \varepsilon\}$$

Wahrscheinlichkeit

$$P(|\overline{X}_n - \mu| > \varepsilon) =?$$

Tschebyschow-Ungleichung

Tschebyschow (Chebychev, Tschebyscheff, Čebyšëv)-Ungleichung

$X_1, \dots, X_n$ i.i.d. mit Varianz $\sigma^2 \in (0, \infty)$ und Erwartungswert μ.
Dann gilt:

$$P(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Diskussion der oberen Schranke:

1. Schranke umso besser, je kleiner σ^2.
2. Schranke umso besser, je größer n.
3. Die Verteilung der X_i geht nur über σ^2 ein!

Das Gesetz der großen Zahlen

Schwaches Gesetz der großen Zahlen

$X_1, \ldots, X_n$ i.i.d. mit Erwartungswert μ und Varianz σ^2, $\sigma^2 \in (0, \infty)$.

Dann konvergiert das arithmetische Mittel $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ im stochastischen Sinne gegen den Erwartungswert μ, d.h. für jede Toleranzabweichung $\varepsilon > 0$ gilt:

$$P(|\overline{X}_n - \mu| > \varepsilon) \to 0,$$

wenn n gegen ∞ strebt.

Konvergenzbegriff: Stochastische Konvergenz

Allgemein definiert man:

Definition: (i) Eine Folge $X_1, X_2, \dots$ von Zufallsvariablen **konvergiert stochastisch** oder **in Wahrscheinlichkeit** gegen die Zufallsvariable X, wenn für alle $\varepsilon > 0$ gilt:

$$P(|X_n - X| > \varepsilon) \to 0, \qquad n \to \infty.$$

Notation: $X_n \stackrel{P}{\to} X$, $n \to \infty$.

(ii) Eine Folge $X_1, X_2, \dots$ von Zufallsvariablen **konvergiert stochastisch** oder **in Wahrscheinlichkeit** gegen die Konstante a, wenn für alle $\varepsilon > 0$ gilt:

$$P(|X_n - a| > \varepsilon) \to 0, \qquad n \to \infty.$$

Konvergenzbegriff: Stochastische Konvergenz

Rechenregeln: $X_1, X_2, \cdots$ und $Y_1, Y_2, \cdots$ seien Folgen von ZVen.

(i) Aus $X_n \overset{P}{\to} a$, $n \to \infty$ und $Y_n \overset{P}{\to} b$, $n \to \infty$ folgt für $\lambda, \mu \in \mathbb{R}$:

$$\lambda \cdot X_n \pm \mu \cdot Y_n \overset{P}{\to} \lambda \cdot a \pm \mu \cdot b, \qquad n \to \infty.$$

(ii) Aus $X_n \overset{P}{\to} X$, $n \to \infty$, und $Y_n \overset{P}{\to} b$, $n \to \infty$ folgt

$$X_n \cdot Y_n \overset{P}{\to} X \cdot b, \qquad n \to \infty.$$

und, falls $b \neq 0$ und ein $n_0 \in \mathbb{N}$ existiert mit $P(Y_n \neq 0)$ für $n > n_0$,

$$\frac{X_n}{Y_n} \overset{P}{\to} \frac{X}{b}, \qquad n \to \infty$$

(iii) Aus $X_n \overset{P}{\to} X$, $n \to \infty$, und $Y_n \overset{P}{\to} Y$, $n \to \infty$, folgt für jede stetige Funktion f: $f(X_n) \overset{P}{\to} f(X), \qquad n \to \infty.$

(iv) Aus $X_n \overset{P}{\to} X$, $n \to \infty$, und $Y_n \overset{P}{\to} Y$, $n \to \infty$, folgt für jede stetige Funktion $f : \mathbb{R}^2 \to \mathbb{R}$: Falls $f(X, Y)$ und $f(X_n, Y_n)$ definiert sind für alle $n \in \mathbb{N}$, dann gilt: $f(X_n, Y_n) \overset{P}{\to} f(X, Y), n \to \infty.$

Beispiele: Gesetz der großen Zahlen/stochastische Konvergenz

Beispiele:

(i) $\overline{X}_n$ und $\overline{Y}_n$ seien die arithmetischen Mittelwerte aus zwei i.i.d-Stichproben $X_1, X_2, \cdots$ und $Y_1, Y_2, \cdots$ mit Erwartungswerten μ_X und $\mu_Y > 0$. Dann gilt $\overline{X}_n / \overline{Y}_n \stackrel{P}{\to} \mu_X / \mu_Y, \qquad n \to \infty.$

(ii) Um die Laufzeiten ℓ_x und ℓ_y von zwei nacheinander geschalteten Algorithmen abzuschätzen, werden die Algorithmen jeweils n mal unabhängig voneinander unter identischen Bedingungen gestartet, so dass die einzelnen Laufzeiten $X_1, \cdots, X_n$ und $Y_1, \cdots, Y_n$ als einfache Stichproben mit Erwartungswerten ℓ_x und ℓ_y angesehen werden können. Man approximiert nun ℓ_x durch $\overline{X}_n$ und ℓ_y durch $\overline{Y}_n$. Da nach dem schwachen Gesetz großer Zahlen $\overline{X}_n \stackrel{P}{\to} \ell_x$ und $\overline{Y}_n \stackrel{P}{\to} \ell_y$, wenn $n \to \infty$, folgt nach Rechenregel (i):

$$\overline{X}_n + \overline{Y}_n \stackrel{P}{\to} \ell_x + \ell_y, \qquad n \to \infty.$$

Das Gesetz der großen Zahlen

Fixiere einen Ausgang $\omega \in \Omega$.

- Die Realisationen

$$\overline{x}_1 = \overline{X}_1(\omega),\ \overline{x}_2 = \overline{X}_2(\omega), \dots$$

  sind eine reelle Zahlenfolge.

- In Abhängigkeit von ω gilt

$$\lim_{n \to \infty} \overline{X}_n(\omega) = \mu$$

  oder nicht.

- Das starke Gesetz macht eine Aussage über die Menge der ω, für die Konvergenz vorliegt:

$$\left\{ \lim_{n \to \infty} \overline{X}_n(\omega) = \mu \right\}$$

Das Gesetz der großen Zahlen

Starkes Gesetz der großen Zahlen

$X_1, \dots, X_n$ i.i.d. mit $E|X_1| < \infty$ und Erwartungswert μ.

Dann konvergiert das arithmetische Mittel mit Wahrscheinlichkeit 1 gegen μ, d.h.

$$P(\overline{X}_n \to \mu) = P(\{\omega | \overline{X}_n(\omega) \text{ konvergiert gegen } \mu\}) = 1.$$

Diskussion:

1. Schwächere Voraussetzungen: Varianz muss nicht existieren.
2. Aussage stärker: Konvergenz mit Wkeit 1.

Hauptsatz der Statistik

1. Der *Hauptsatz der Statistik* macht eine Aussage über die Konvergenz der empirischen Verteilungsfunktion.
2. Dieses Ergebnis ist fundamental, da sehr viele Funktion von Stichprobenvariablen $X_1, \ldots, X_n$ hierüber ausgedrückt werden können.
3. Also: Der Hauptsatz der Statistik liefert die rigorose Begründung, warum Statistik 'funktioniert'.

Hauptsatz der Statistik

Die Zufallsvariablen $X_1, \ldots, X_n, \ldots$ seien unabhängig und identisch (i.i.d.) nach der Verteilungsfunktion F verteilt.

Dann konvergiert der (maximale) Abstand zwischen der **empirischen Verteilungsfunktion** $F_n(x)$ und der **wahren Verteilungsfunktion** $F(x)$ mit Wahrscheinlichkeit 1 gegen 0:

$$P\left(\lim_{n\to\infty} \max_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0\right) = 1.$$

Konvergenzbegriff: Fast sichere Konvergenz

Allgemein definiert man:
Definition: (i) Eine Folge $X_1, X_2, \ldots$ von Zufallsvariablen **konvergiert fast sicher** oder **mit Wahrscheinlichkeit 1** gegen die Zufallsvariable X, wenn gilt:

$$P(\lim_{n\to\infty} |X_n - X| = 0) = 1$$

Notation: $X_n \stackrel{f.s.}{\to} X$, $n \to \infty$.
(ii) Eine Folge $X_1, X_2, \ldots$ von Zufallsvariablen **konvergiert fast sicher** oder **mit Wahrscheinlichkeit 1** gegen die Konstante a, wenn gilt:

$$P(\lim_{n\to\infty} |X_n - a| = 0) = 1$$

Konvergenzbegriff: Fast sichere Konvergenz

Rechenregeln: $X_1, X_2, \cdots$ und $Y_1, Y_2, \cdots$ seien Folgen von ZVen.

(i) Aus $X_n \stackrel{f.s.}{\to} a$, $n \to \infty$ und $Y_n \stackrel{f.s.}{\to} b$, $n \to \infty$ folgt für $\lambda, \mu \in \mathbb{R}$:

$$\lambda \cdot X_n \pm \mu \cdot Y_n \stackrel{f.s.}{\to} \lambda \cdot a \pm \mu \cdot b, \qquad n \to \infty.$$

(ii) Aus $X_n \stackrel{f.s.}{\to} X$, $n \to \infty$, und $Y_n \stackrel{f.s.}{\to} b$, $n \to \infty$ folgt

$$X_n \cdot Y_n \stackrel{f.s.}{\to} X \cdot b, \qquad n \to \infty.$$

und, falls $b \neq 0$ und ein $n_0 \in \mathbb{N}$ existiert mit $P(Y_n \neq 0) = 1$ für $n > n_0$,

$$\frac{X_n}{Y_n} \stackrel{f.s.}{\to} \frac{X}{b}, \qquad n \to \infty$$

(iii) Aus $X_n \stackrel{f.s.}{\to} X$, $n \to \infty$, und $Y_n \stackrel{f.s.}{\to} Y$, $n \to \infty$, folgt für jede stetige Funktion f: $f(X_n) \stackrel{f.s.}{\to} f(X)$, $\qquad n \to \infty$.

(iv) Aus $X_n \stackrel{f.s.}{\to} X$, $n \to \infty$, und $Y_n \stackrel{f.s.}{\to} Y$, $n \to \infty$, folgt für jede stetige Funktion $f : \mathbb{R}^2 \to \mathbb{R}$: Falls $f(X, Y)$ und $f(X_n, Y_n)$ definiert sind für alle $n \in \mathbb{N}$, dann gilt: $f(X_n, Y_n) \stackrel{f.s.}{\to} f(X, Y), n \to \infty$.

Anwendung: Konvergenz der Stichprobenvarianz

Sind $X_1, \cdots, X_n$ i.i.d. mit Erwartungswert μ und Varianz $\sigma^2 \in (0, \infty)$, so heißt

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \left(\overline{X}_n\right)^2$$

Stichprobenvarianz.[1] Es gilt
$\mathrm{Var}(X_1) = \sigma^2 = E(X_1^2) - \mu^2 \Leftrightarrow E(X_1^2) = \sigma^2 + \mu^2 \in \mathbb{R}$. $X_1^2, \cdots, X_n^2$ erfüllen die Voraussetzungen des starken Gesetzes großer Zahlen:

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \overset{f.s.}{\to} E(X_1^2) = \sigma^2 + \mu^2, \quad n \to \infty.$$

Ferner: $\overline{X}_n \overset{f.s.}{\to} \mu$ und $f(x) = x^2$ stetig $\overset{(iii)}{\Rightarrow} \left(\overline{X}_n\right)^2 = f(\overline{X}_n) \overset{f.s.}{\to} f(\mu) = \mu^2, n \to \infty$.
Rechenregel (i) liefert nun die f.s. Konvergenz der Stichprobenvarianz:

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \left(\overline{X}_n\right)^2 \overset{f.s.}{\to} (\sigma^2 + \mu^2) - \mu^2 = \sigma^2.$$

[1] Die zweite Formel folgt durch Ausmultiplizieren von $(X_i - \overline{X}_n)^2$ und zusammenfassen.

Der zentrale Grenzwertsatz: Vorbereitung

$X_1, X_2, \cdots$ i.i.d.-Zufallsvariablen mit Erwartungswert μ und Varianz $\sigma^2 \in (0, \infty)$.

Gesetz der großen Zahlen:

$$\overline{X}_n \stackrel{f.s.}{\to} \mu, \qquad n \to \infty$$

Äquivalent:

$$\overline{X}_n - \mu \stackrel{f.s.}{\to} 0, \qquad n \to \infty$$

Skaliert man die Abweichung $\overline{X}_n - \mu$ mit $\sqrt{n}$, so 'kollabiert' der Ausdruck nicht mehr gegen 0.

Die mit $\sqrt{n}$ skalierte Abweichung folgt einem <u>universellem Fehlergesetz</u>.

Der zentrale Grenzwertsatz

Die Simulation der Verteilung von $\overline{X}_n^*$ legt folgende Vermutung für Ereignisse A nahe.

Vermutung: In großen Stichproben gilt für Intervalle $[A, B]$:

$$P(\overline{X}_n^* \in [A, B]) \approx P(Z \in [a, b]), \qquad (Z \sim N(0, 1)),$$

und somit:

$$P(\overline{X}_n \in [a, b]) \approx P(Z_n \in [a, b]), \qquad Z_n \sim N(\mu, \sigma^2/n).$$

Zusammenhang: $P\left(\overline{X}_n \in [a, b]\right) = P\left(\overline{X}_n^* \in [\frac{a-\mu}{\sigma/\sqrt{n}}, \frac{b-\mu}{\sigma/\sqrt{n}}]\right)$.

Der zentrale Grenzwertsatz

Zentrale Grenzwertsatz (ZGWS)

$X_1, \ldots, X_n$ i.i.d. mit

$$\mu = E(X_1), \qquad \sigma^2 = \mathrm{Var}(X_1) \in (0, \infty).$$

Dann gilt: $\overline{X}_n$ ist asymptotisch $N(\mu, \sigma^2/n)$-verteilt,

$$\overline{X}_n \sim_{approx} N(\mu, \sigma^2/n),$$

in dem Sinne, dass die Verteilungsfunktion der standardisierten Version gegen die Verteilungsfunktion der $N(0,1)$-Verteilung konvergiert:

$$P\left(\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \leq x\right) \to \Phi(x), \qquad n \to \infty.$$

Der zentrale Grenzwertsatz

Die Aussage des zentralen Grenzwertsatzes bleibt gültig, wenn die in der Praxis meist unbekannte Streuung σ durch die empirische Standardabweichung

$$\widehat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}$$

ersetzt wird.

Szenarioanalyse

Fallgestaltung und Aufgabe:

Ein Autohersteller betrachtet im Rahmen einer Szenarioanalyse die Wahrscheinlichkeit, dass der Gesamtgewinn von n Autohäusern die Benchmark b erreicht.

Hierzu wurden n Autohäuser ausgewählt, die in verschiedenen - aber vergleichbaren - Großstädten unabhängig voneinander operieren und von vergleichbarer Größe sind.

Es ist zu klären, wie die gesuchte Wahrscheinlichkeit - zumindest näherungsweise - ermittelt werden kann.

Modellbildung:

Die zukünftigen zufallsbehafteten Quartalsgewinne $X_1, \dots, X_n$ werden als einfache (i.i.d.-) Zufallsstichprobe aufgefasst mit Erwartungswert μ und Varianz $\sigma^2 \in (0, \infty)$.

Die genaue Verteilung der Quartalsgewinne ist unbekannt. Lediglich der erwartete Quartalsgewinn wird im Rahmen der Szenarien spezifiert. Als relevanter Bereich wird festgelegt:

$$1.2 \leq \mu \leq 1.6$$

ZGWS: Fallgestaltung

Problemformulierung
Bestimme (approximativ) die Wahrscheinlichkeit, dass der Gesamtgewinn

$$G = X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

die Benchmark b übersteigt, also

$$P(G > b) = P\left(\sum_{i=1}^{n} X_i > b\right).$$

Wir werden sehen, dass der ZGWS eine praktikable Lösung ermöglicht, sofern n nicht zu klein ist.

Der zentrale Grenzwertsatz

1. $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F(x)$ mit

$$\mu = E(X_1), \qquad \sigma^2 = \operatorname{Var}(X_1) \in (0, \infty).$$

2. Arithmetisches Mittel:

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

   erfüllt: $E(\overline{X}_n) = \mu$ und $\operatorname{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$.

3. Dann erfüllt die standardisierte Version

$$\overline{X}_n^* = \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} = \sqrt{n} \frac{\overline{X}_n - \mu}{\sigma}$$

   $E(\overline{X}_n^*) = 0$ und $\operatorname{Var}(\overline{X}_n^*) = 1$ für alle $n \in \mathbb{N}$!

ZGWS: Anwendung auf die Fallgestaltung

Der Zentrale Grenzwertsatz liefert die Näherung

$$
\begin{aligned}
P\left(\sum_{i=1}^{n} X_i > b\right) &= P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} > \frac{b - n\mu}{\sqrt{n\sigma^2}}\right) \\
&\approx 1 - \Phi\left(\frac{b - n\mu}{\sqrt{n}\sigma}\right) \\
&\approx 1 - \Phi\left(\frac{b - n\mu}{\sqrt{n}\widehat{\sigma}_n}\right)
\end{aligned}
$$

Schritt 1: Standardisieren.
Schritt 2: Rückführung auf die (tabellierte) Verteilungsfunktion der $N(0,1)$-Verteilung.
Schritt 3: Ersetzen von σ durch $\widehat{\sigma}_n$.

Hierbei ist

$$
\begin{aligned}
b &: \text{Benchmark} \\
\mu &: \text{erwarteter Quartalsgewinn im Szenario} \\
\widehat{\sigma}_n &: \text{Schätzung für } \sigma
\end{aligned}
$$

ZGWS: Anwendung auf die Fallgestaltung

Zahlenbeispiel:
Sei $n = 36$. Für die Standardabweichung liege eine Schätzung aus historischen Daten vor:

$$\widehat{\sigma}_n = 0.2 \text{ Mio Euro.}$$

Benchmark: $b = 50.4$, d.h. 1.4 Mio pro Händler.

Als relevante Szenarien werden erwartete Quartalsgewinne zwischen 1.2 und 1.6 Mio Euro betrachtet.

Die folgende Grafik zeigt die zugehörige approximative Wahrscheinlichkeit, die Benchmark zu erreichen oder sogar zu schlagen.

ZGWS: Anwendung auf die Fallgestaltung