Statistik für Informatiker, SS 2019

1.5 Gesetz der großen Zahlen und zentraler Grenzwertsatz

Matthias Birkner

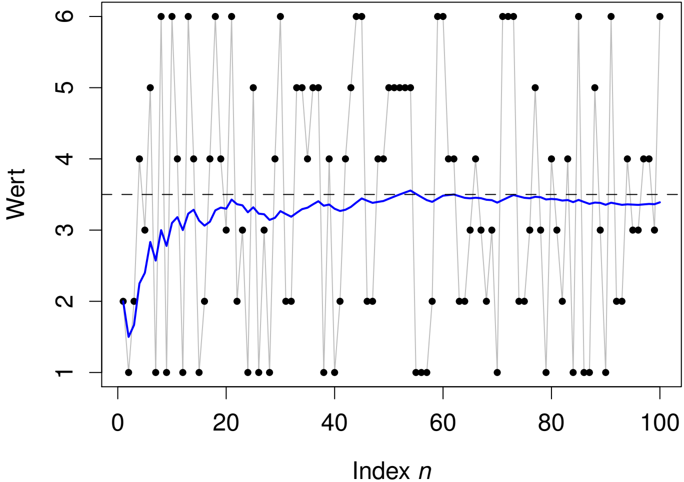
http://www.staff.uni-mainz.de/birkner/StatInfo19/

17.6.2019

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Erinnerung. Wir hatten bereits in Bem. 1.68, 3 das Gesetz der großen Zahlen illustriert:

$X_1, X_2, \ldots$ unabh., uniform auf $\{1,2,3,4,5,6\}$, X_n sind die schwarzen Punkte, $(X_1 + \cdots + X_n)/n$ die blaue Linie



Satz 1.88 ((Schwaches) Gesetz der großen Zahlen)

Seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte (u.i.v.) reellwertige ZVn mit $\mathbb{E}[X_1] = \mu$ und $\mathrm{Var}[X_1] < \infty$, dann gilt für jedes $\varepsilon > 0$

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{\mathrm{Var}[X_1]}{\varepsilon^2 n} \xrightarrow[n\to\infty]{} 0$$

Beweis. Sei $Y_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$, es ist

$$\mathrm{Var}[Y_n] = \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{Var}[X_i - \mu] + \sum_{\substack{1\leq i,j\leq n\\ i\neq j}}\mathrm{Cov}[X_i - \mu, X_j - \mu]\right)$$

$$= \frac{1}{n^2}\mathrm{Var}[X_i] = \frac{1}{n}\mathrm{Var}[X_1]$$

Satz 1.88 ((Schwaches) Gesetz der großen Zahlen)

Seien $X_1, X_2, \dots$ unabhängige und identisch verteilte (u.i.v.) reellwertige ZVn mit $\mathbb{E}[X_1] = \mu$ und $\mathrm{Var}[X_1] < \infty$, dann gilt für jedes $\varepsilon > 0$

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{\mathrm{Var}[X_1]}{\varepsilon^2 n} \underset{n\to\infty}{\longrightarrow} 0$$

$$Y_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \text{ hat } \mathrm{Var}[Y_n] = \frac{1}{n}\mathrm{Var}[X_1],$$

somit

$$P(|Y_n| \geq \varepsilon) \leq \frac{\mathrm{Var}[Y_n]}{\varepsilon^2} = \frac{\mathrm{Var}[X_1]}{\varepsilon^2 n}$$

gemäß Chebyshev-Ungleichung.

Bemerkung 1.89

1. Wir entnehmen dem Beweis von Satz 1.88 folgende kleine Verallgemeinerung:
   Sind $X_1, X_2, \dots \in \mathcal{L}^2$ seien paarweise unkorreliert mit

   $$\sup_n \mathrm{Var}[X_n] \leq \theta < \infty,$$

   dann gilt für $\varepsilon > 0$

   $$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mathbb{E}[X_i]\right)\right| > \varepsilon\right) \leq \frac{\theta}{\varepsilon^2 n} \;\left(\xrightarrow[n\to\infty]{} 0\right).$$

   (Das Argument geht genauso wie im Beweis von Satz 1.88, wenn wir $Y_n := \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mathbb{E}[X_i]\right)$ setzen.)

Bemerkung 1.89 (Fortsetzung)

1. Seien $Y_n, n \in \mathbb{N}$ und Y reellwertigen ZVn, die auf einem gemeinsamen Wahrscheinlichkeitsraum definiert sind.
Man sagt die Folge $(Y_n)_{n\in\mathbb{N}}$ *konvergiert stochastisch* gegen Y, auch geschrieben

$$Y_n \xrightarrow[n\to\infty]{\text{stoch.}} Y,$$

(auch $Y_n \xrightarrow[n\to\infty]{} Y$ stoch. oder $Y_n \xrightarrow[n\to\infty]{P} Y$), wenn gilt

$$\forall\, \varepsilon > 0 : \quad \lim_{n\to\infty} P\big(|Y_n - Y| > \varepsilon\big) = 0.$$

Man spricht damit Satz 1.88 oft folgendermaßen aus:

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{stoch.}} \mu$$

Bericht 1.90 (Nur der Vollständigkeit halber)

Die Konvergenzaussage in Satz 1.88 sieht (zumindest mit Blick auf die in der Analysis übliche Definition der Konvergenz) vielleicht etwas merkwürdig aus.

Tatsächlich gilt für $X_1, X_2, \dots$ u.i.v. mit $\mathbb{E}[X_1] = \mu$ auch:

Für jedes $\varepsilon > 0$ gibt es ein (vom Zufall abhängiges) N_0 mit

$$\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \leq \varepsilon \quad \text{für alle } n \geq N_0.$$

In der Literatur heißt dies manchmal das *starke Gesetz der großen Zahlen*, man sagt auch $(X_1 + \cdots + X_n)/n$ *konvergiert fast sicher* gegen μ.

Wir werden dies im weiteren Verlauf der Vorlesung nicht verwenden.

Vorbemerkung. Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilt (u.i.v.), $\mathbb{E}[X_1] = \mu$, $\mathrm{Var}[X_1] = \sigma^2 < \infty$.

Wir haben gesehen, dass $X_1 + \cdots + X_n \approx n\mu$ mit hoher Wahrscheinlichkeit, denn

$$\frac{X_1 + \cdots + X_n}{n} - \mu \xrightarrow[n\to\infty]{} 0 \quad \text{stochastisch}$$

gemäß dem Gesetz der großen Zahlen (Satz 1.88), aber feiner gefragt:

Wie groß ist $X_1 + \cdots + X_n - n\mu$ typischerweise?

Für $A \gg \sqrt{n}$ ist (mit Chebyshev-Ungleichung) zumindest

$$P\big(|X_1 + \cdots + X_n - n\mu| > A\big) \leq \frac{n\sigma^2}{A^2} \quad \text{(sehr) klein.}$$

Tatsächlich ist $\sqrt{n}$ die korrekte Größenordnung der typischen Abweichungen von $X_1 + \cdots + X_n$ von $n\mu$, beachte dazu

$$\mathbb{E}\big[(X_1 + \cdots + X_n) - n\mu\big] = 0$$

und $\mathrm{Var}\big[(X_1 + \cdots + X_n) - n\mu\big] = n\mathrm{Var}\big[X_1\big] = n\sigma^2$, also

$$S_n := \frac{(X_1 + \cdots + X_n) - n\mu}{\sqrt{n\sigma^2}} \quad \text{erfüllt} \quad \mathrm{Var}\big[S_n\big] = 1$$

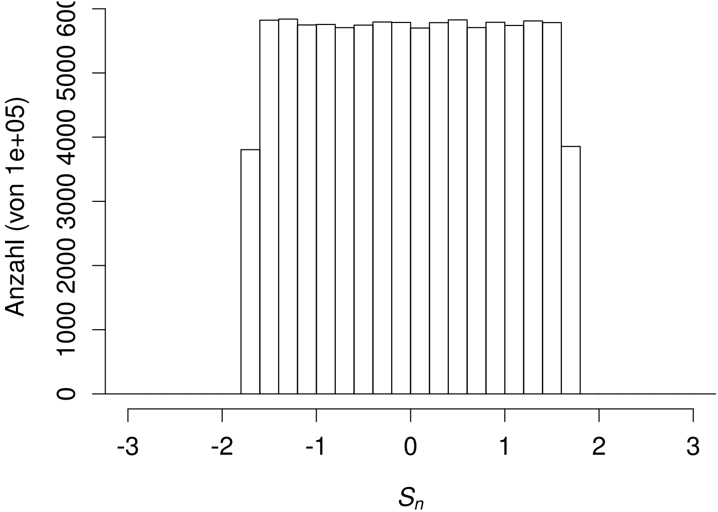
Demnach: Mit dieser Skalierung hängen zumindest Erwartungswert und Varianz nicht mehr von n ab.

Wie sieht es aber mit der „ganzen“ Verteilung aus?

Wir betrachten dazu Simulationen ...

$X_i \sim \mathrm{unif}_{[0,1]}$, $\mathbb{E}[X_1] = 1/2$, $\mathrm{Var}[X_1] = \frac{1}{12}$

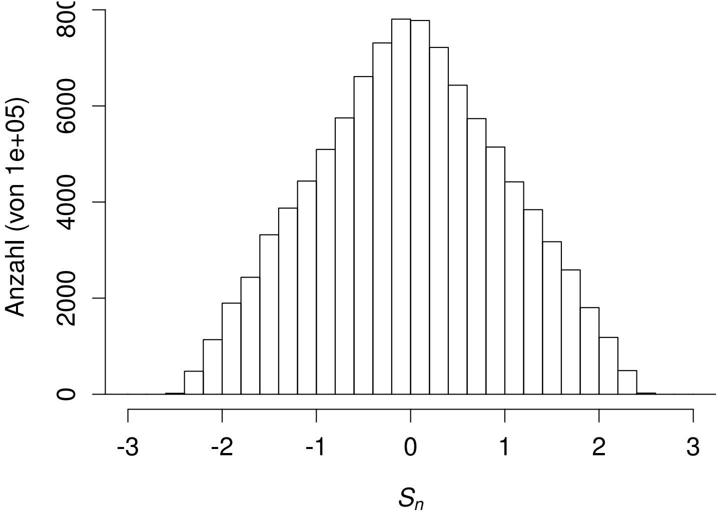
$n = 1$



(Histogramme jeweils basierend auf 10^5 Simulationen von S_n)

$X_i \sim \mathrm{unif}_{[0,1]}$, $\mathbb{E}[X_1] = 1/2$, $\mathrm{Var}[X_1] = \frac{1}{12}$

$n = 2$



(Histogramme jeweils basierend auf 10^5 Simulationen von S_n)

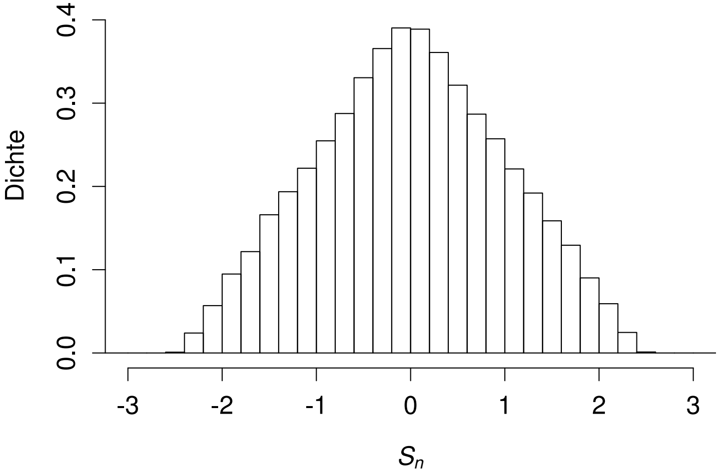
Zur Vergleichbarkeit gehen wir von den absoluten Anzahlen als Balkenhöhen zur sogenannten Dichte über, d.h. die Balkenhöhe ist nun jeweils

$$\frac{\text{Anzahl}}{100.000 \times \text{Balkenbreite}}.$$

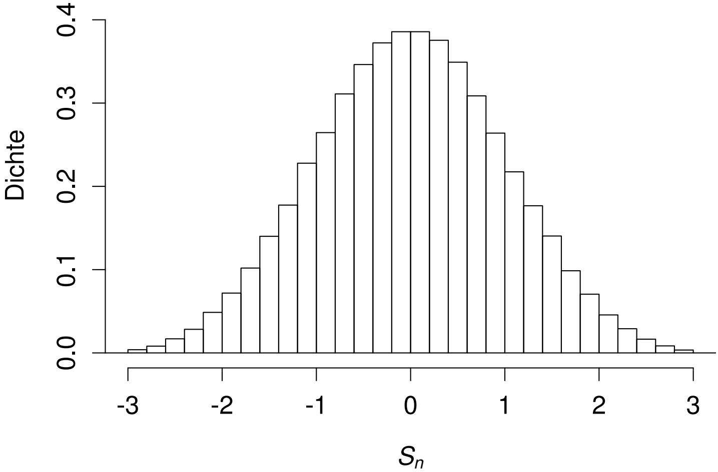
Damit wird die Gesamtfläche der Balken $= 1$ (wie bei einer Wahrscheinlichkeitsdichte).

(Da wir gleich breite Balken verwendet haben, entspricht dies einfach der Wahl einer anderen Skala auf der y-Achse)

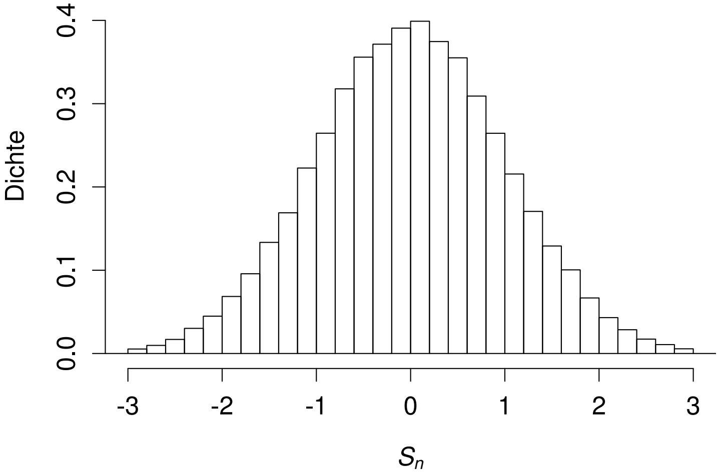
$n = 2$



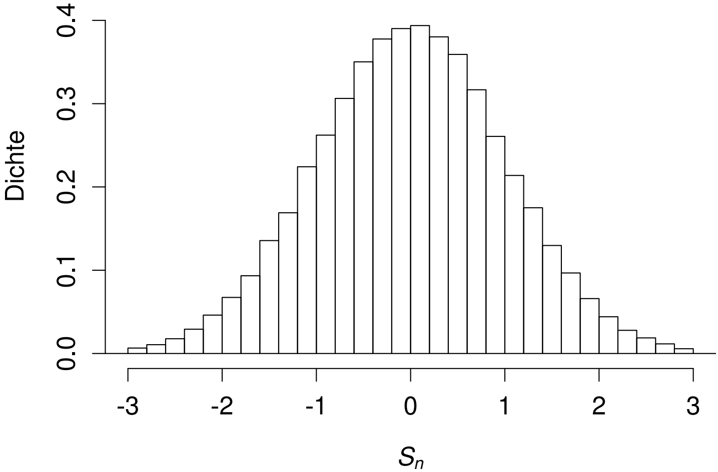
$n = 5$



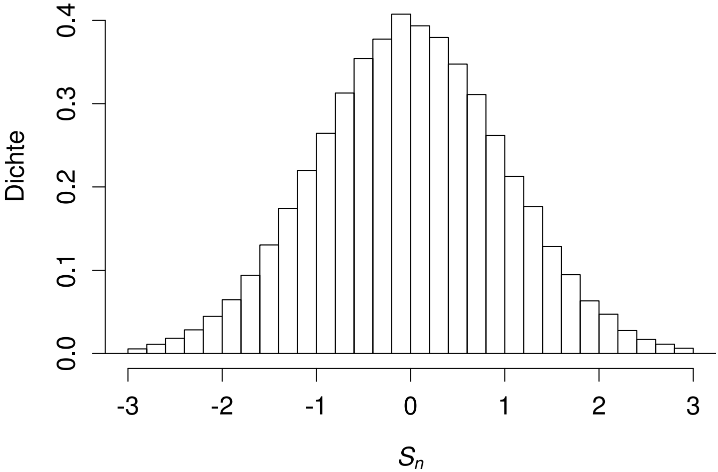
$n = 20$



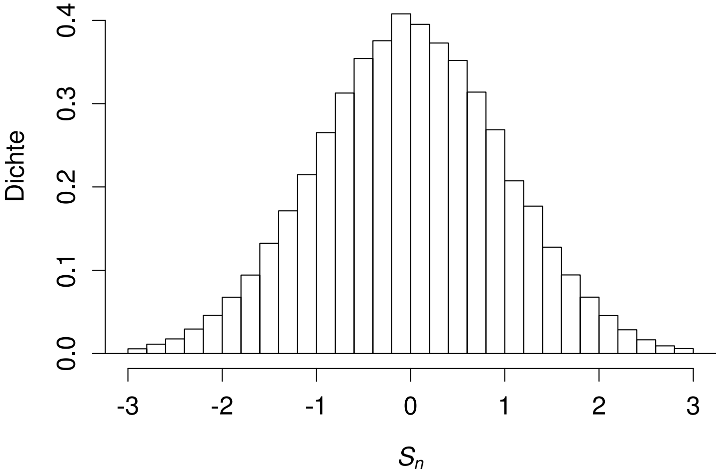
$n = 100$



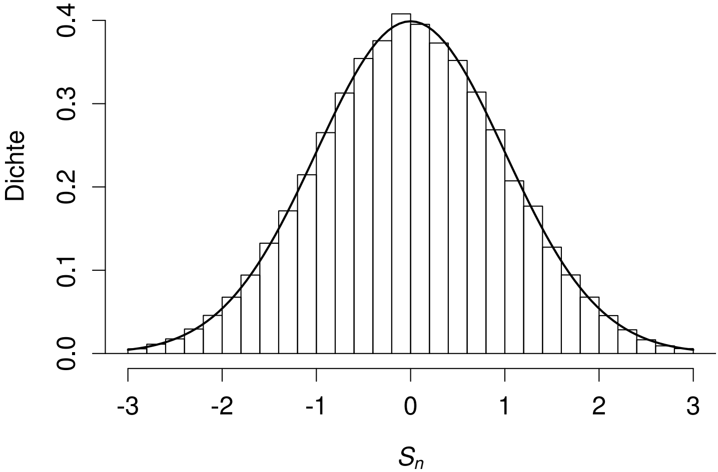
$n = 200$



$n = 500$



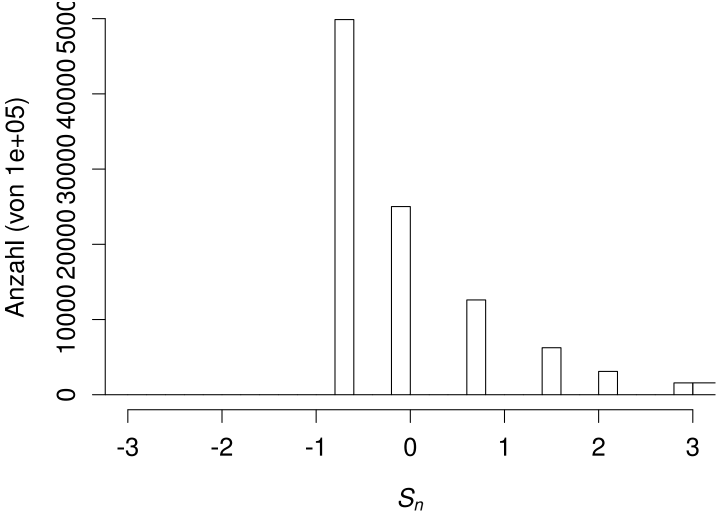
$n = 500$



Die schwarze Kurve ist die Dichte der Standard-Normalverteilung, $f(x) = (2\pi)^{-1/2} \exp(-x^2/2)$.

Nun dasselbe nochmal mit $X_i \sim \mathrm{geom}_{1/2}$
($\mathbb{E}[X_1] = 1$, $\mathrm{Var}[X_1] = 2$):

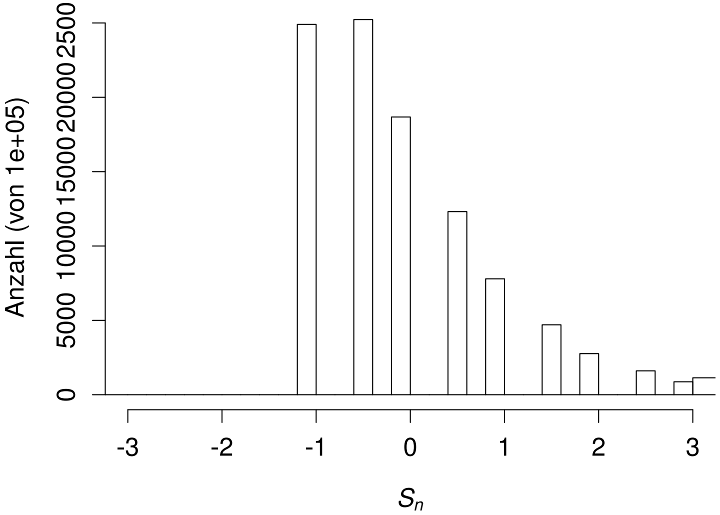
$n = 1$



(Histogramme jeweils basierend auf 10^5 Simulationen von S_n)

Nun dasselbe nochmal mit $X_i \sim \mathrm{geom}_{1/2}$
$(\mathbb{E}[X_1] = 1, \mathrm{Var}[X_1] = 2)$:

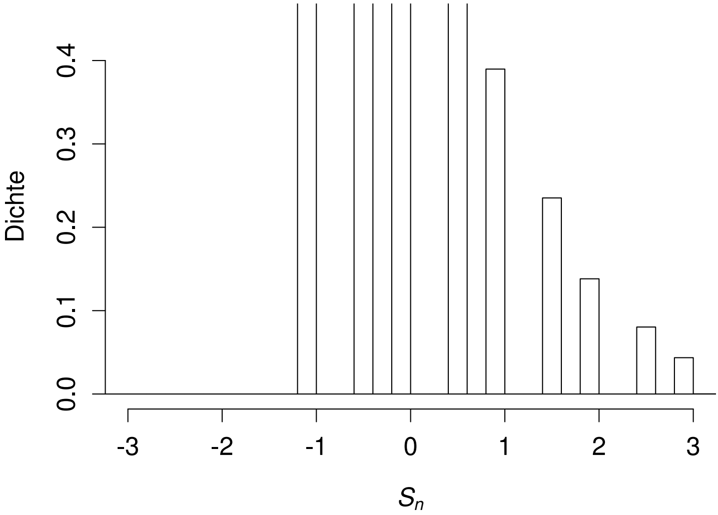
$n = 2$



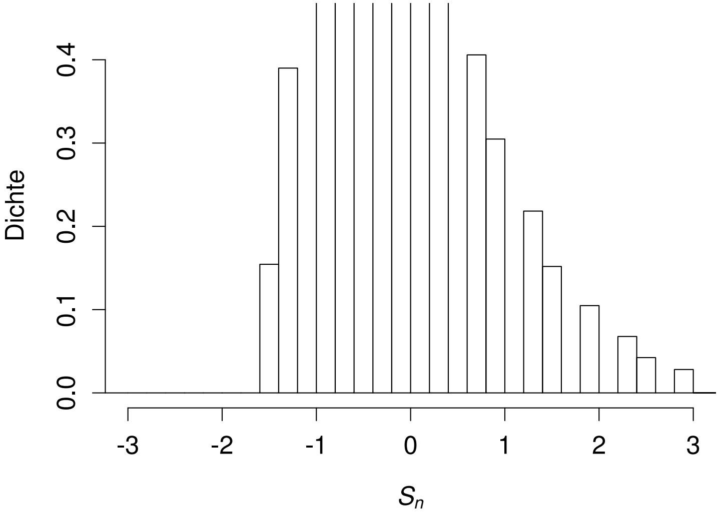
(Histogramme jeweils basierend auf 10^5 Simulationen von S_n)

Zur Vergleichbarkeit gehen wir wieder von den absoluten Anzahlen als Balkenhöhen zur sogenannten Dichte über.

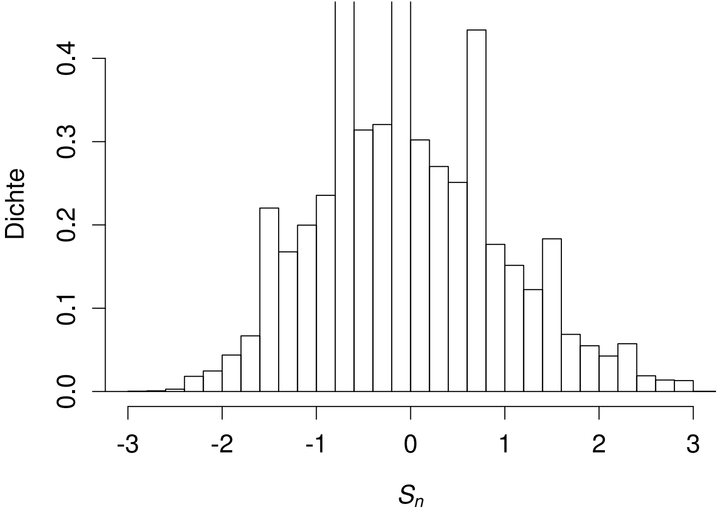
$n = 2$



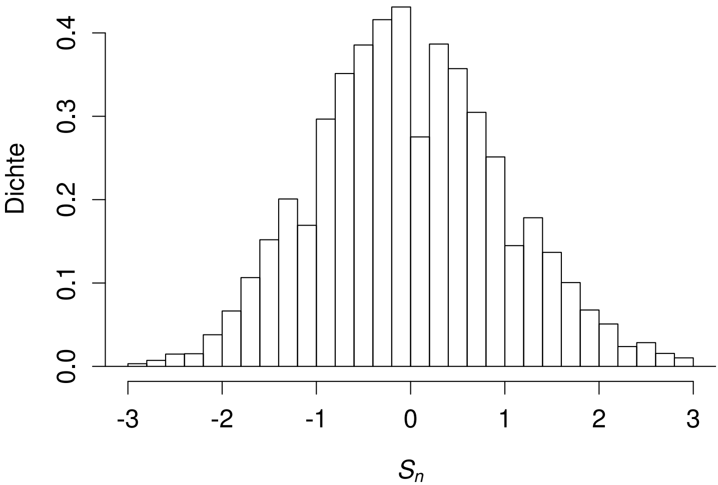
$n = 5$



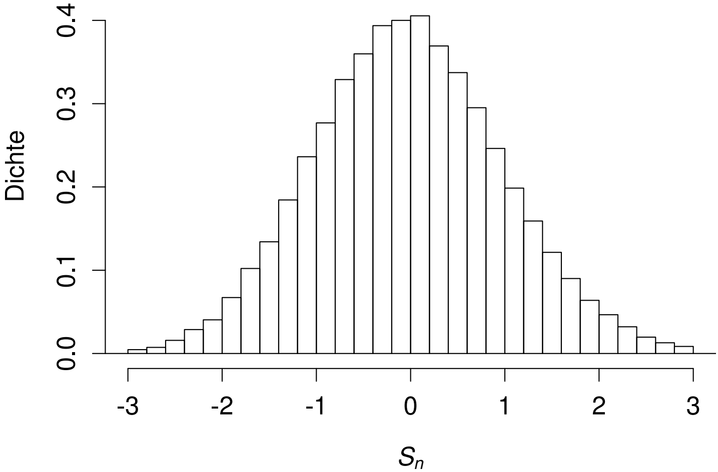
$n = 20$



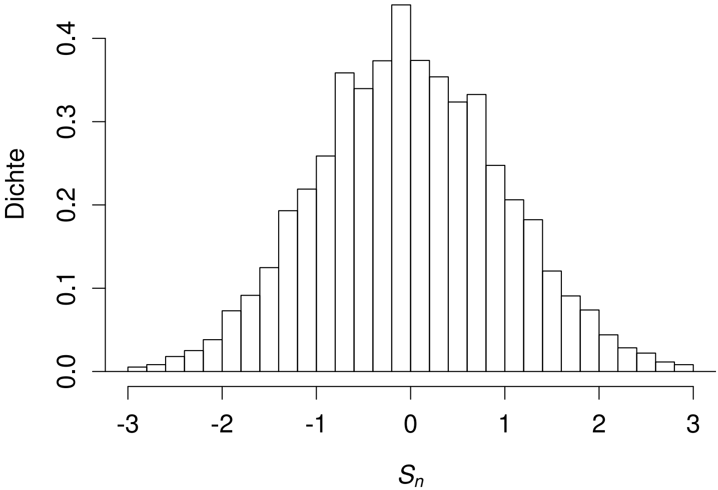
$n = 100$



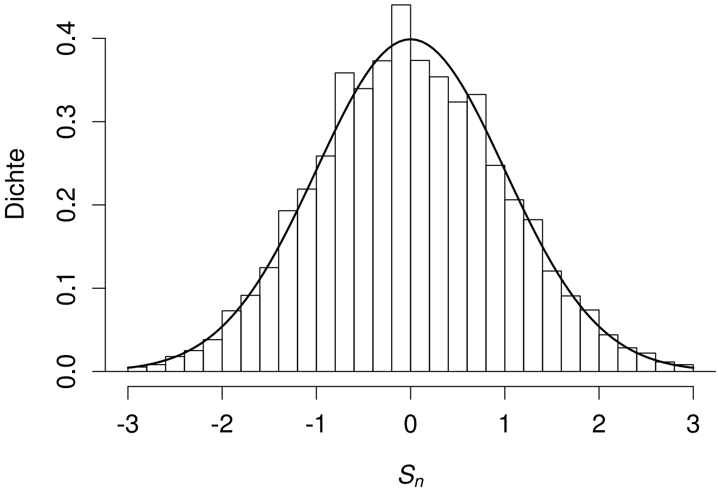
$n = 200$



$n = 500$



$n = 500$



Die schwarze Kurve ist die Dichte der Standard-Normalverteilung, $f(x) = (2\pi)^{-1/2} \exp(-x^2/2)$.

Wir sehen: Für genügend großes n ist die Verteilung von

$$\frac{(X_1 + \cdots + X_n) - n\mu_{X_1}}{\sqrt{n\sigma^2_{X_1}}} \overset{d}{\approx} Z \quad \text{mit } Z \sim \mathcal{N}_{0,1}. \qquad (*)$$

Übrigens: Da die Summe unabhängiger, normalverteilter ZVn wieder normalverteilt ist, gilt für $X_i \sim \mathcal{N}_{\mu,\sigma^2}$

$$\frac{(X_1 + \cdots + X_n) - n\mu}{\sqrt{n\sigma^2}} \sim \mathcal{N}_{0,1},$$

d.h. dann gilt $(*)$ exakt

(dies folgt aus Bsp. 1.65 zusammen mit Bsp. 1.32, 1.)

Satz 1.91 („Zentraler Grenzwertsatz“)

Seien $X_1, X_2, \ldots$ u.i.v. reelle ZVn $\in \mathscr{L}^2$ mit $\mathrm{Var}[X_1] \in (0, \infty)$, dann gilt für $-\infty \le a < b \le \infty$

$$\lim_{n\to\infty} P\left(a \le \frac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n\mathrm{Var}[X_1]}} \le b\right) = P(a \le Z \le b) \quad (*)$$

mit $Z \sim \mathcal{N}_{0,1}$.

Die wichtige Botschaft von Satz 1.91 lautet:

Eine Summe von vielen unabhängigen und identisch verteilten zufälligen Summanden ist (approximativ) normalverteilt.

Bemerkung 1.92

Die Eigenschaft $(*)$ im zentralen Grenzwertsatz wird auch ausgesprochen als „Konvergenz in Verteilung“: X, X_n reellwertige ZVn, so sagt man

$$X_n \underset{n\to\infty}{\longrightarrow} X \ \text{ in Verteilung}$$

(auch $X_n \xrightarrow[n\to\infty]{d} X$ oder $X_n \xrightarrow[n\to\infty]{\mathscr{L}} X$ geschrieben), wenn gilt

$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x) \ \big(= F_X(x)\big)$$

für jedes $x \in \mathbb{R}$ (an dem F_X stetig ist).

Satz 1.91 besagt also: $\dfrac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n\mathrm{Var}[X_1]}} \xrightarrow[n\to\infty]{d} Z$

Bericht 1.93

Es gibt viele Verallgemeinerungen von Satz 1.91, die die Annahme, dass die X_i u.i.v. sind, (stark) abschwächen.

Beweise des zentralen Grenzwertsatzes finden sich in der Lehrbuch-Literatur, wir betrachten hier nur ein heuristisches Argument:

> *„Warum taucht im zentralen Grenzwertsatz*
> *die Normalverteilung auf?“*

1.5.3 Eine Heuristik zum zentralen Grenzwertsatz

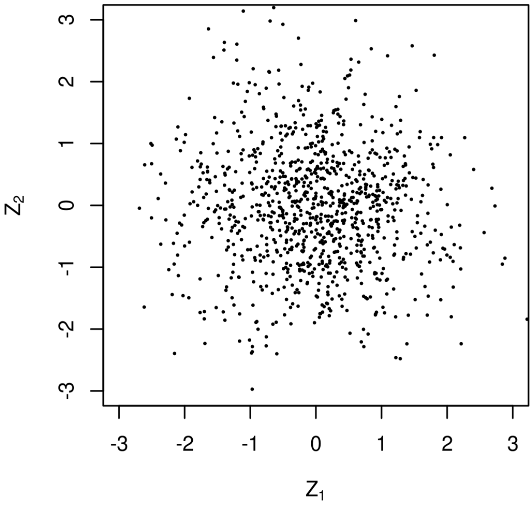
Beobachtung. In der Situation des zentralen Grenzwertsatzes sei

$$(Z_1, Z_2) := \left(\frac{X_1 + X_2 + \cdots + X_n - n\mu_{X_1}}{\sqrt{n\sigma_{X_1}^2}}, \frac{X_{n+1} + X_{n+2} + \cdots + X_{2n} - n\mu_{X_1}}{\sqrt{n\sigma_{X_1}^2}} \right),$$

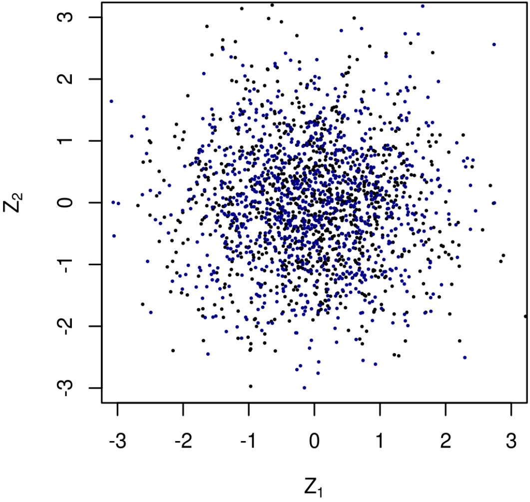
offenbar sind Z_1 und Z_2 unabhängig und identisch verteilt.

Betrachten wir die gemeinsame Verteilung von Z_1 und Z_2 (Simulationen mit $n = 200$, $X_i \sim \mathrm{unif}_{[0,1]}$)

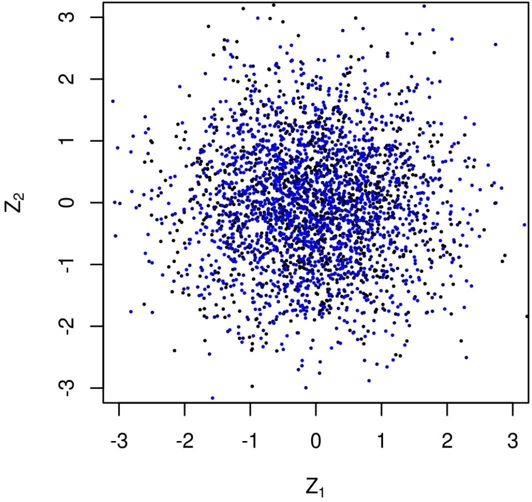
1000 simulierte Werte von (Z_1, Z_2)



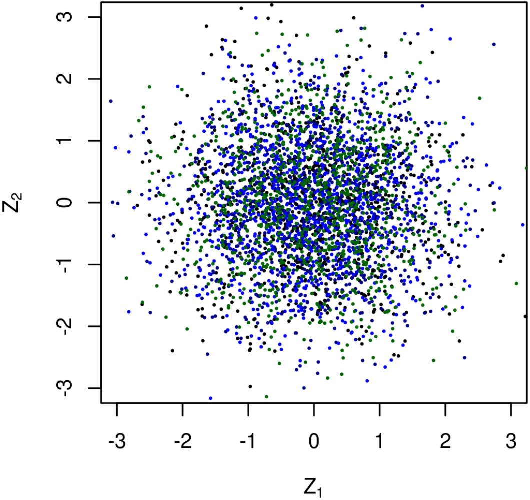
2000 simulierte Werte von (Z_1, Z_2)



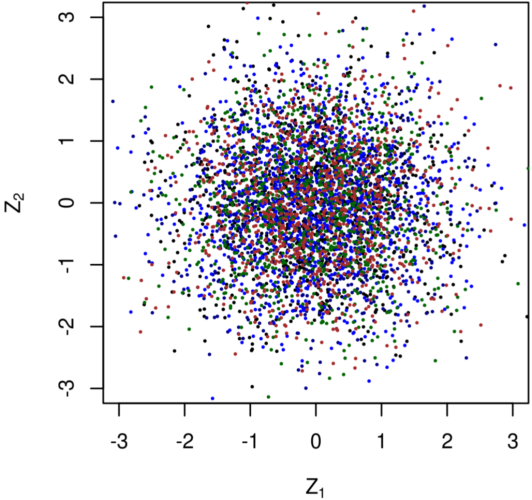
3000 simulierte Werte von (Z_1, Z_2)



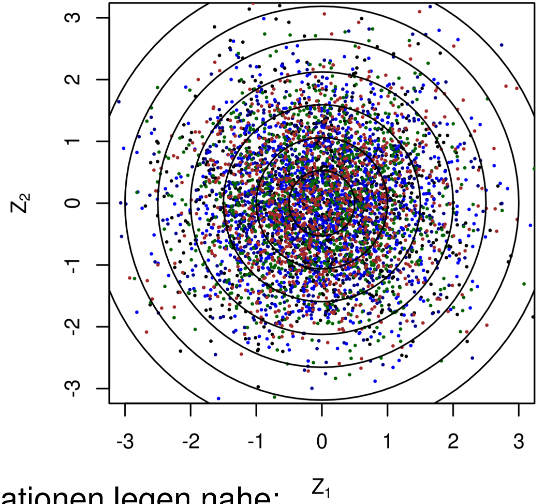
4000 simulierte Werte von (Z_1, Z_2)



5000 simulierte Werte von (Z_1, Z_2)



5000 simulierte Werte von (Z_1, Z_2)



Die Simulationen legen nahe:
(Z_1, Z_2) ist (approximativ) rotationssymmetrisch verteilt.

Beobachtung 1.94 (Eine Charakterisierung der Normalverteilung)

Seien Z_1, Z_2 unabhängige, reellwertige ZVn mit derselben Dichte f, so dass (Z_1, Z_2) rotationssymmetrisch verteilt ist.

Dann muss f eine (zentrierte) Normaldichte sein, d.h.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

für ein $\sigma^2 \in (0, \infty)$.

Beweis. Die (gemeinsame) Dichte $f_{(Z_1,Z_2)}$ ist rotationssymmetrisch, also gilt

$$f_{(Z_1,Z_2)}(z_1, z_2) = g\Big(\sqrt{z_1^2 + z_2^2}\Big)$$

für eine gewisse Funktion g (vgl. Beob. 1.58), andererseits gilt wegen Unabhängigkeit (vgl. Bericht 1.56)

$$f_{(Z_1,Z_2)}(z_1, z_2) = f(z_1)f(z_2)$$

Mit der Wahl $z_1 = r \geq 0, z_2 = 0$ folgt

$$f(r)f(0) = g\big(\sqrt{r^2}\big) = g(r)$$

(insbesondere muss f symmetrisch sein: $f(-r) = f(r)$).

$$f(r)f(0) = g\left(\sqrt{r^2}\right) = g(r)$$

Somit erfüllt f folgende Funktionalgleichung (setze oben $r = \sqrt{z_1^2 + z_2^2}$):

$$f(z_1)f(z_2) = f(0)f\left(\sqrt{z_1^2 + z_2^2}\right), \quad z_1, z_2 \in \mathbb{R}.$$

(Eine mögliche Lösung ist $f(z) = e^{-z^2}$, denn $e^{-z_1^2} \cdot e^{-z_1^2} = e^{-(z_1^2+z_2^2)} = 1 \cdot e^{-(\sqrt{z_1^2+z_2^2})^2}$.)

Zur allgemeinen Lösung: $w(x) := f\left(\sqrt{|x|}\right)$ erfüllt

$$w(a^2)w(b^2) = w(0)w\left(a^2 + b^2\right), \quad a, b \in \mathbb{R}$$

also gilt

$$w(u)w(v) = w(0)w(u + v), \quad u, v \geq 0$$

Für $w(x) := f\big(\sqrt{|x|}\big)$ gilt

$$w(u)w(v) = w(0)w(u+v), \quad u, v \geq 0 \quad (*)$$

Die allgemeine Lösung von $(*)$ hat die Form

$$w(u) = c_1 e^{-c_2 u},$$

also ist

$$f(z) = w(z^2) = c_1 \cdot \exp(-c_2 z^2)$$

für gewisse $c_1, c_2 > 0$;
wegen Normierung $\int_{-\infty}^{\infty} f(x)\,dx = 1$ muss dann
$c_1 = (2\pi\sigma^2)^{-1/2}$, $c_2 = 1/(2\sigma^2)$ für ein $\sigma^2 \in (0, \infty)$ gelten.