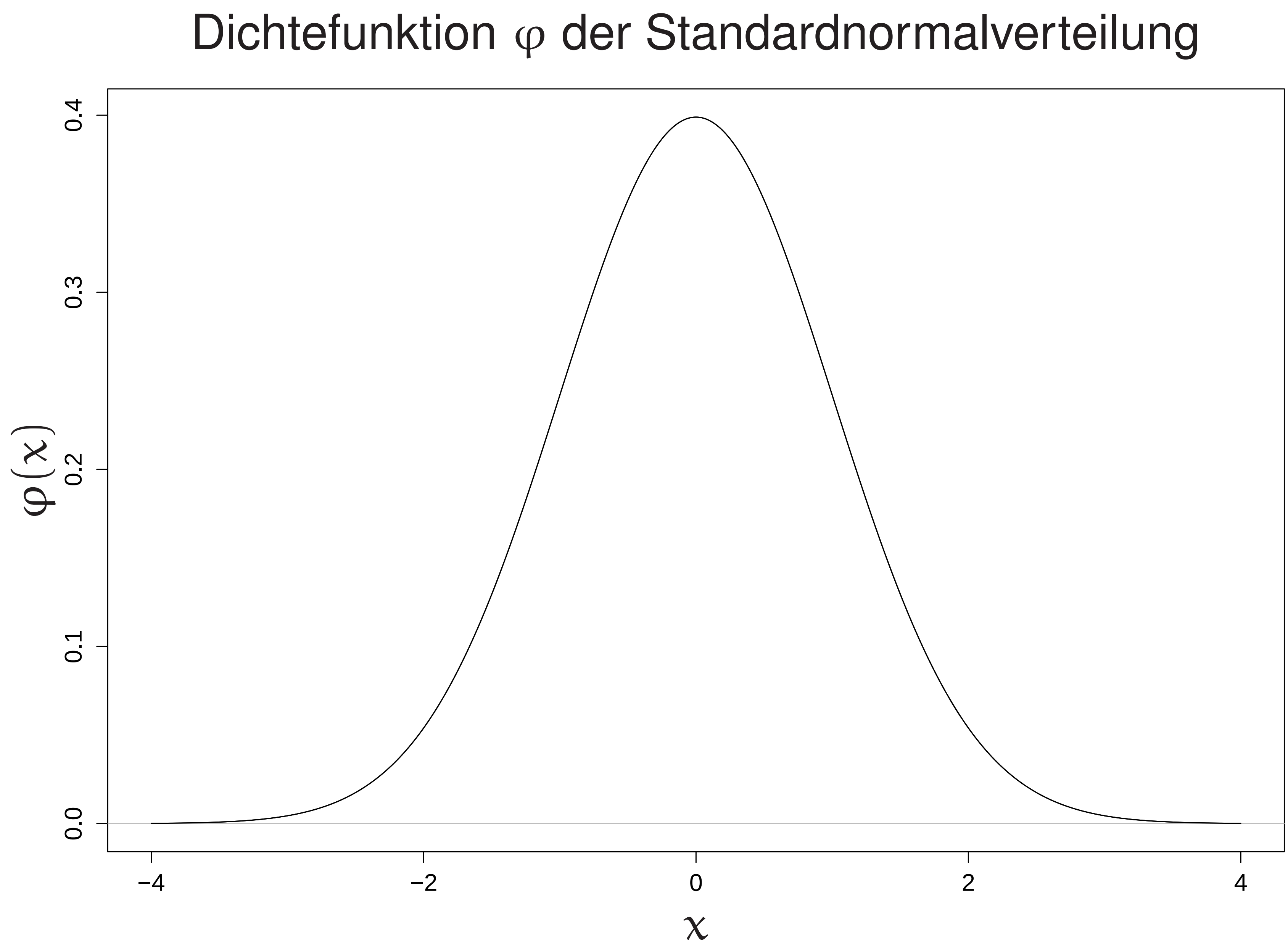
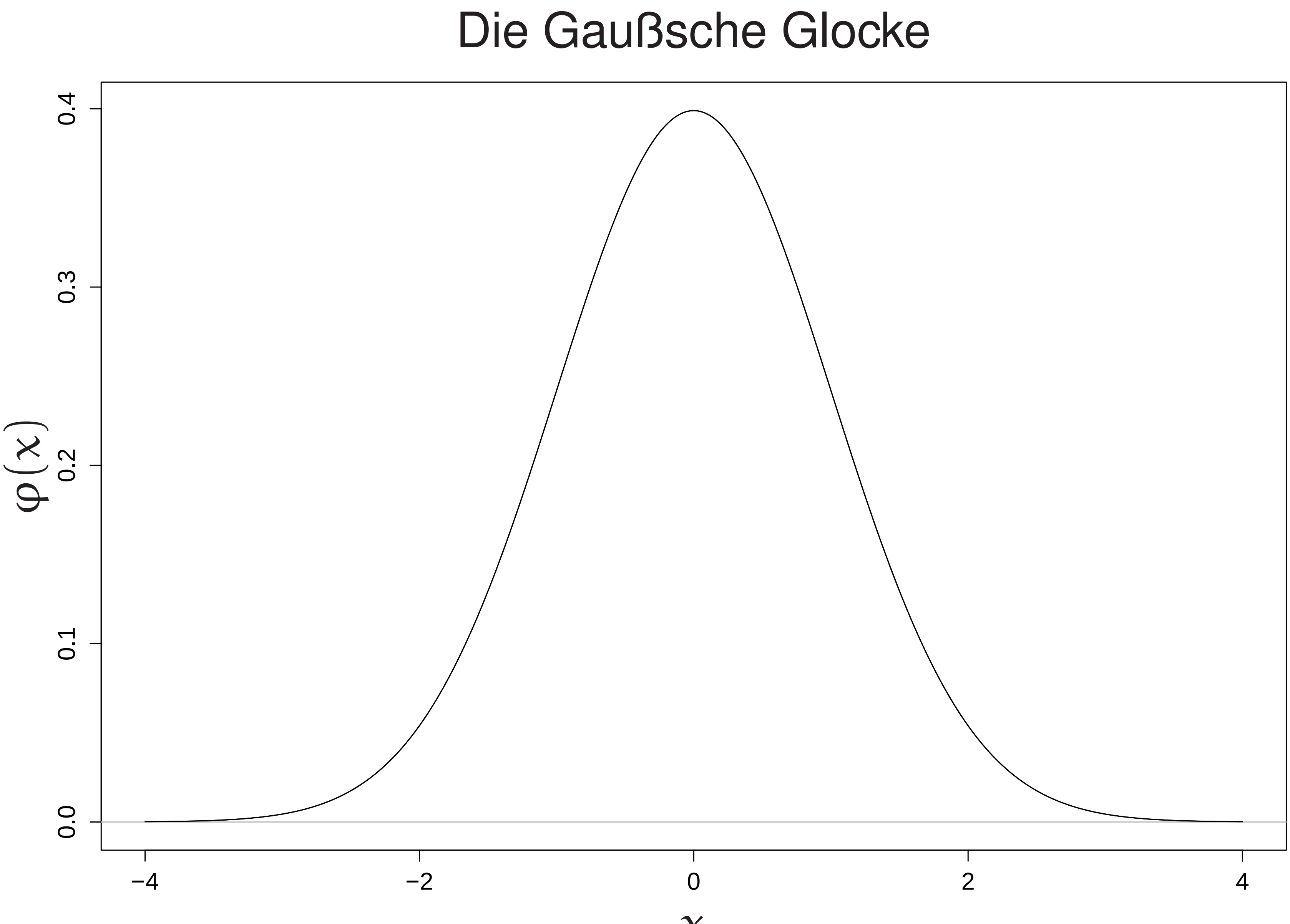
Der zentrale Grenzwertsatz

Wiederholung:

Die Normalverteilung



Dichtefunktion φ der Standardnormalverteilung
0.0
0.1
0.2
0.3
0.4
φ(x)
−4
−2
0
2
4
x


Die Gaußsche Glocke
0.4
0.3
0.2
0.1
0.0
φ(x)
−4
−2
0
2
4
x

Die Standardnormalverteilung

Dichtefunktion:

$$\varphi(x) := \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

Verteilungsfunktion:

$$\Phi(a) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx$$

Z standard-normalverteilt:

$$\mathbf{P}(\ Z \leq a\ ) = \Phi(a)$$

Für standard-normalverteiltes Z gilt:

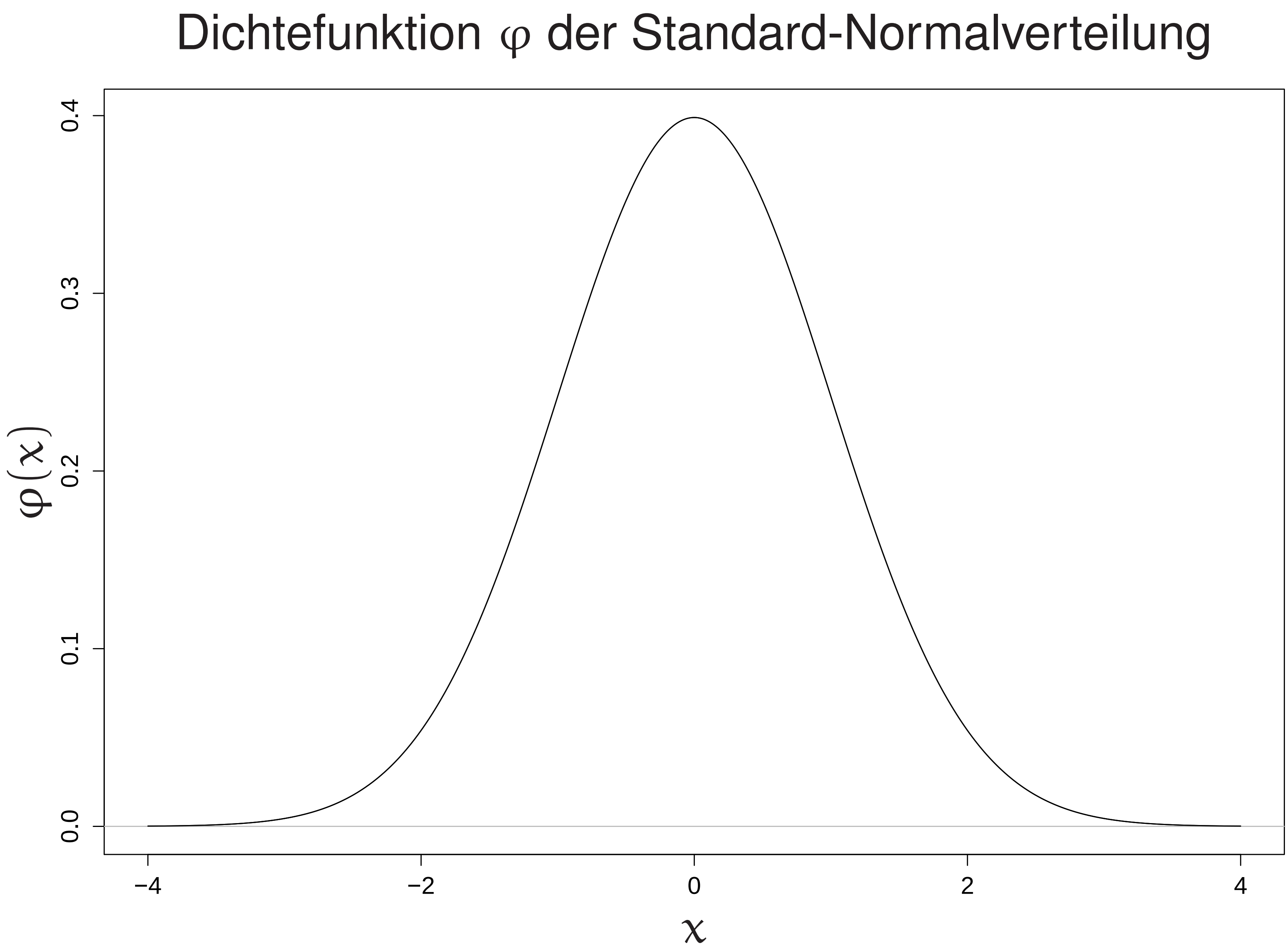
$$\mathbf{E}Z = 0 \qquad \sigma_Z = 1$$

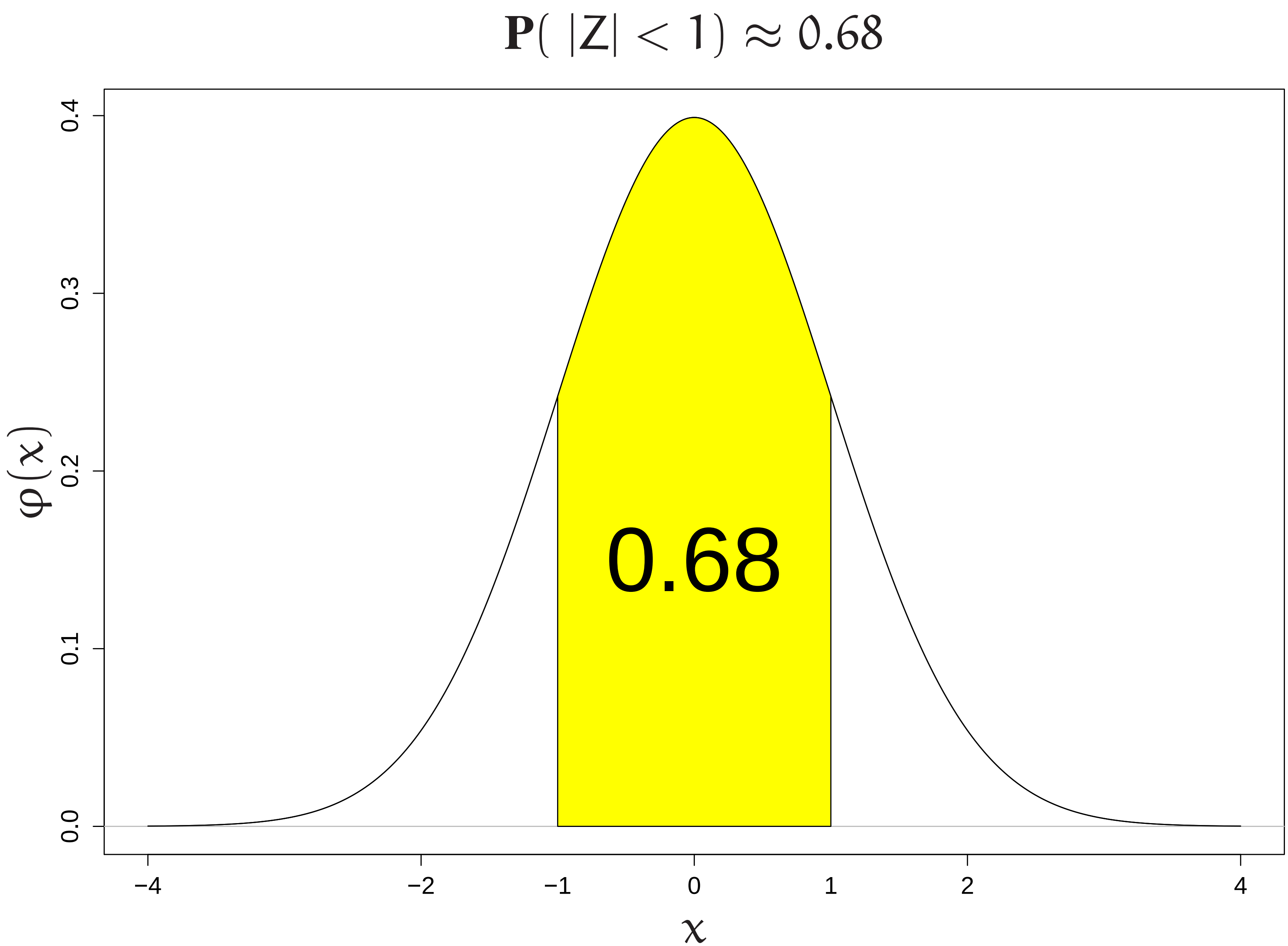
Allgemeine normalverteilte Zufallsvariable entstehen so:

$$N = \mu + \sigma Z$$

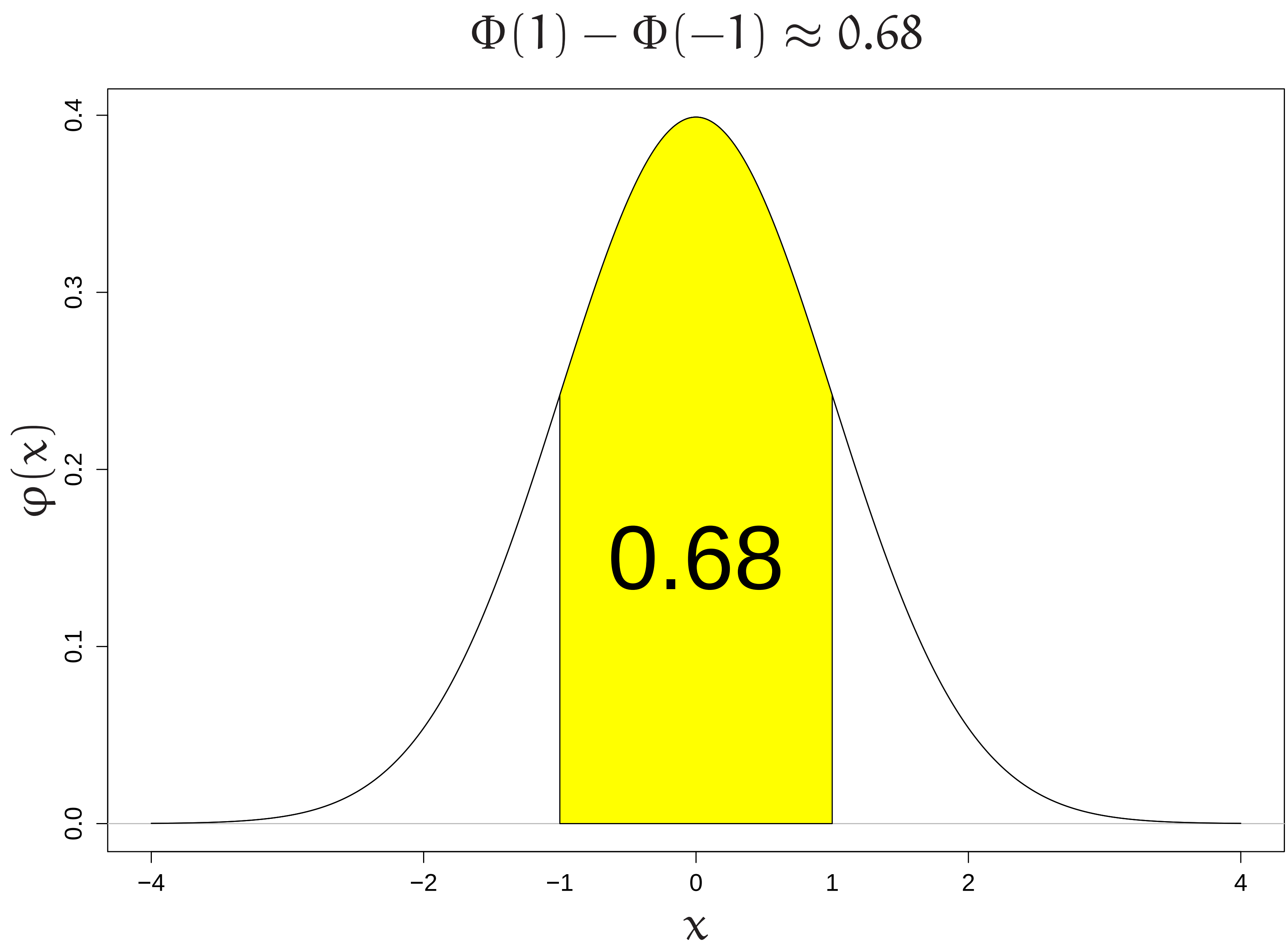
Für sie gilt:

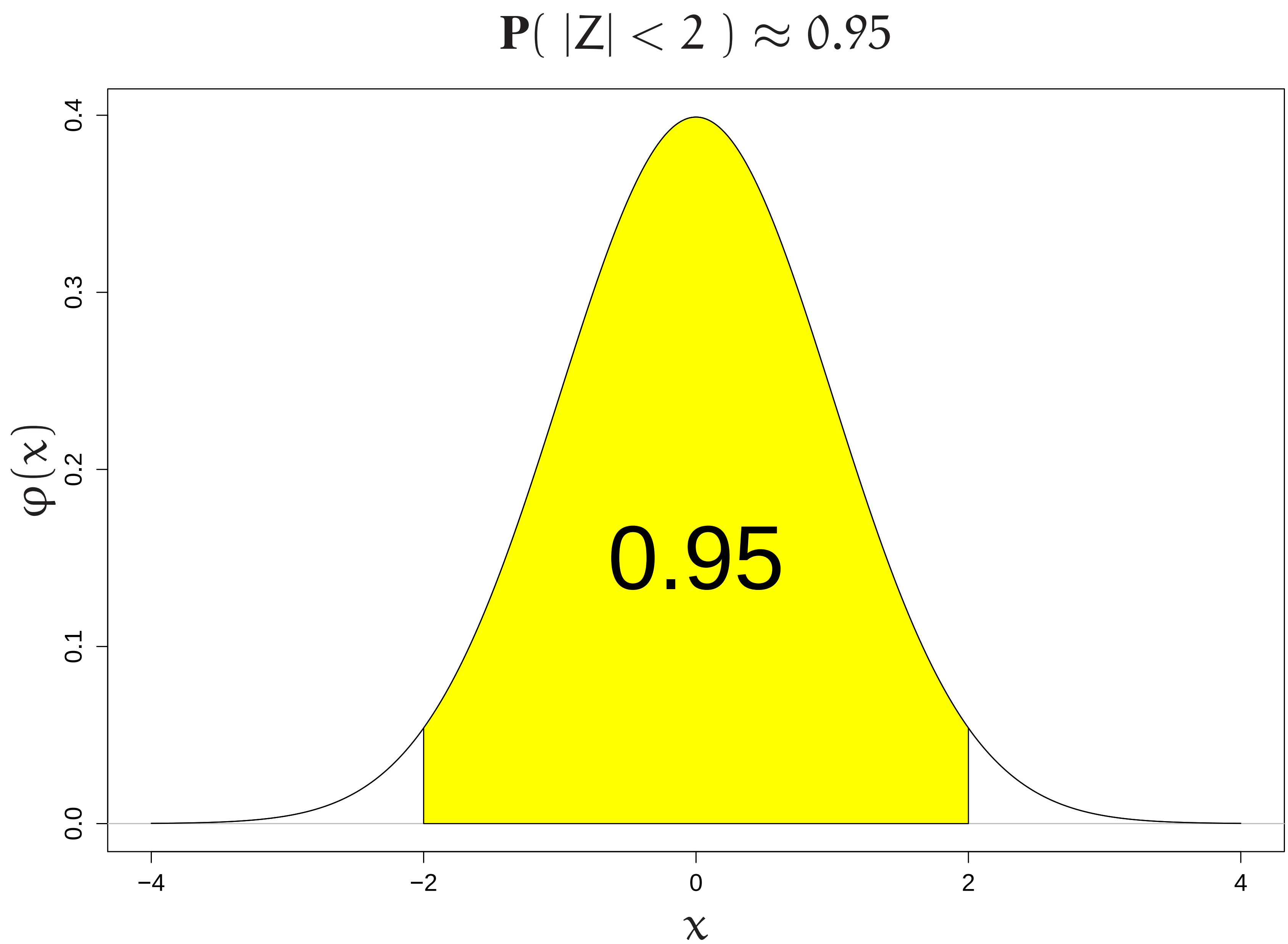
$$\mathbf{E}N = \mu \qquad \sigma_N = \sigma$$


Dichtefunktion φ der Standard-Normalverteilung
0.0
0.1
0.2
0.3
0.4
φ(x)
−4
−2
0
2
4
x

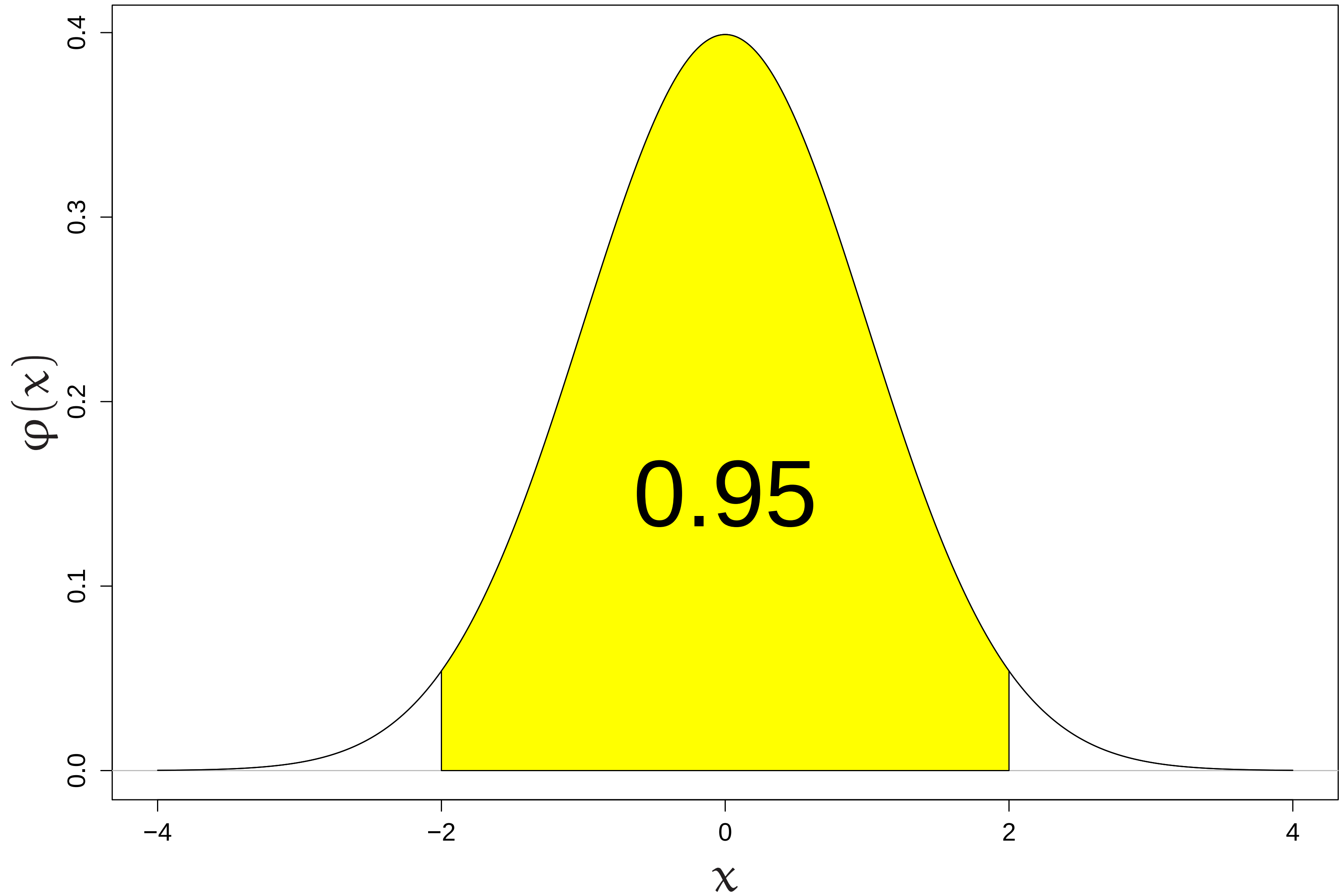


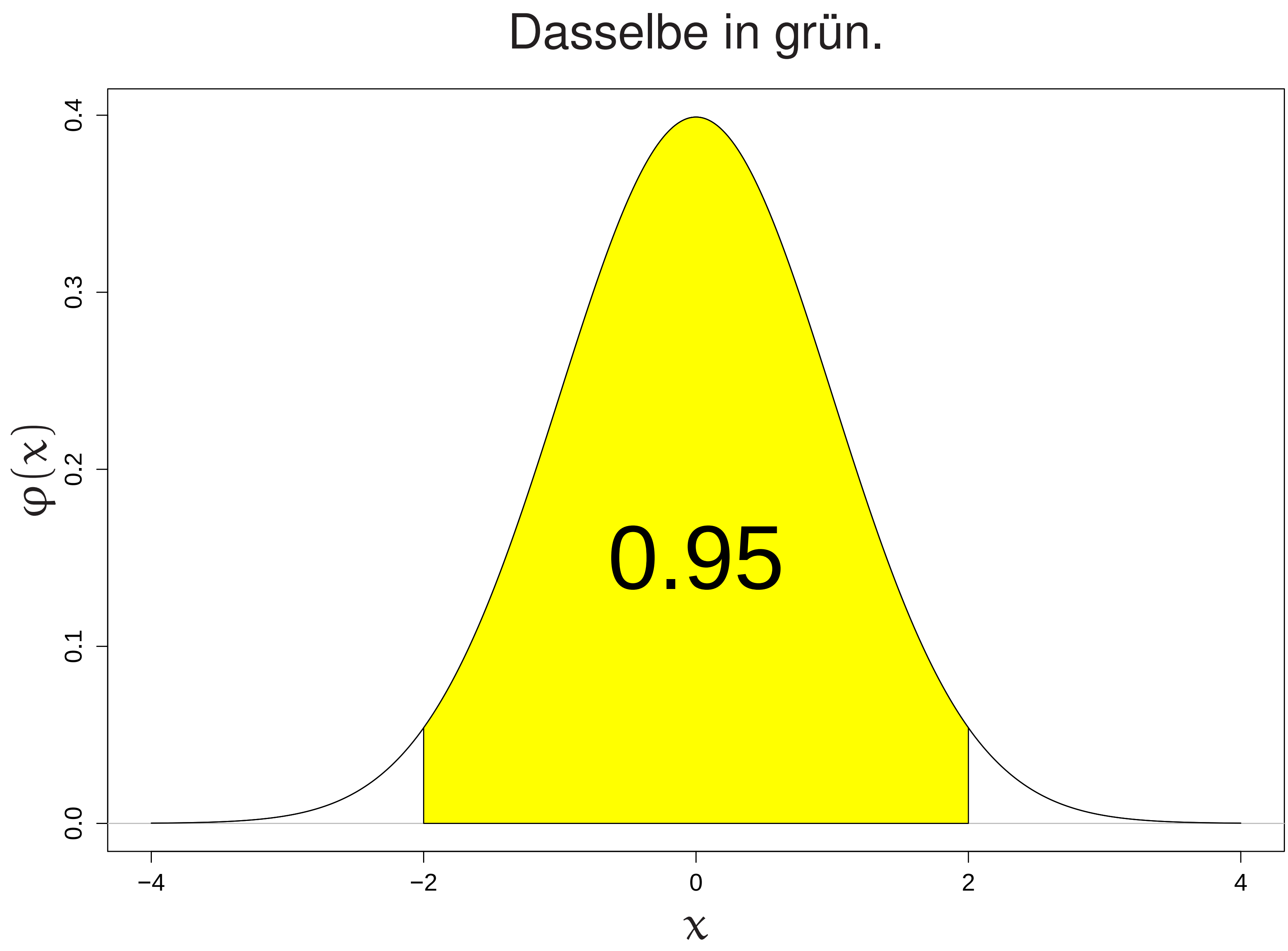
P(|Z| < 1) ≈ 0.68
0.68
φ(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
−1
0
1
2
4
x


$\Phi(1) - \Phi(-1) \approx 0.68$
0.68
$\varphi(x)$
x
0.0
0.1
0.2
0.3
0.4
−4
−2
−1
0
1
2
4

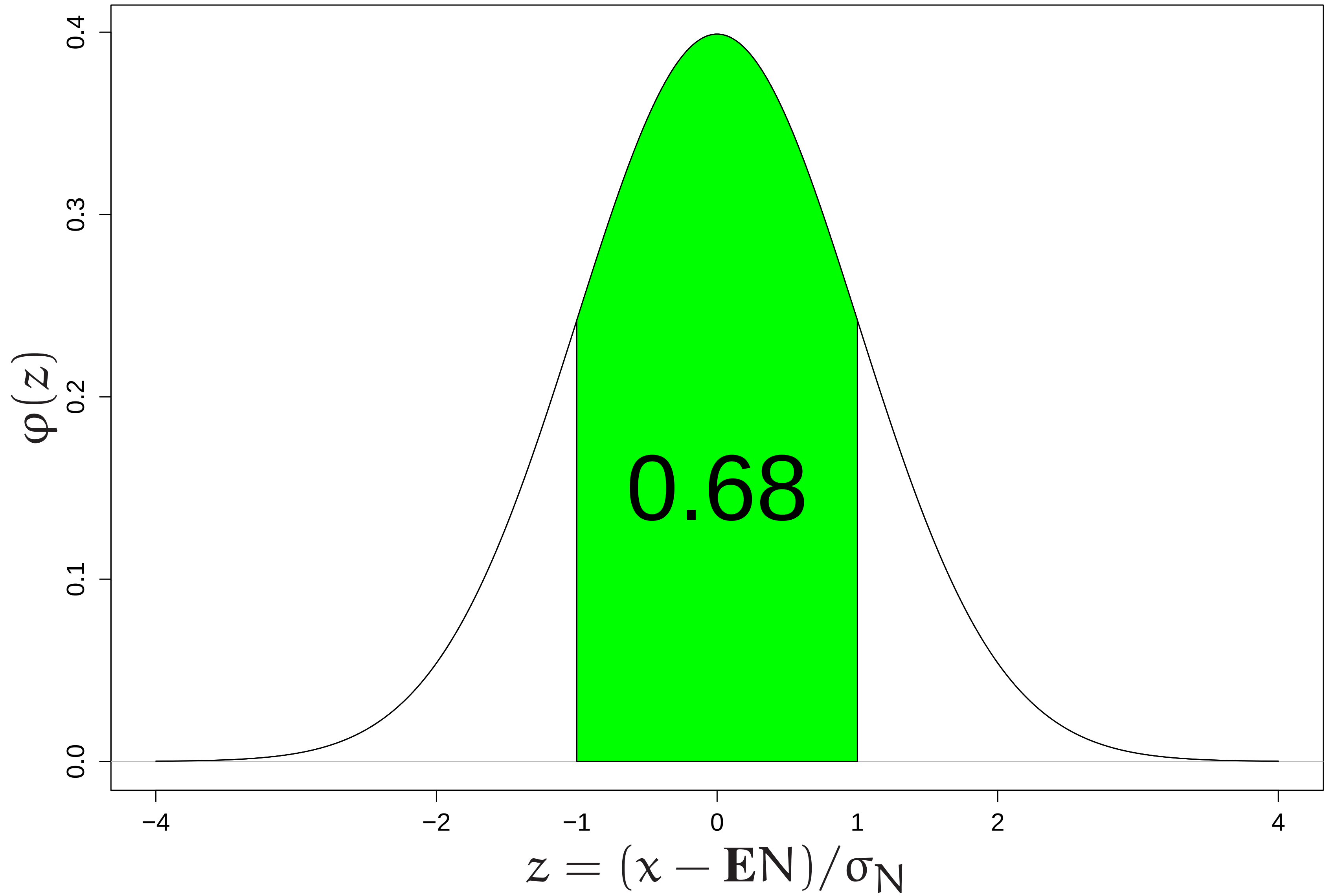

P(|Z| < 2) ≈ 0.95
0.95
φ(x)
x
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

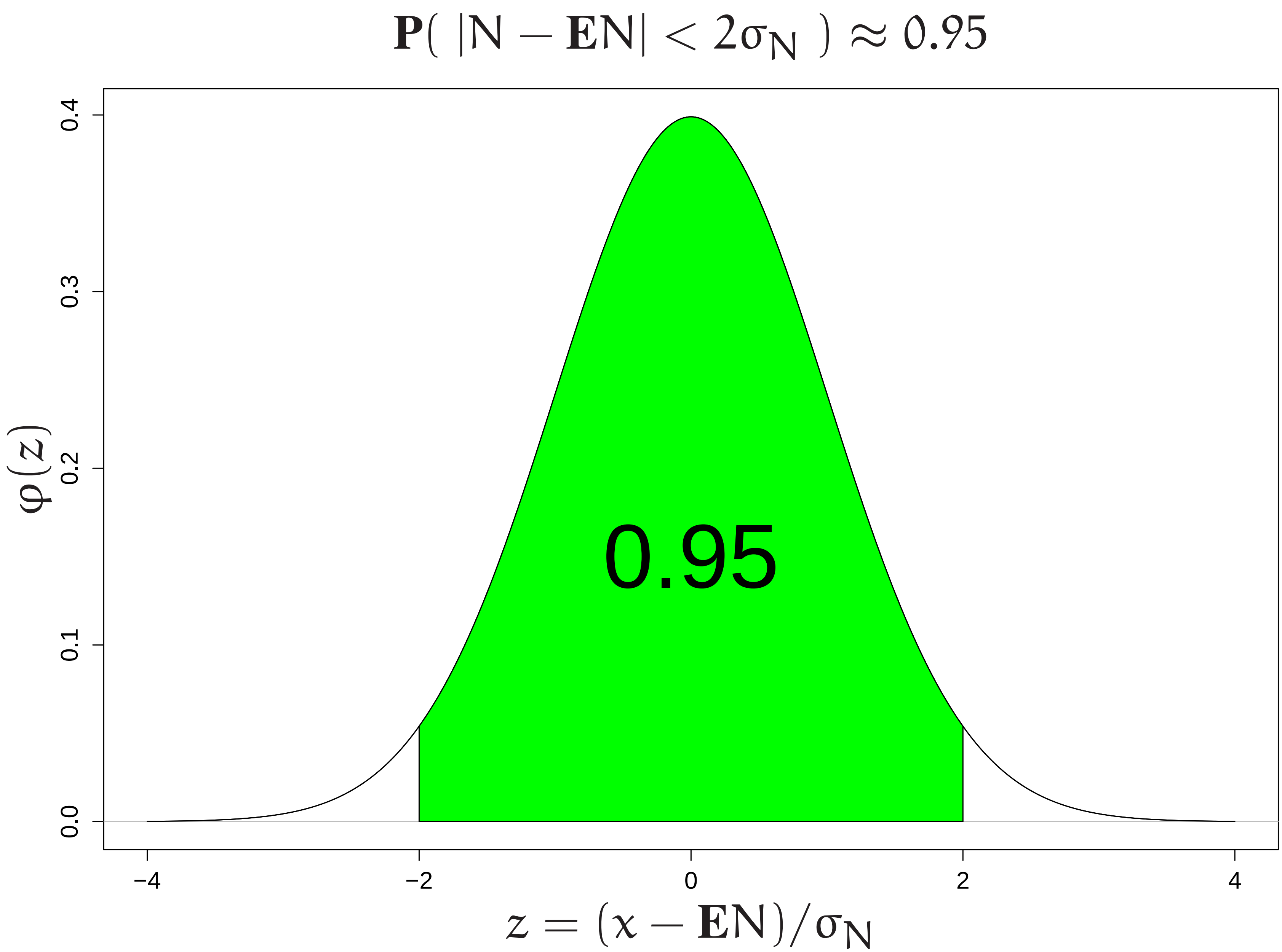
Und für allgemeine normalverteilte Zufallsgrößen $\mathcal{N}$?




Dasselbe in grün.
0.95
φ(x)
x
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

$$\mathbf{P}(\ |N - \mathbf{E}N| < \sigma_N\ ) \approx 0.68$$





P(|N − EN| < 2σ_N) ≈ 0.95
0.95
φ(z)
z = (x − EN)/σ_N
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Eine wichtige Botschaft der letzten Stunde war:

Die standardisierte Summe von unabhängigen,

identisch normalverteilten Zufallsvariablen

ist standard-normalverteilt

Mit anderen Worten: Sind $N_1, N_2, \ldots, N_n$

unabhängig und $N(\mu, \sigma^2)$-verteilt, dann ist

$$\frac{N_1 + \cdots + N_n - n\mu}{\sqrt{n\sigma^2}} \quad \text{standard-normalverteilt.}$$

Der Zentrale Grenzwertsatz liefert hiervon

eine gewaltige Verallgemeinerung

(asymptotisch für große n):

Zentraler Grenzwertsatz:

“Die standardisierte Summe von VIELEN

unabhängigen, identisch verteilten

nicht notwendig normalverteilten

$\mathbb{R}$-wertigen Zufallsvariablen

mit endlicher Varianz

ist annähernd standard-normalverteilt”

Formal:

Seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariable mit endlichem Erwartungswert μ und endlicher Varianz $\sigma^2 > 0$. Dann gilt für alle $c < d \in \mathbb{R}$

$$\mathbf{P}\Big(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \in [c, d]\Big) \underset{n\to\infty}{\longrightarrow} \mathbf{P}(Z \in [c, d]).$$

Dabei ist Z standard-normalverteilt.

In Worten:

Die standardisierte Summe von n unabhängigen,
identisch verteilten $\mathbb{R}$-wertigen Zufallsvariablen
mit endlicher Varianz
konvergiert für $n \to \infty$ in Verteilung
gegen eine standard-normalverteilte Zufallsvariable.

Ein (erster) Hinweis darauf, dass die standardisierte Summe
von n unabhängigen, identisch verteilten $\mathbb{R}$-wertigen Zufallsvariablen
etwas mit asymptotischer Normalität zu tun haben könnte:

Sei X eine Zufallsvariable mit $\mathbf{E}[X] = 0$ und $\mathbf{Var}[X] = 1$.

$$X_1, X_2, \ldots, \quad X_1', X_2', \ldots$$

seien unabhängige, identisch verteilte Kopien von X.

$$Z_n := \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n), \quad Z_n' := \frac{1}{\sqrt{n}}(X_1' + \cdots + X_n').$$

Dann ist $\frac{1}{\sqrt{2}}(Z_n + Z_n')$ so verteilt wie Z_{2n}.

Und für unabhängige, N(0,1)-verteilte Zufallsvariable Z, Z' wissen wir:

$\frac{1}{\sqrt{2}}(Z + Z')$ ist so verteilt wie Z

weil ja, wie wir gesehen haben, (Z, Z') rotationssymmetrisch verteilt ist!

Vielleicht ist für große n auch

(Z_n, Z_n') annähernd rotationssymmetrisch verteilt?

Eine kurze Geschichte

des

Zentralen Grenzwertsatzes

Abraham de Moivre:

Der faire Münzwurf (1733)

Pierre-Simon Laplace:

Allgemeine binomiale Zufallsgrößen (1812)

Pafnuty Lvovich Chebyshev:

Skizze eines Beweises für den allgemeinen Fall (1887)

Aleksandr Mikhailovich Lyapunov:

Noch allgemeiner (1906)

Allgemeiner zentraler Grenzwertsatz (1901)

Andrei Andreyevich Markov:

weitere Verallgemeinerungen (~ 1910)

Nehmen wir an,

diese Herren hätten sich

auf ihre vielen anderen Interessen

beschränkt.

ZENTRALER GRENZWERTSATZ

Nehmen wir an,

diese Herren hätten sich

auf ihre vielen anderen Interessen

beschränkt.

ZENTRALER GRENZWERTSATZ

Unbekannt.

Könnten <u>wir</u> ihn entdecken?

Wie kämen wir auf φ?

Warum gerade $e^{-x^2/2}$?

BEISPIEL

Rundungsfehler bei Addition

In Wirklichkeit

$$\pi =$$

$$3.14159265358979323846264338327950288419716939937510 5\ldots$$

Im Rechner

$$\pi \leftarrow 3.14159265358979$$

MODELL

Zahl = Rechnerdarstellung + Rundungsfehler.

$$A = a^{[R]} + \varepsilon X \qquad \varepsilon = 10^{-15}$$

Annahme: X uniform verteilt auf $[-0.5, 0.5]$.

$$\sum_{i=1}^{n} A_i = ?$$

$$\sum_{i=1}^{n} A_i = \sum_{i=1}^{n} a_i^{[R]} + \varepsilon \sum_{i=1}^{n} X_i$$

Wie groß ist der Fehler?

$$\sum_{i=1}^{n} X_i \approx ?$$

Ein Beispiel:

$X_1, X_2, \ldots$ unabhängig
und uniform auf $[-0.5, 0.5]$ verteilt

Empirische Verteilung von

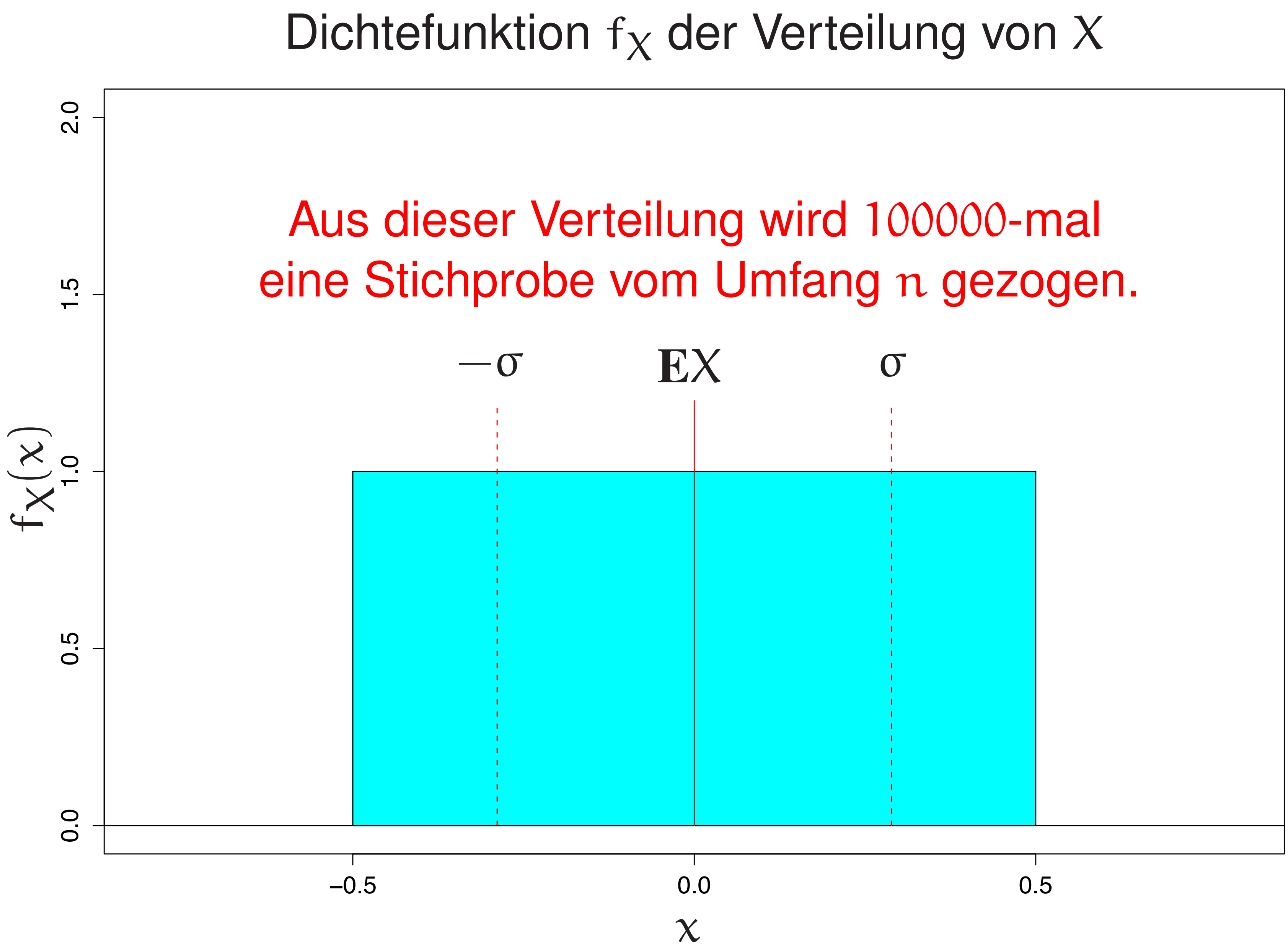
$$S_n := X_1 + ... + X_n$$

100000 Simulationen

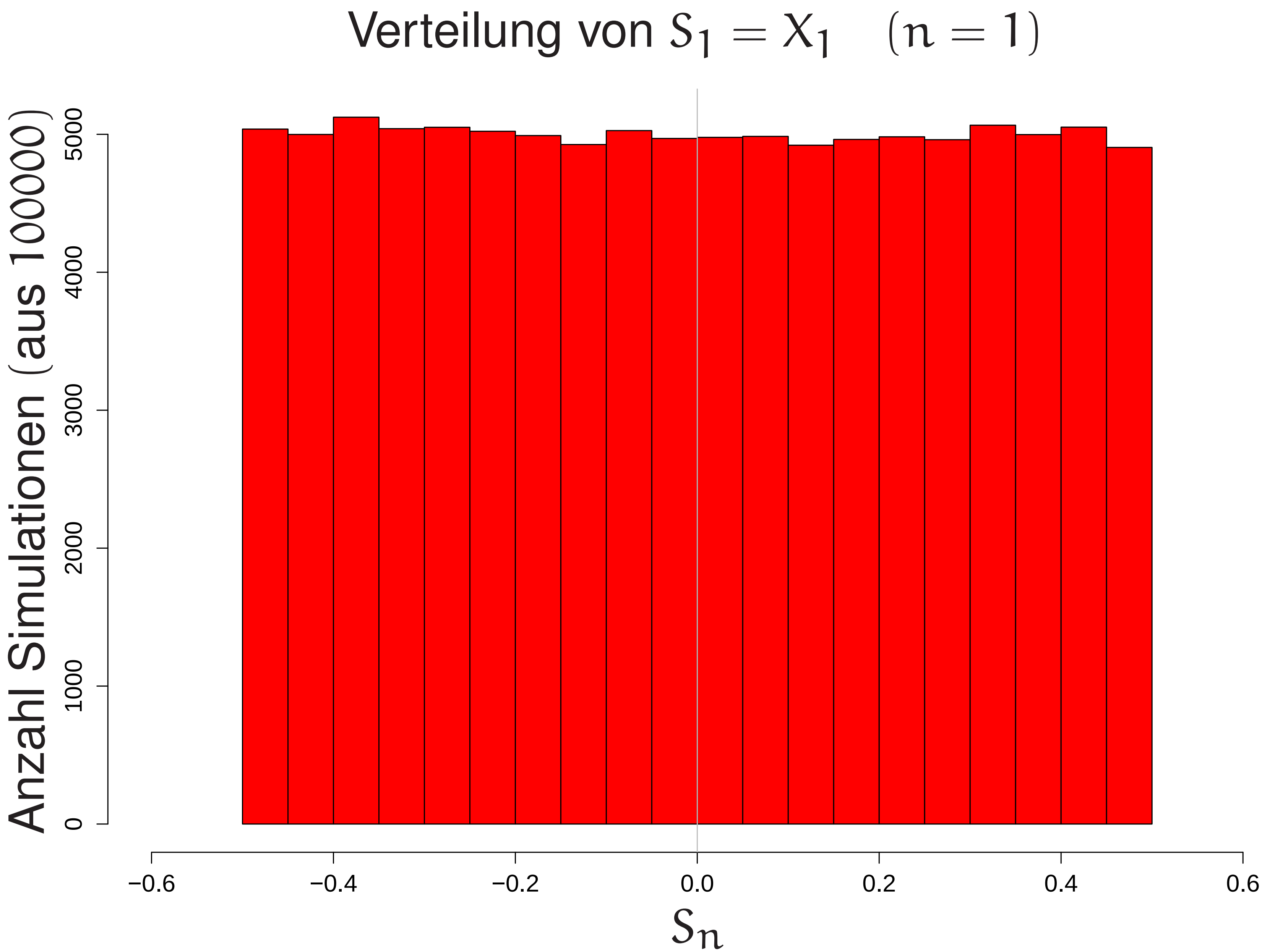
jeweils für

$$n = 1, 2, ..., 10$$

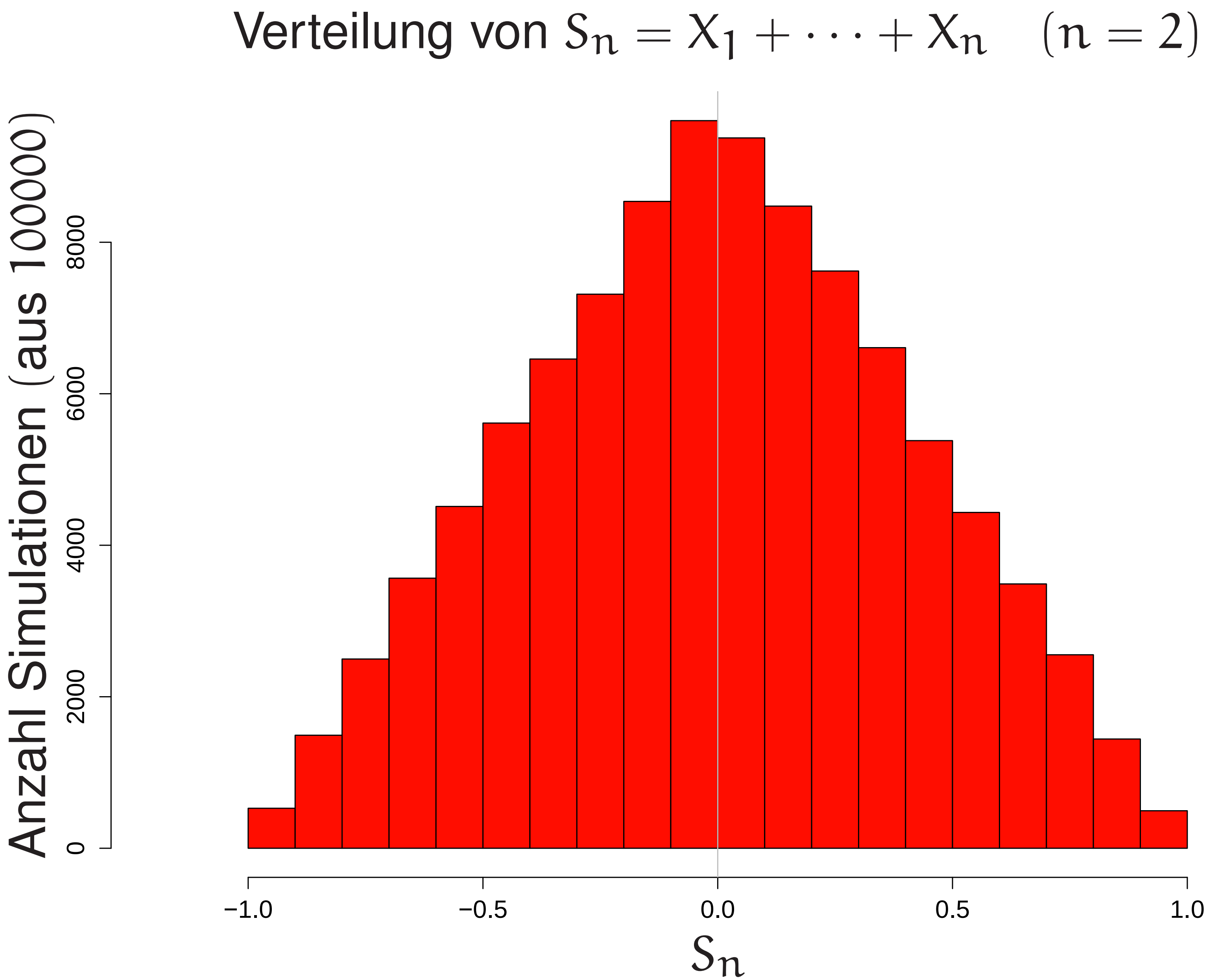
$$n = 15, 20, ..., 100$$

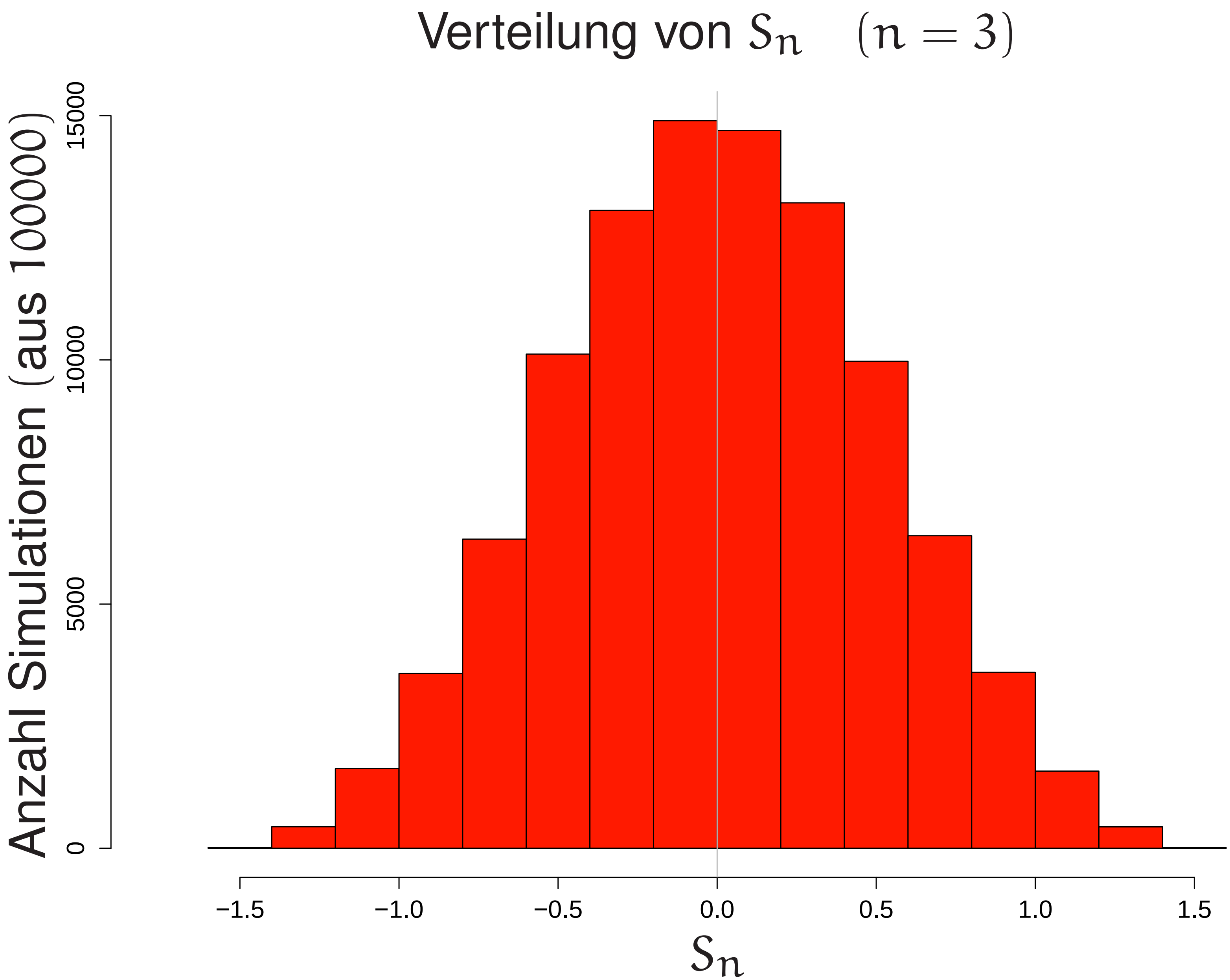


Dichtefunktion f_X der Verteilung von X
Aus dieser Verteilung wird 100000-mal
eine Stichprobe vom Umfang n gezogen.
$-\sigma$
$\mathbf{E}X$
σ
$f_X(x)$
x
0.0
0.5
1.0
1.5
2.0
−0.5
0.0
0.5

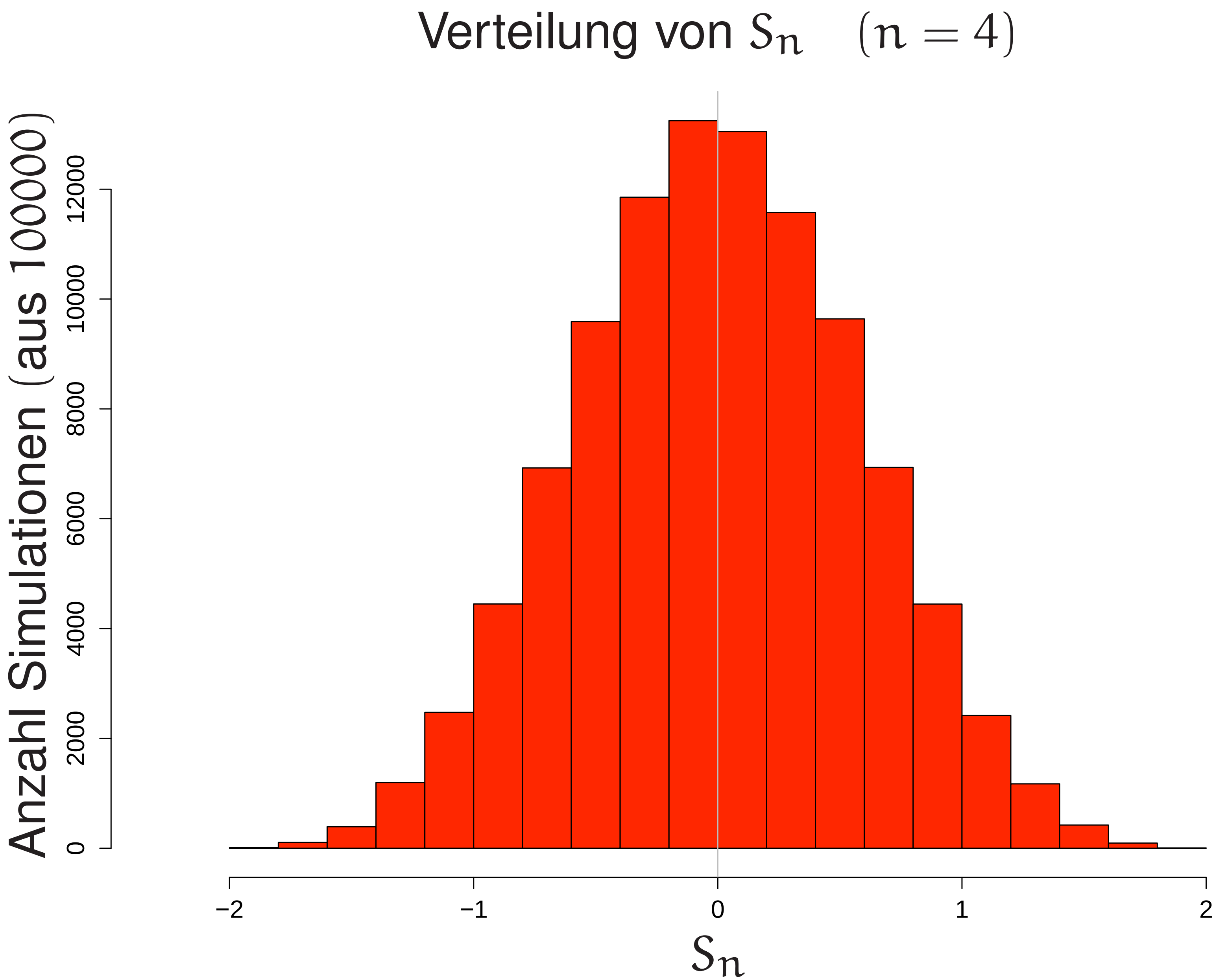


Verteilung von $S_1 = X_1 \quad (n = 1)$
Anzahl Simulationen (aus 100000)
0
1000
2000
3000
4000
5000
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
S_n

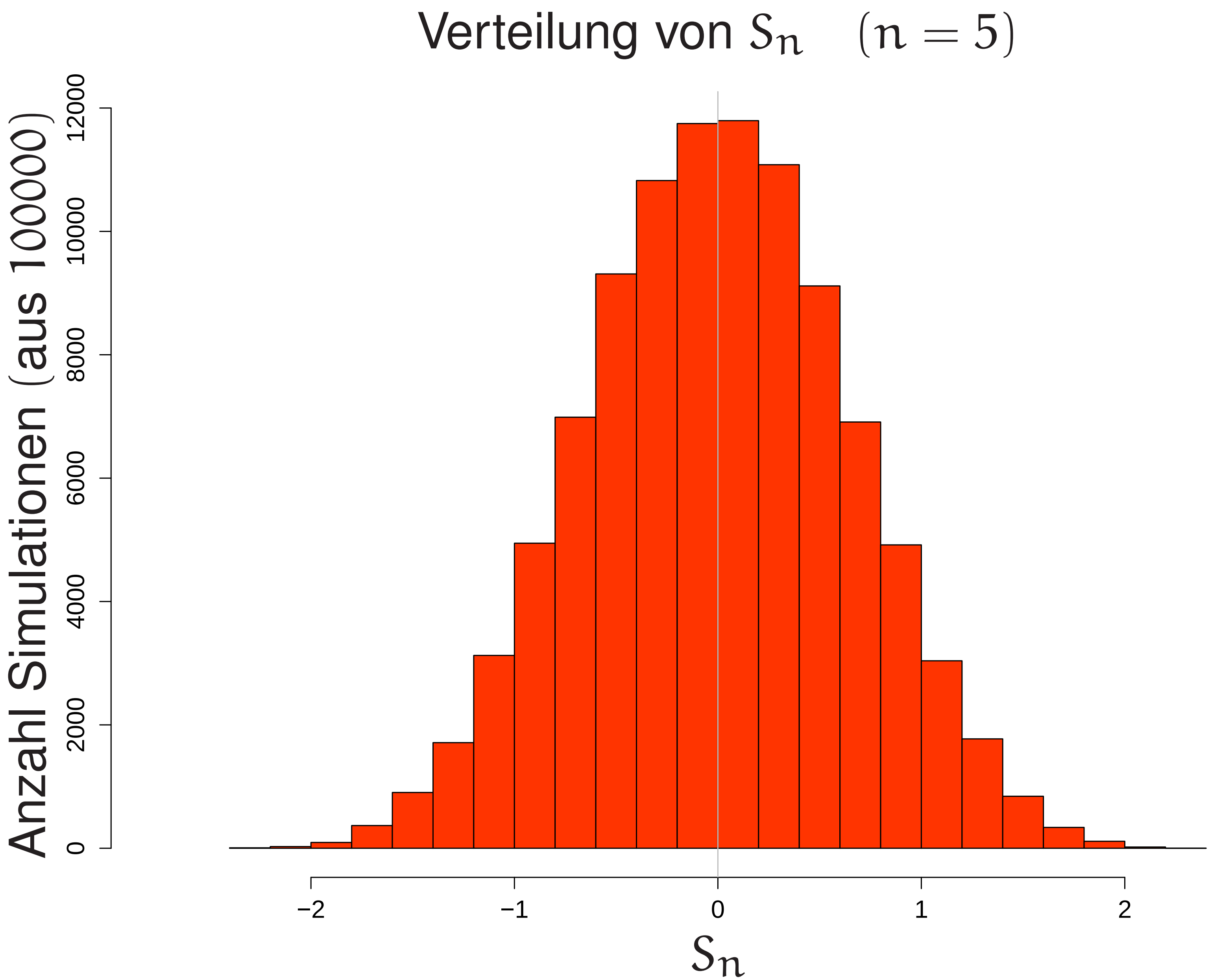

Verteilung von $S_n = X_1 + \cdots + X_n \quad (n = 2)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−1.0
−0.5
0.0
0.5
1.0
S_n



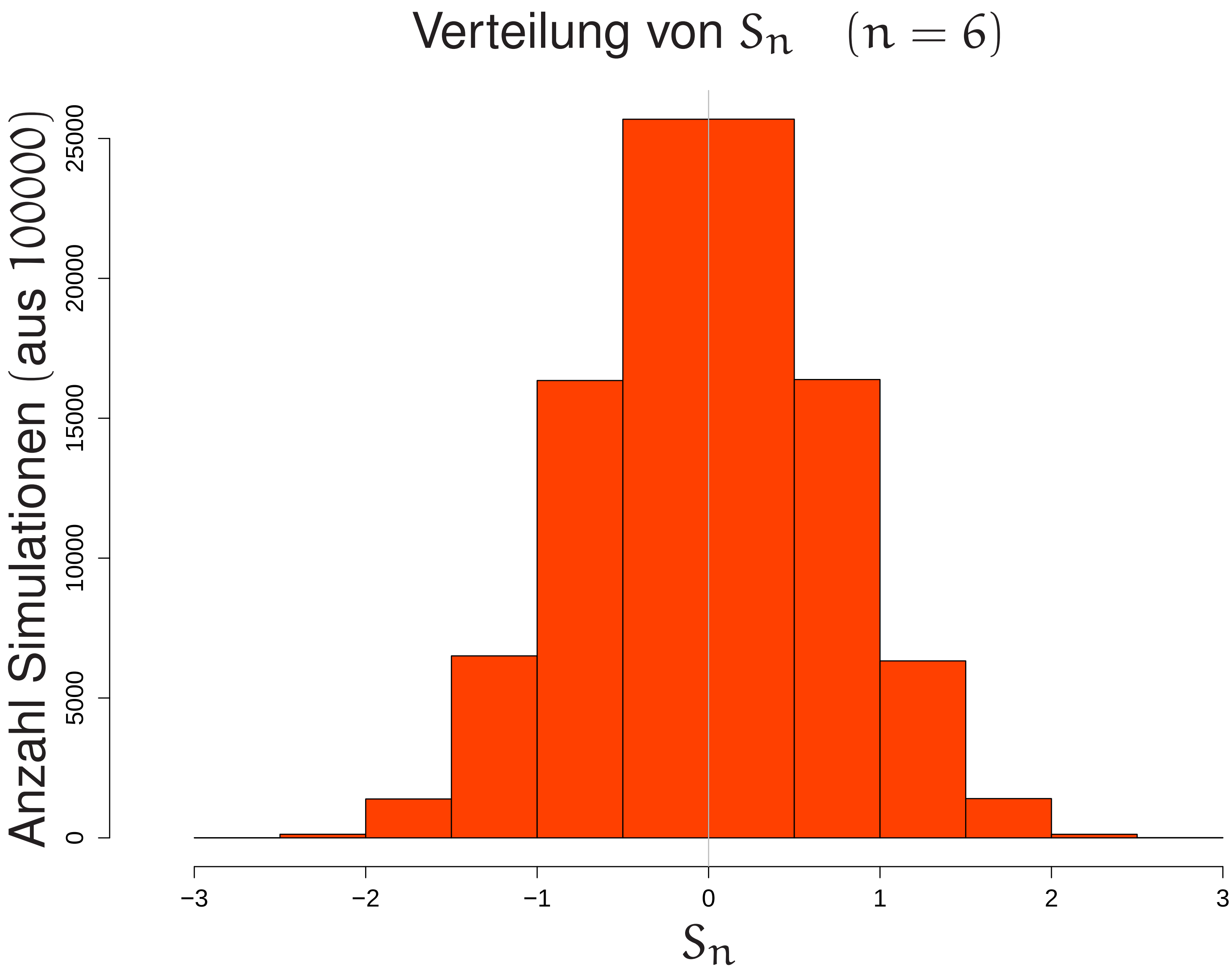
Verteilung von S_n $(n = 3)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
S_n



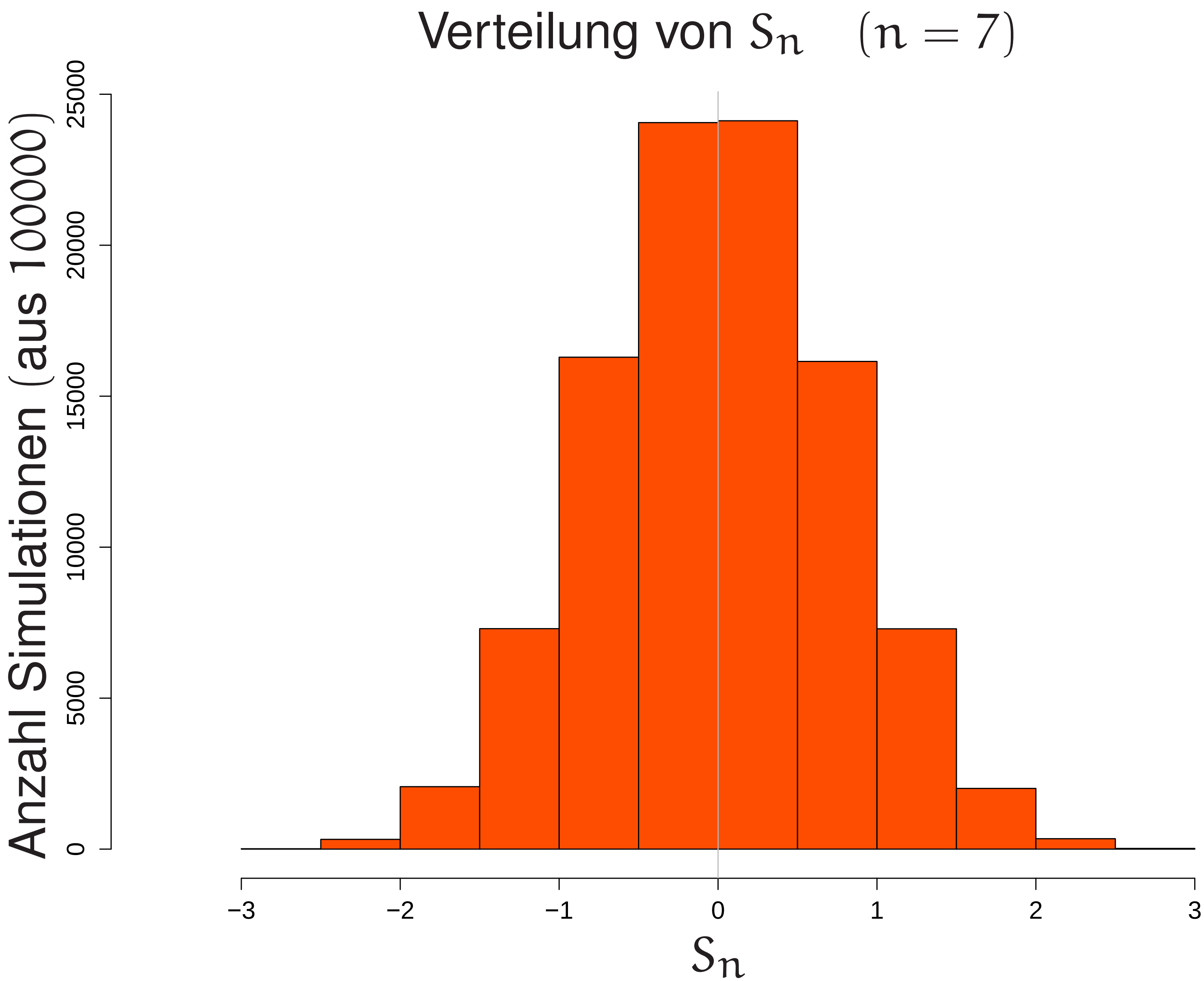
Verteilung von S_n $(n = 4)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
10000
12000
−2
−1
0
1
2
S_n



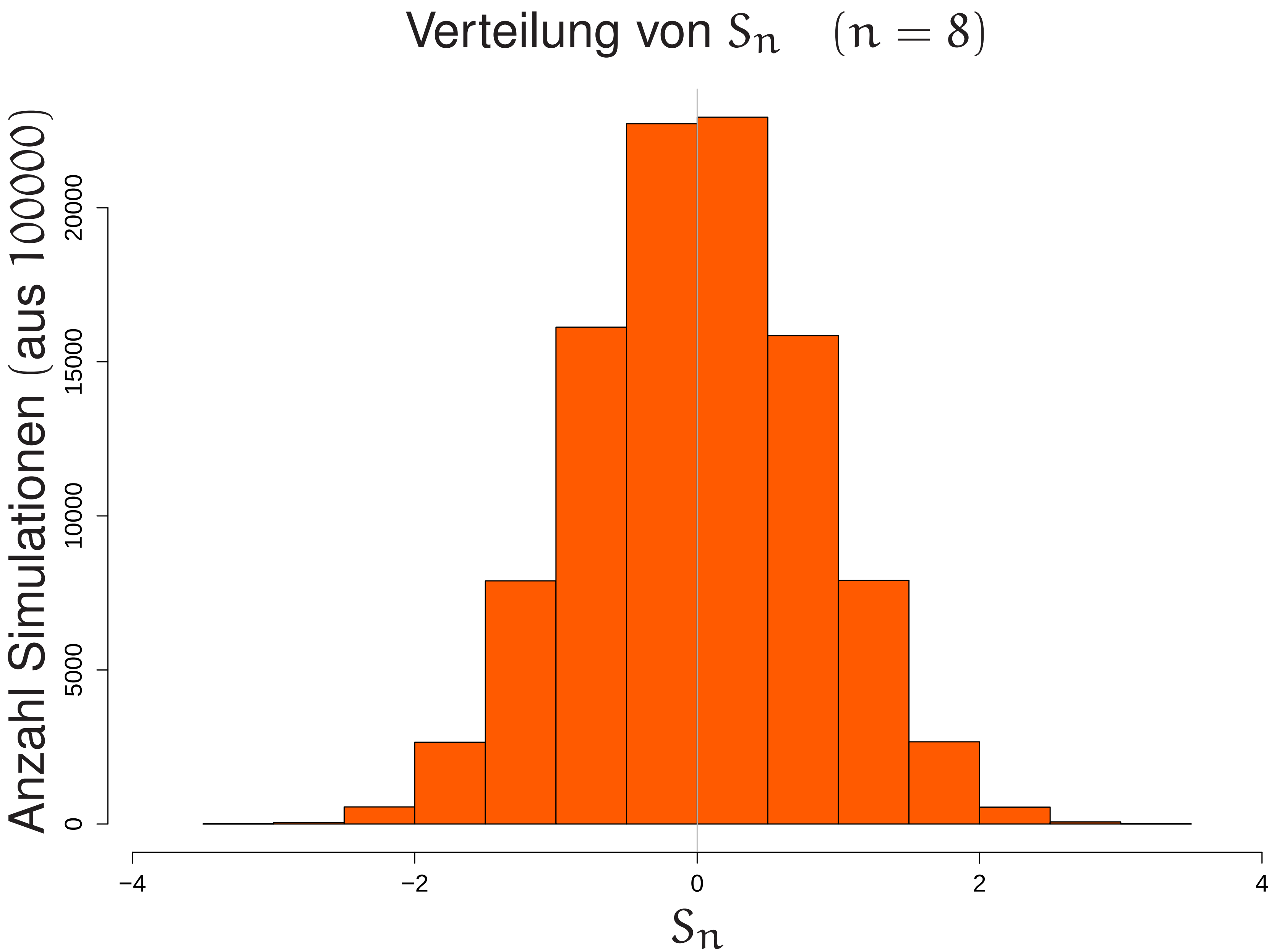
Verteilung von $S_n \quad (n = 5)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
10000
12000
−2
−1
0
1
2
S_n

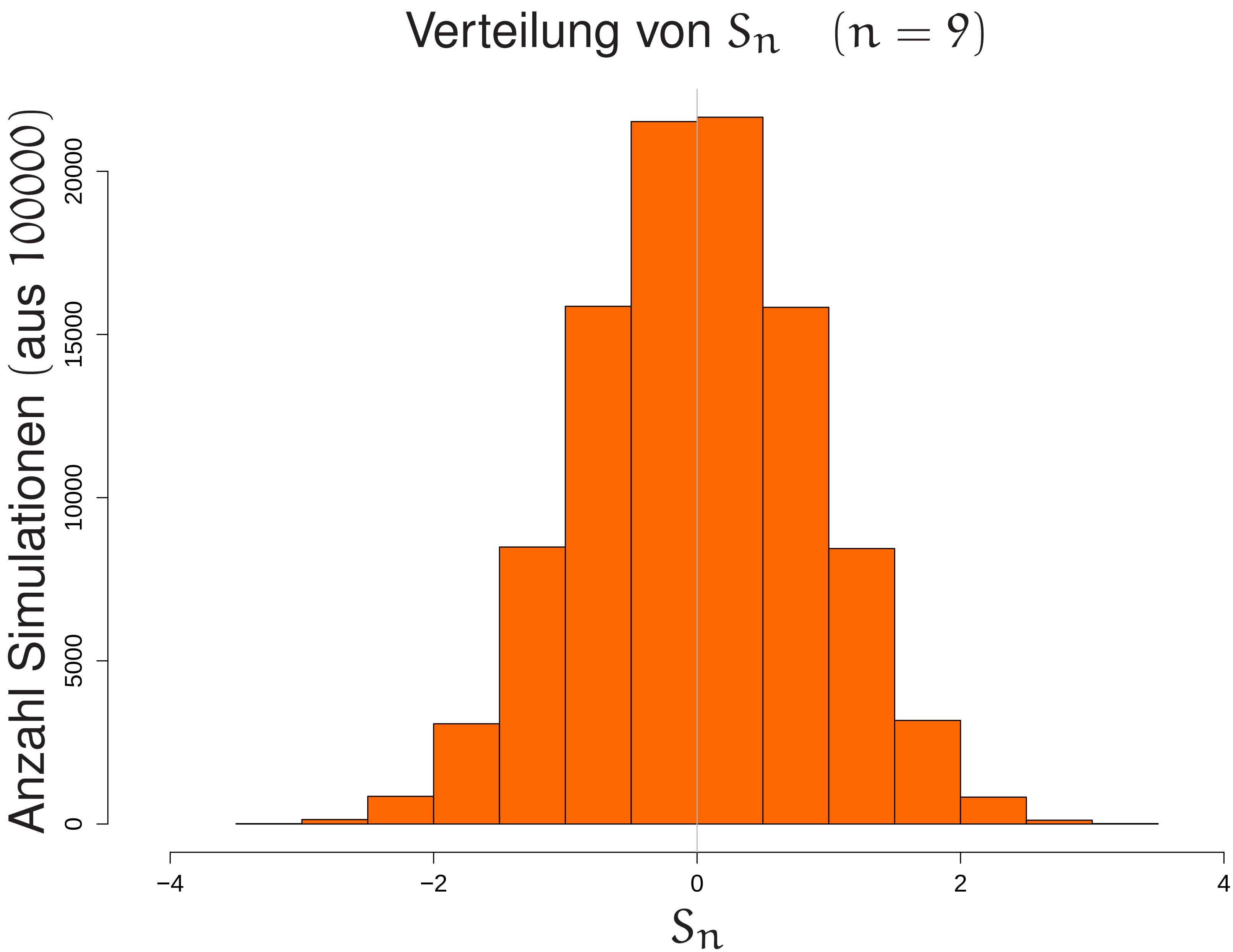


Verteilung von S_n $(n = 6)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
20000
25000
−3
−2
−1
0
1
2
3
S_n

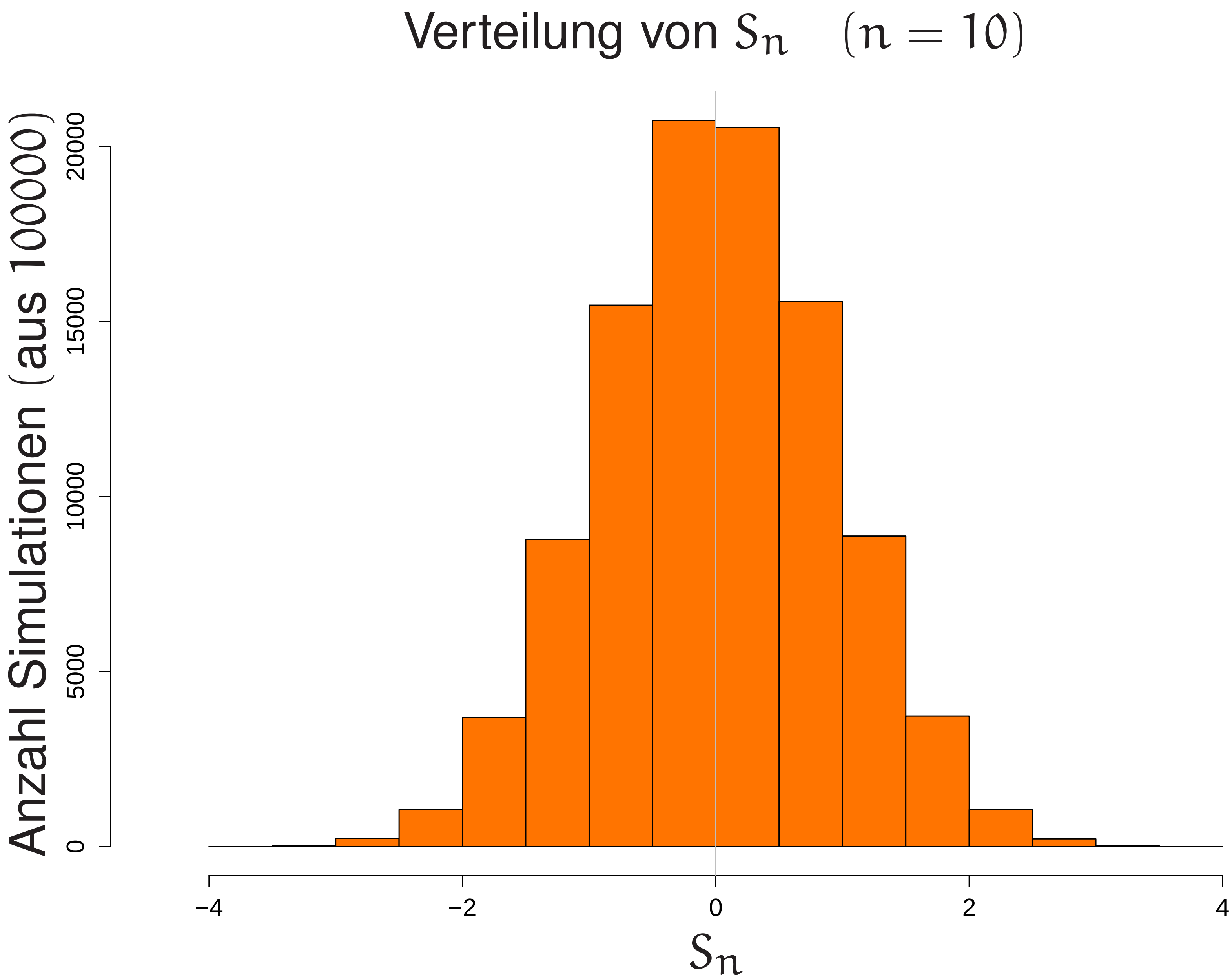


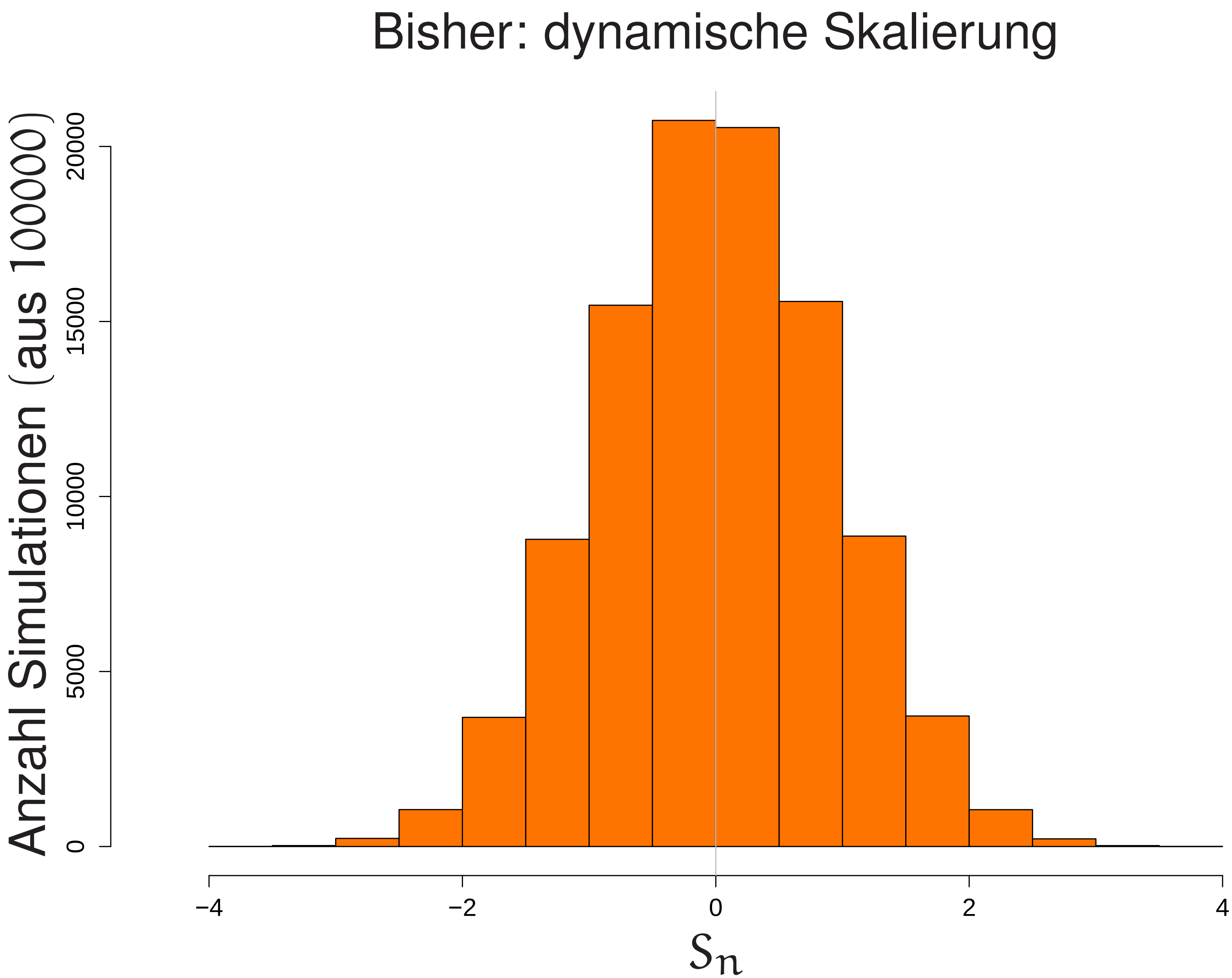
Verteilung von $S_n \quad (n = 7)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
20000
25000
−3
−2
−1
0
1
2
3
S_n

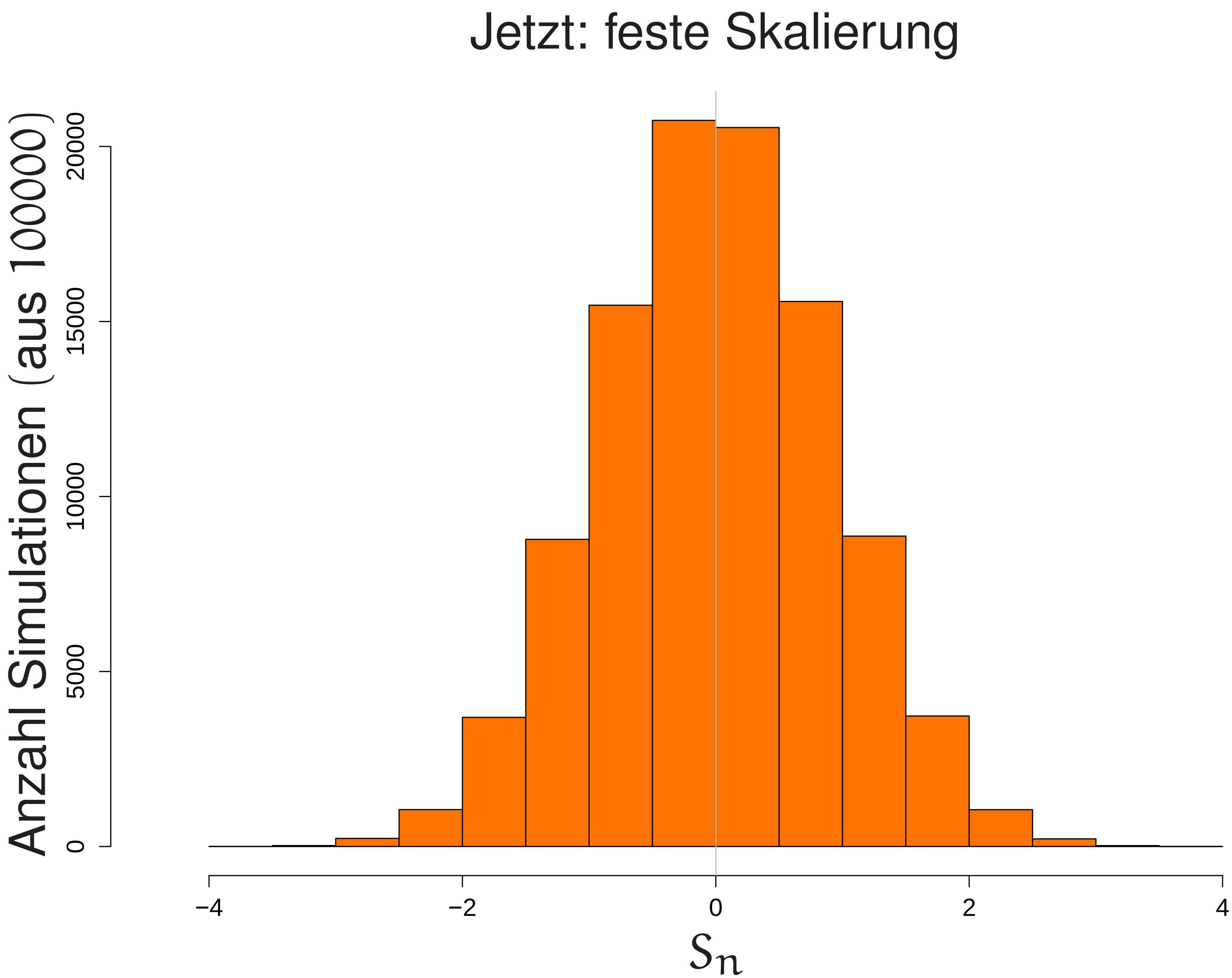

Verteilung von S_n $(n = 8)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
20000
−4
−2
0
2
4
S_n



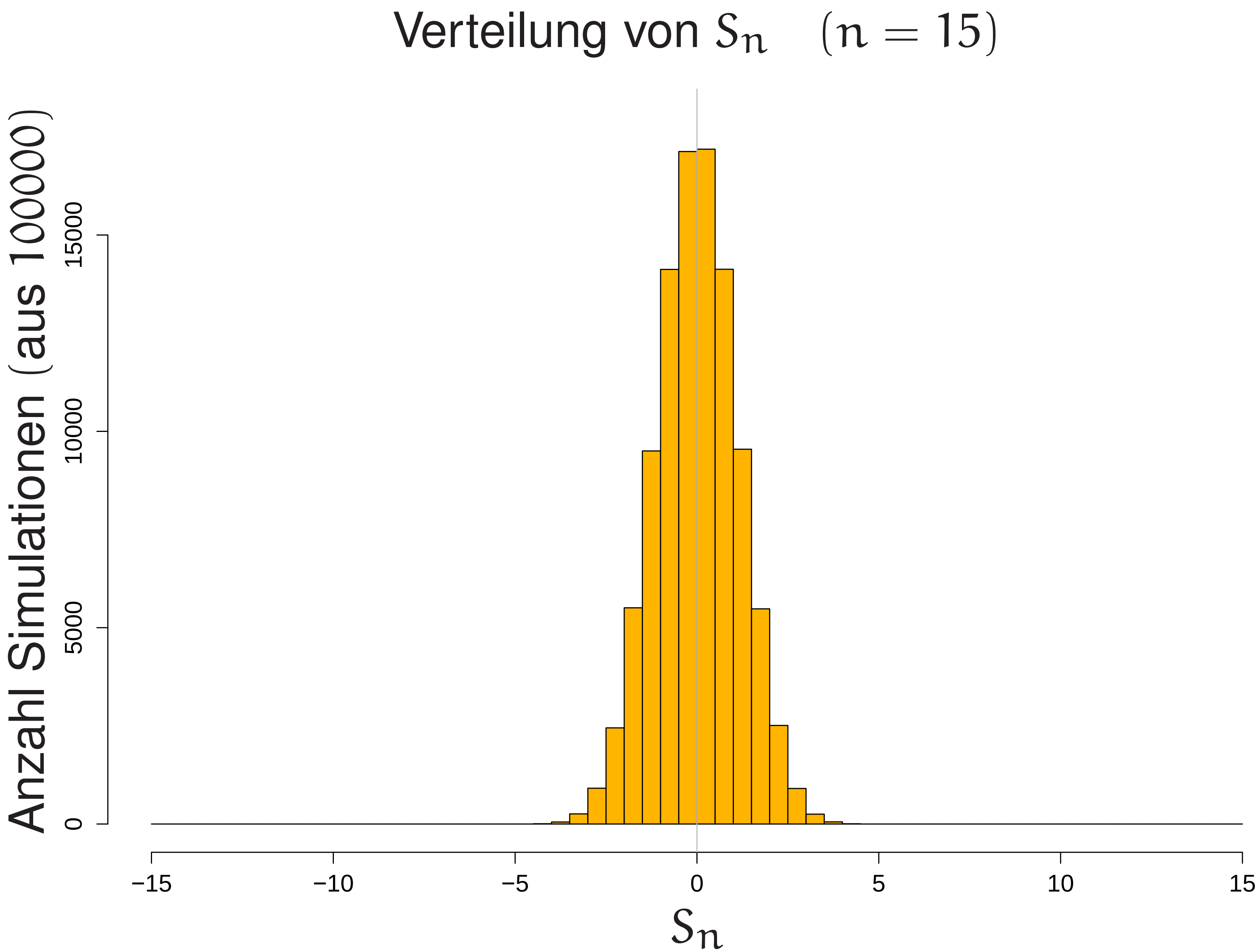
Verteilung von S_n $(n = 9)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
20000
−4
−2
0
2
4
S_n


Verteilung von S_n $(n = 10)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
20000
−4
−2
0
2
4
S_n

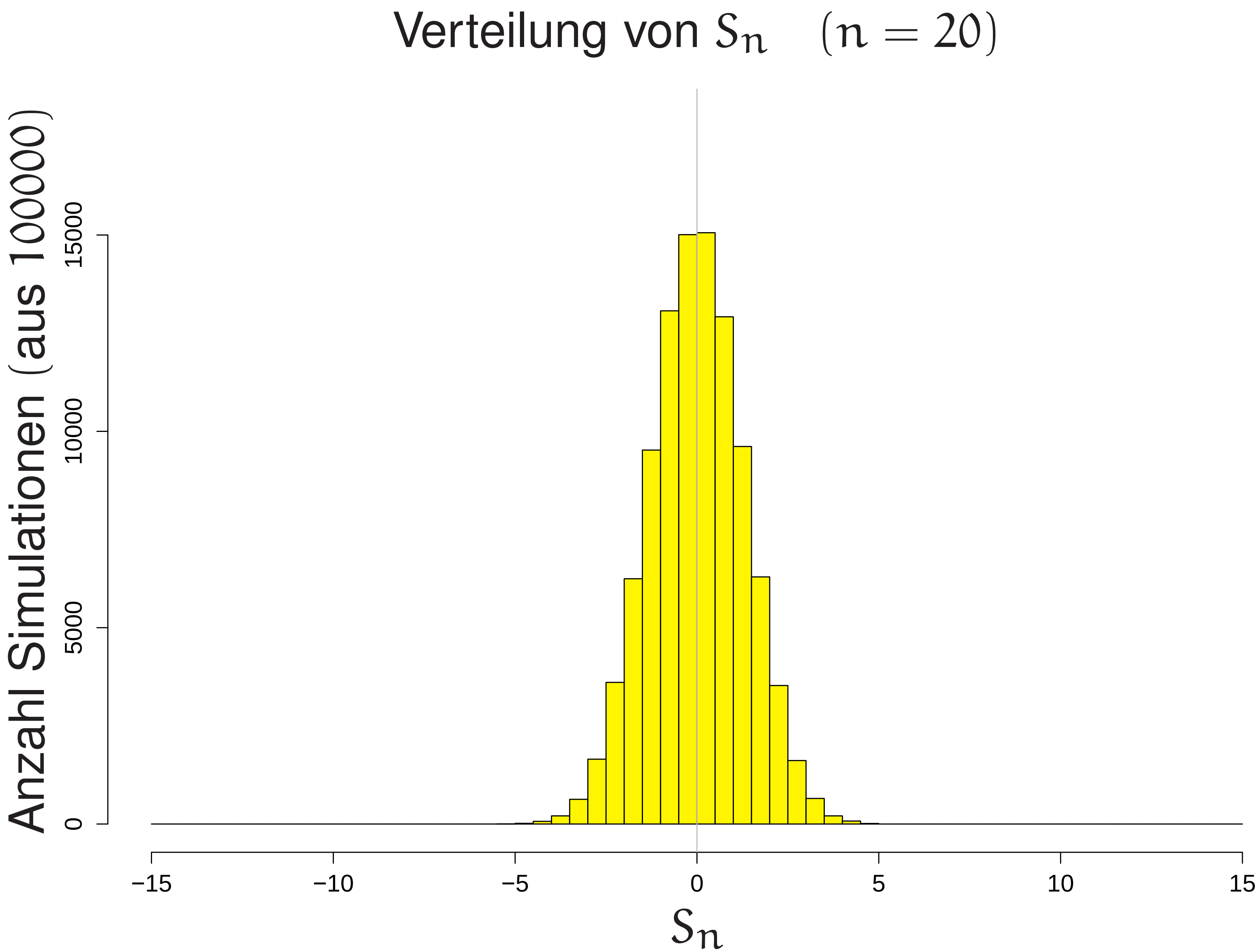

Bisher: dynamische Skalierung
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
20000
−4
−2
0
2
4
S_n



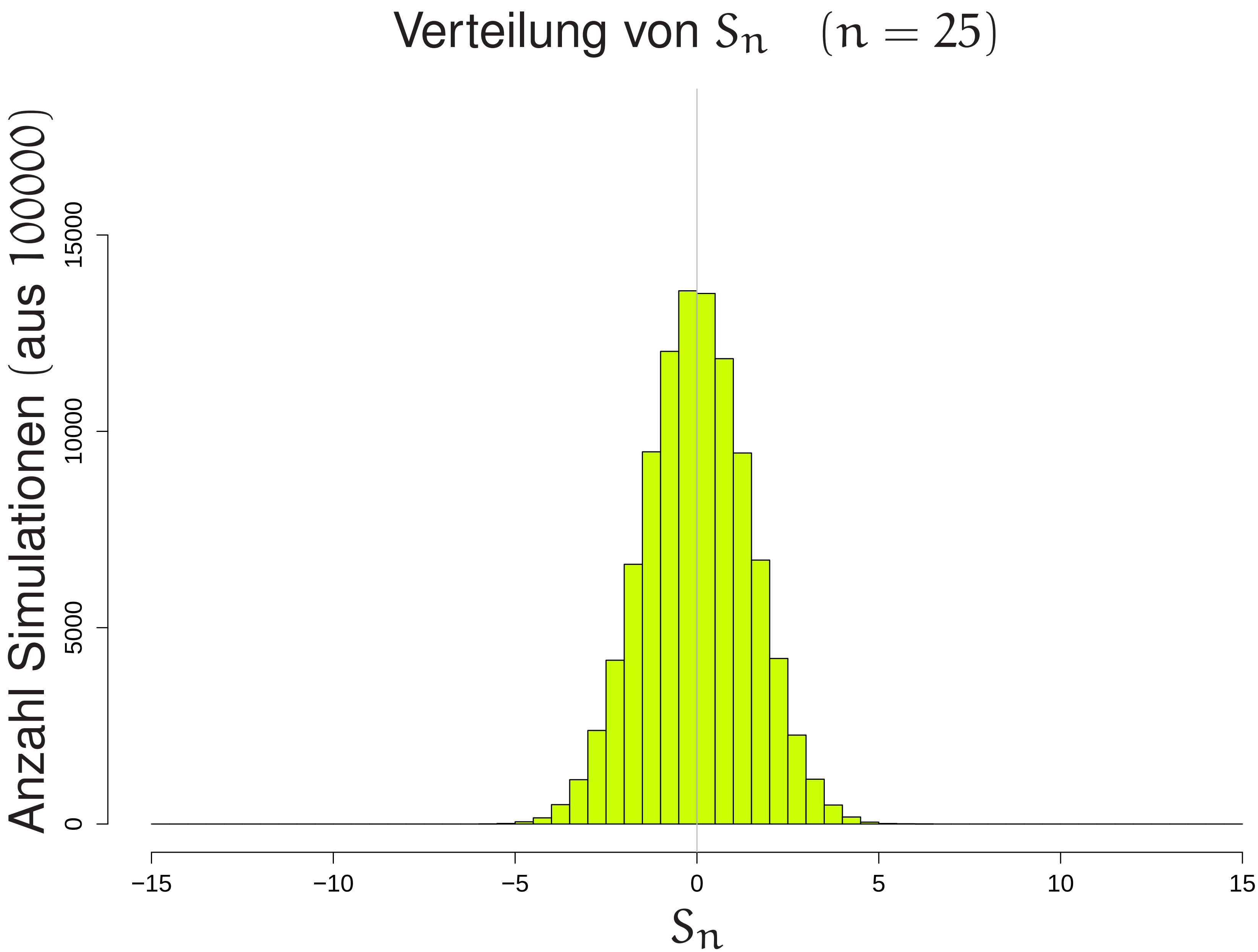
Jetzt: feste Skalierung
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
20000
−4
−2
0
2
4
S_n



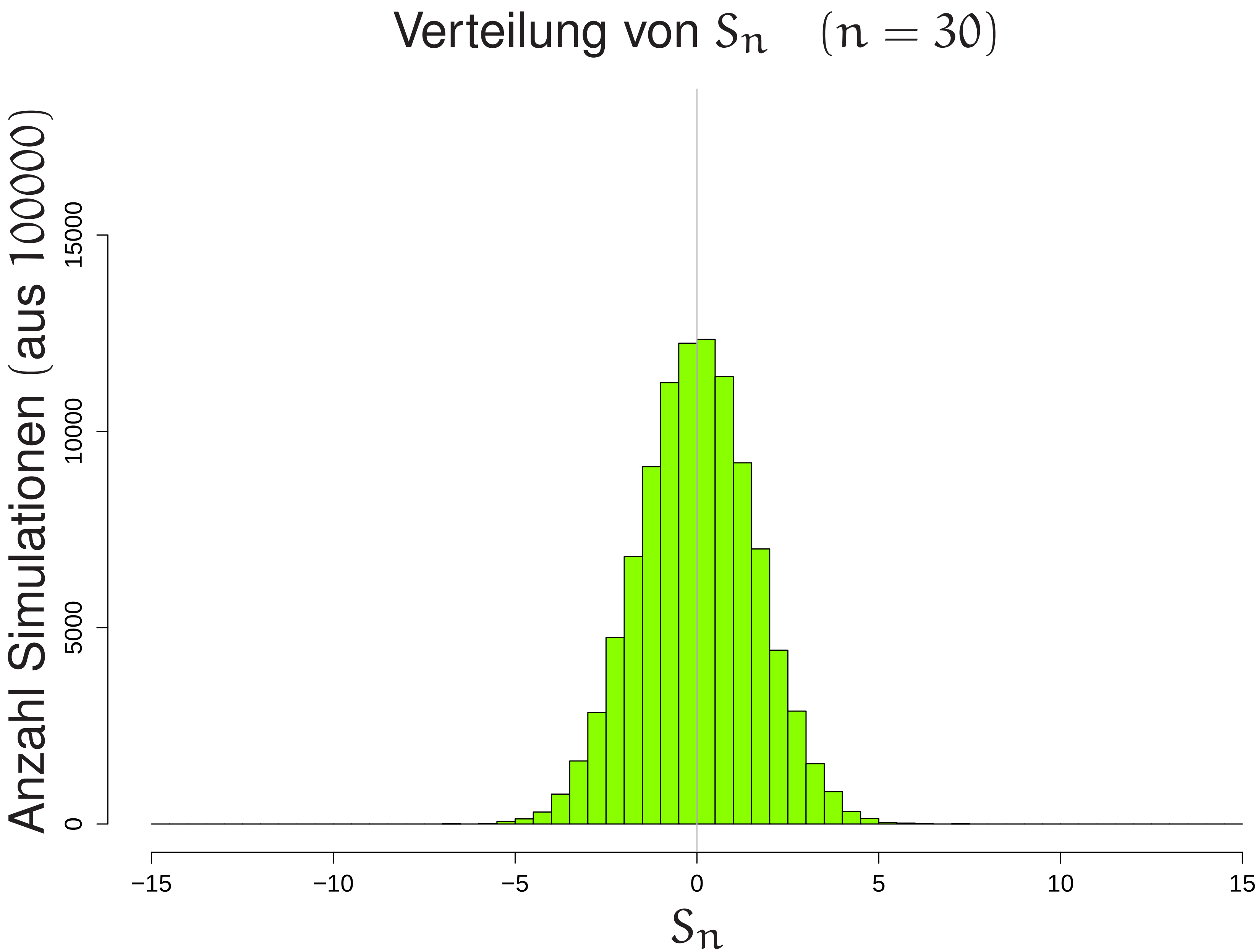
Verteilung von S_n $(n = 15)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n



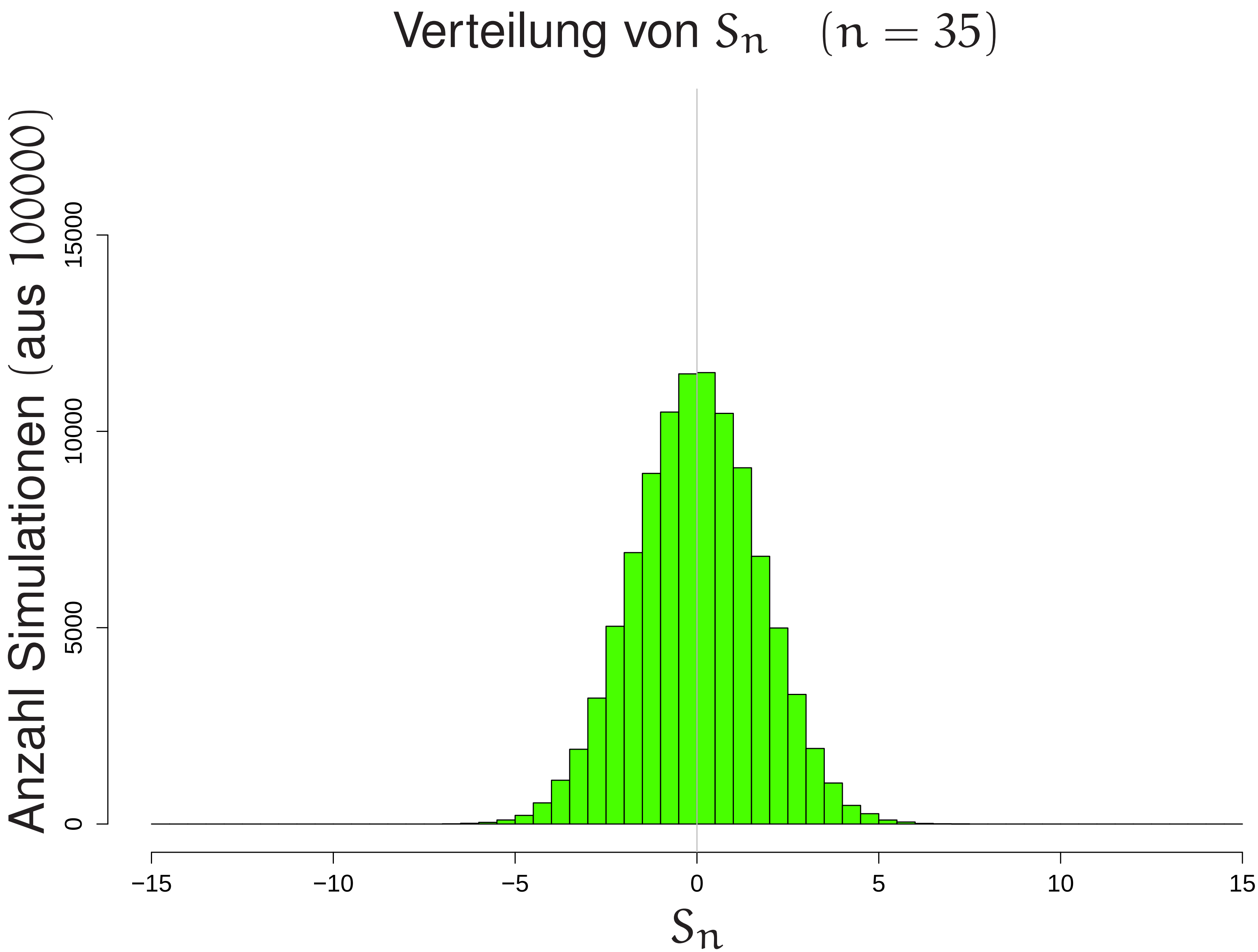
Verteilung von S_n $(n = 20)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n



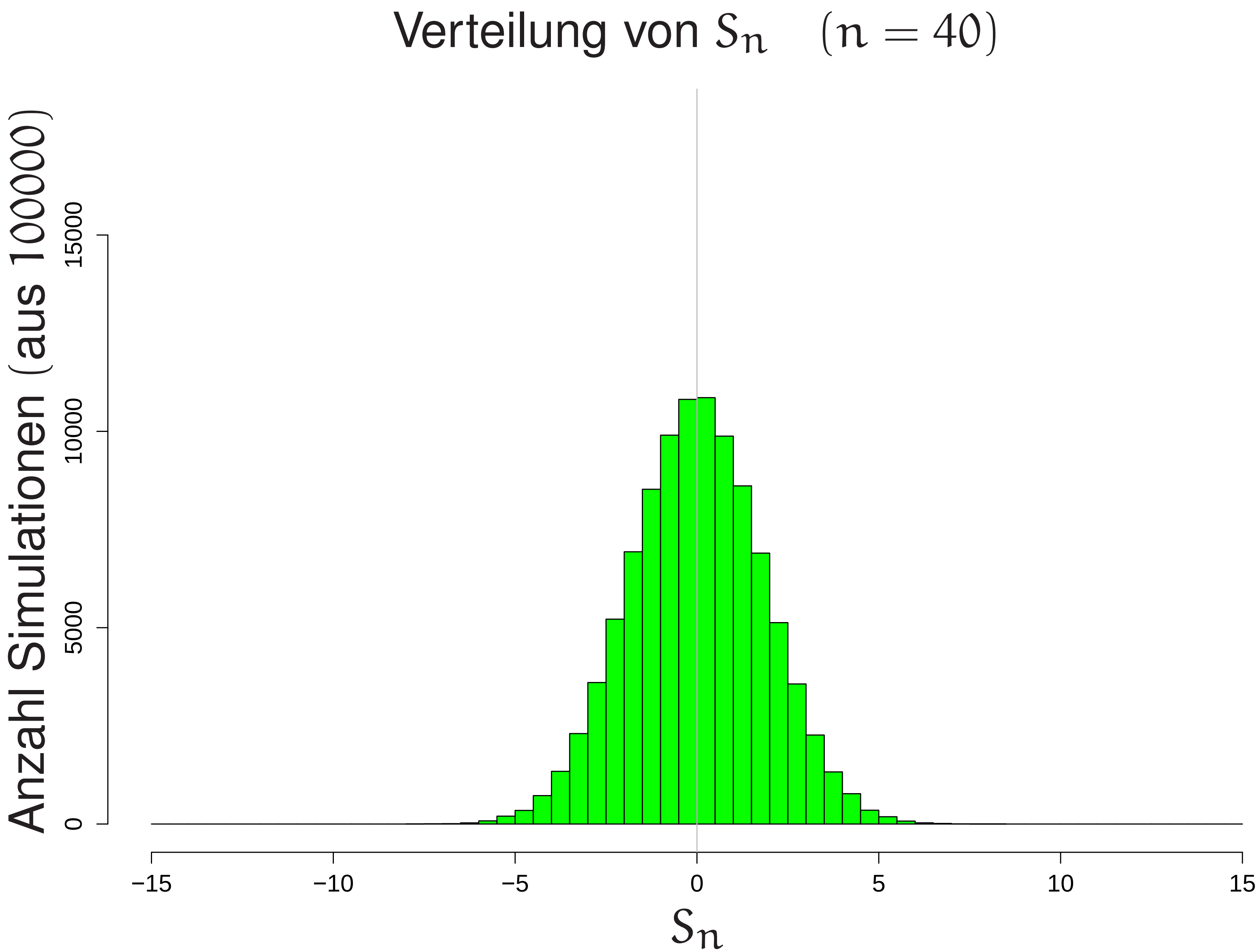
Verteilung von S_n $(n = 25)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n

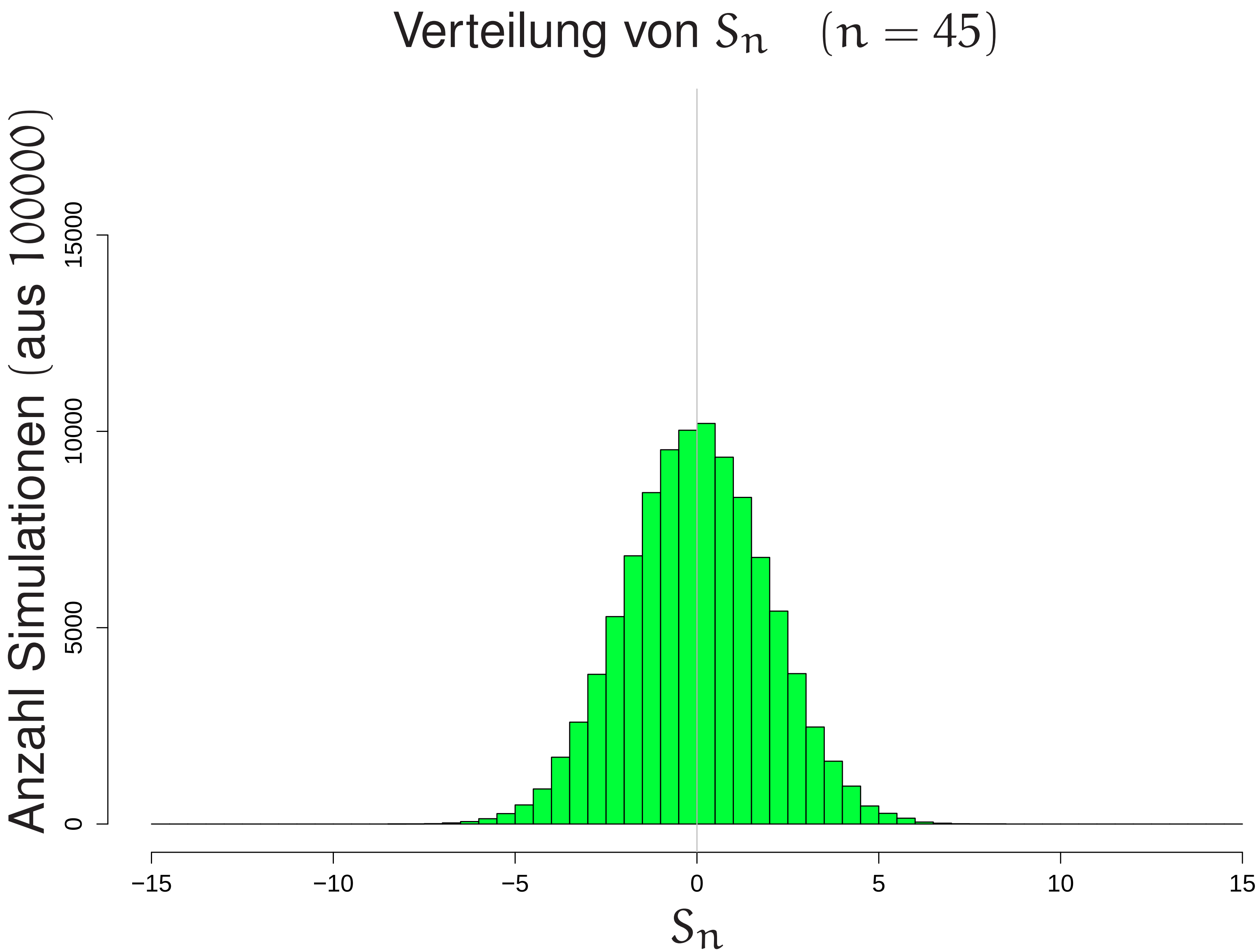


Verteilung von S_n $(n = 30)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n

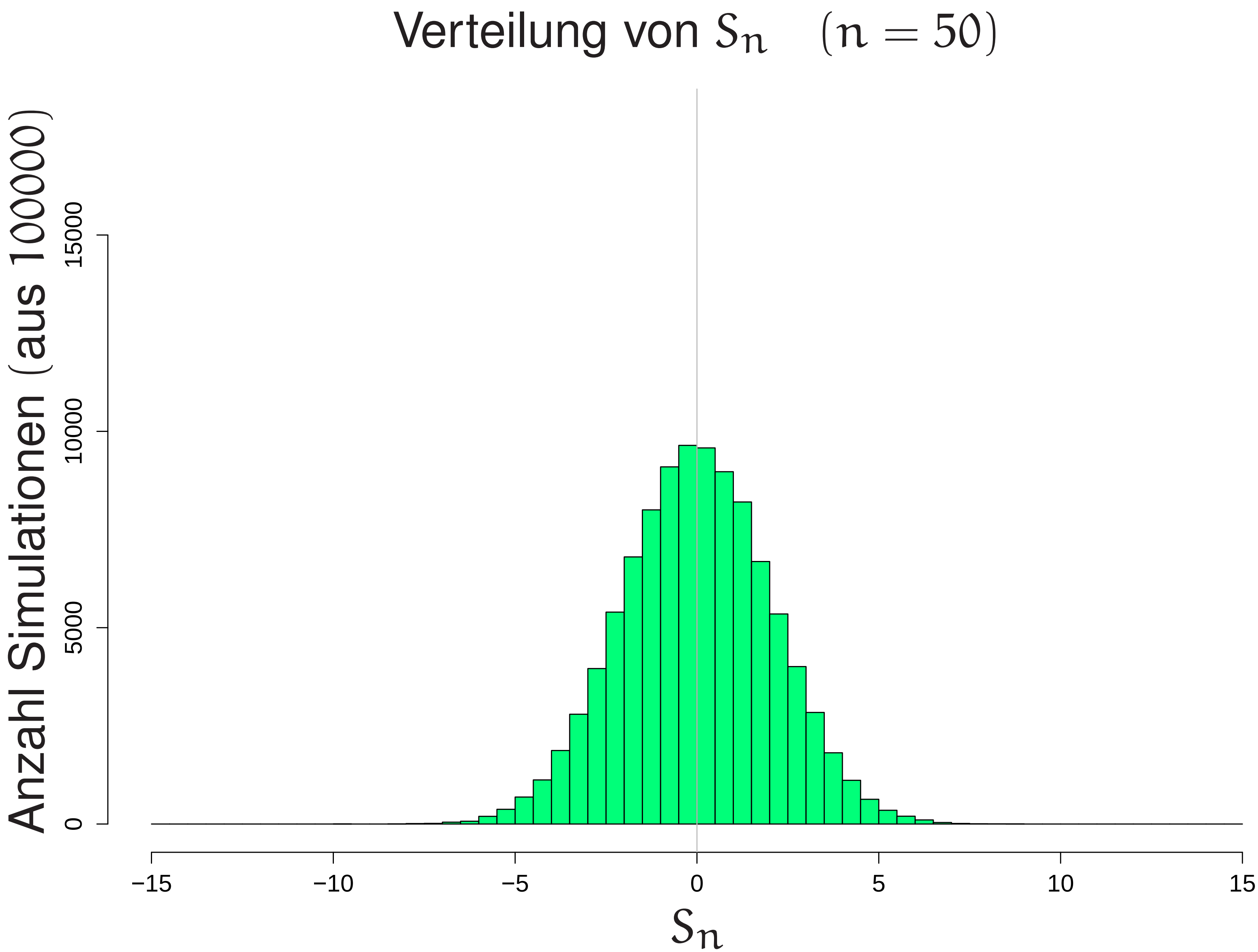


Verteilung von S_n $(n = 35)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n

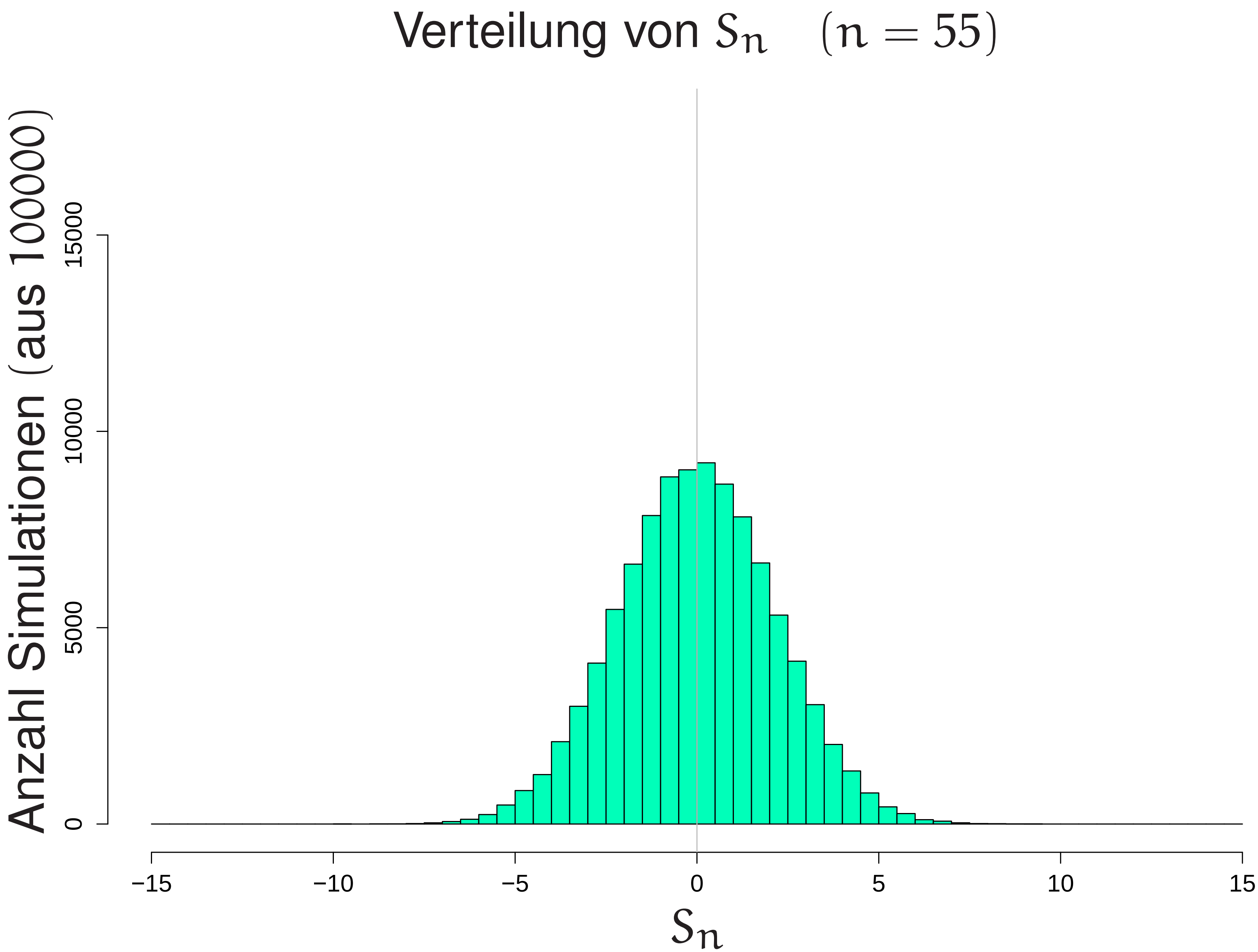

Verteilung von S_n $(n = 40)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n



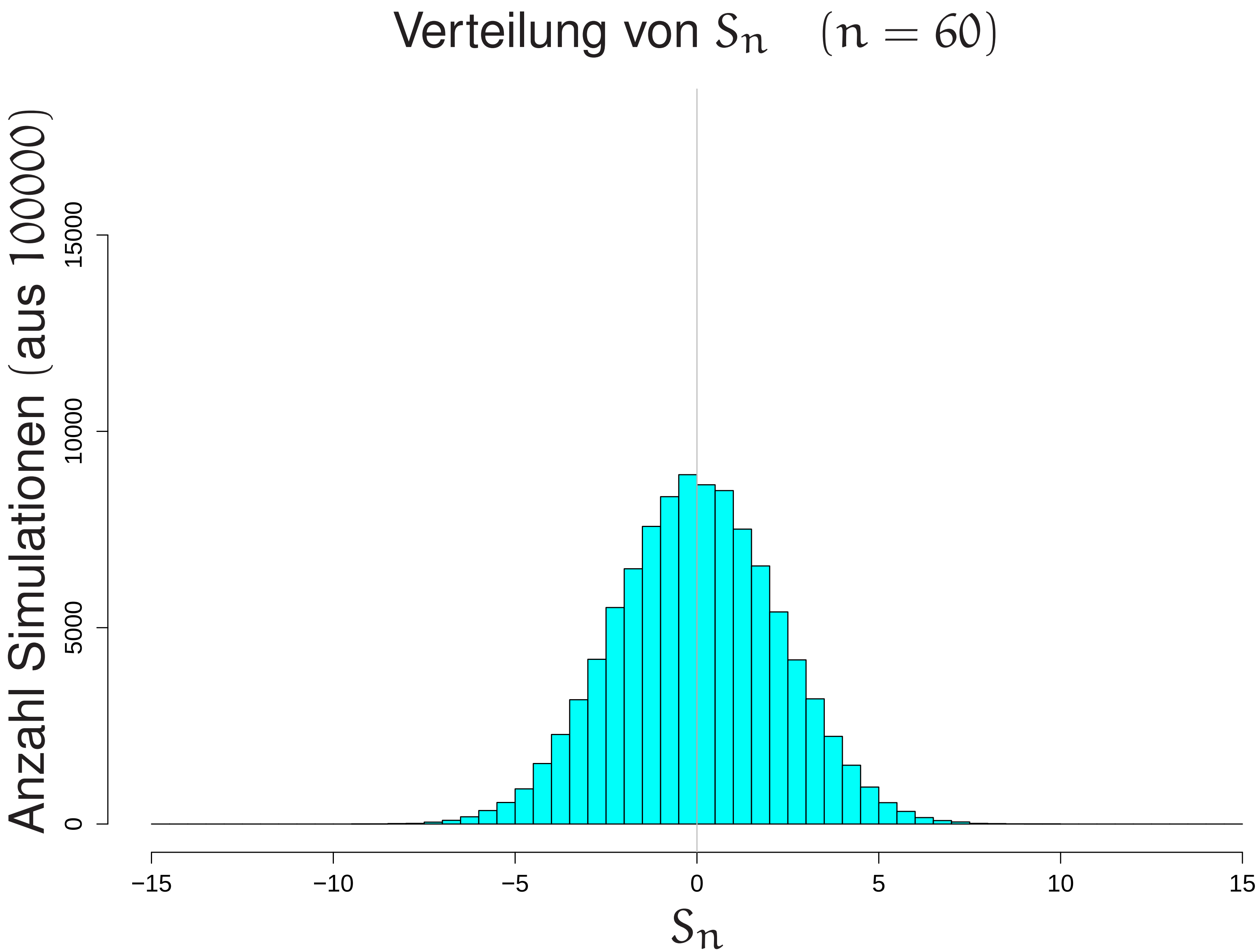
Verteilung von $S_n \quad (n = 45)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n



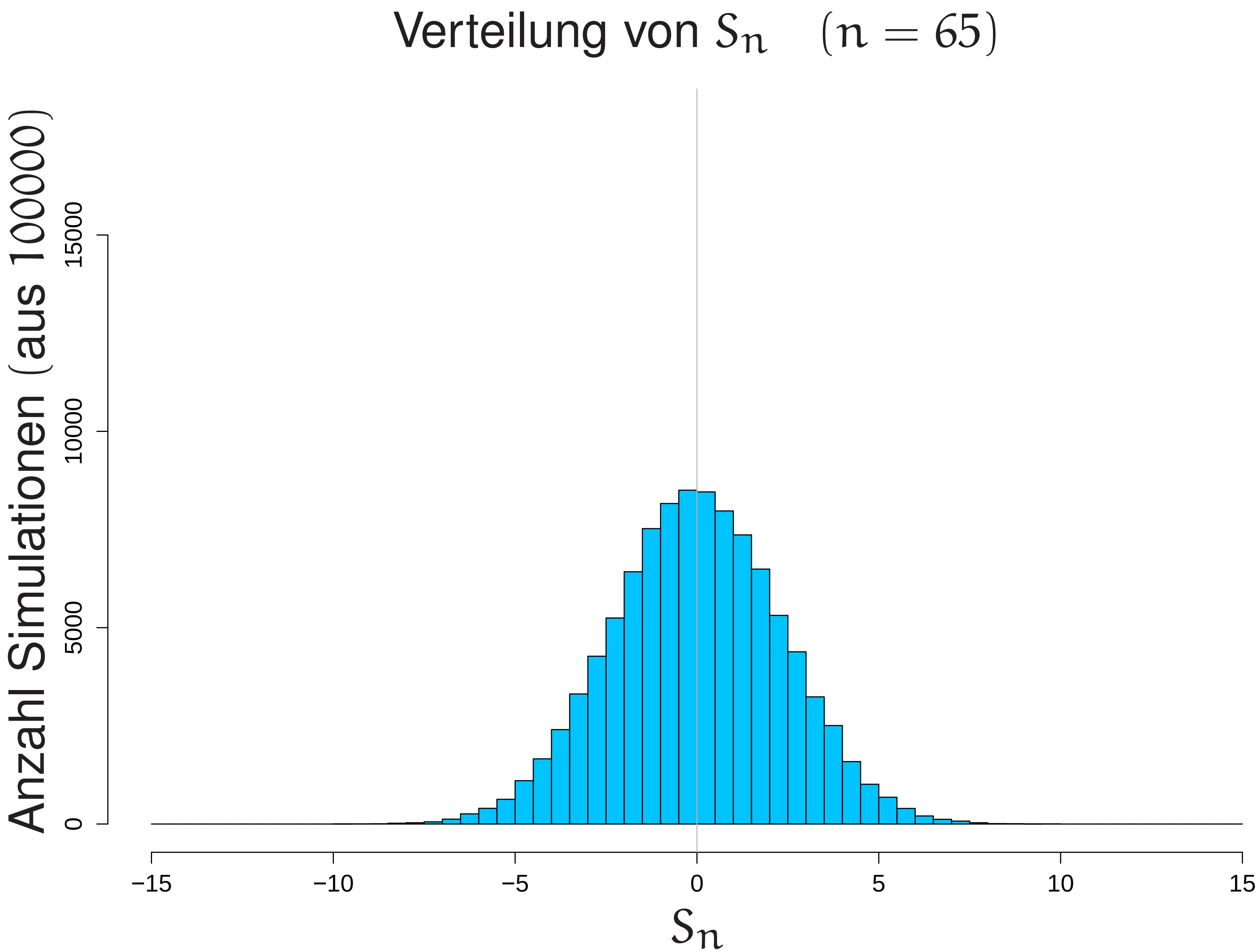
Verteilung von S_n $(n = 50)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n



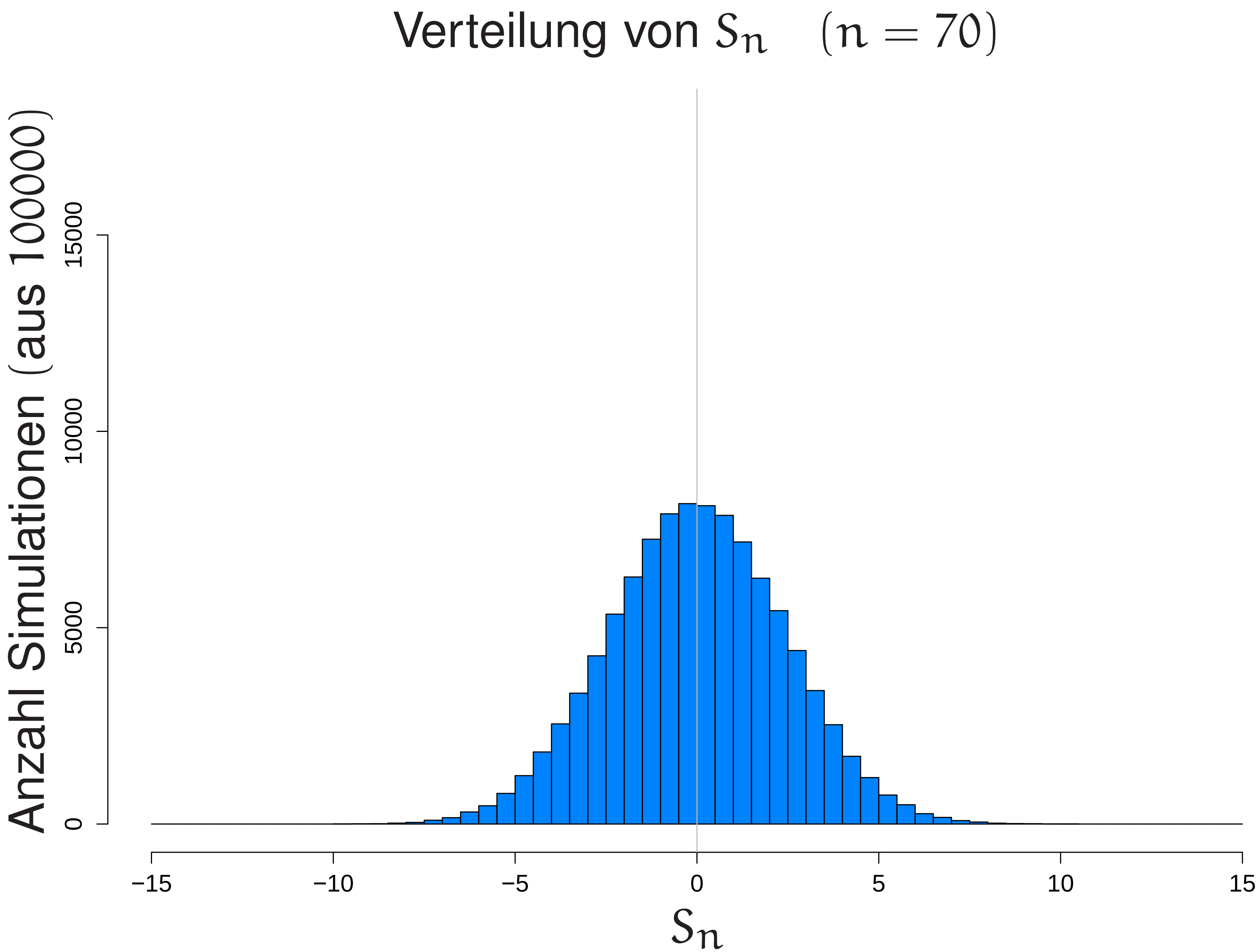
Verteilung von S_n $(n = 55)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n



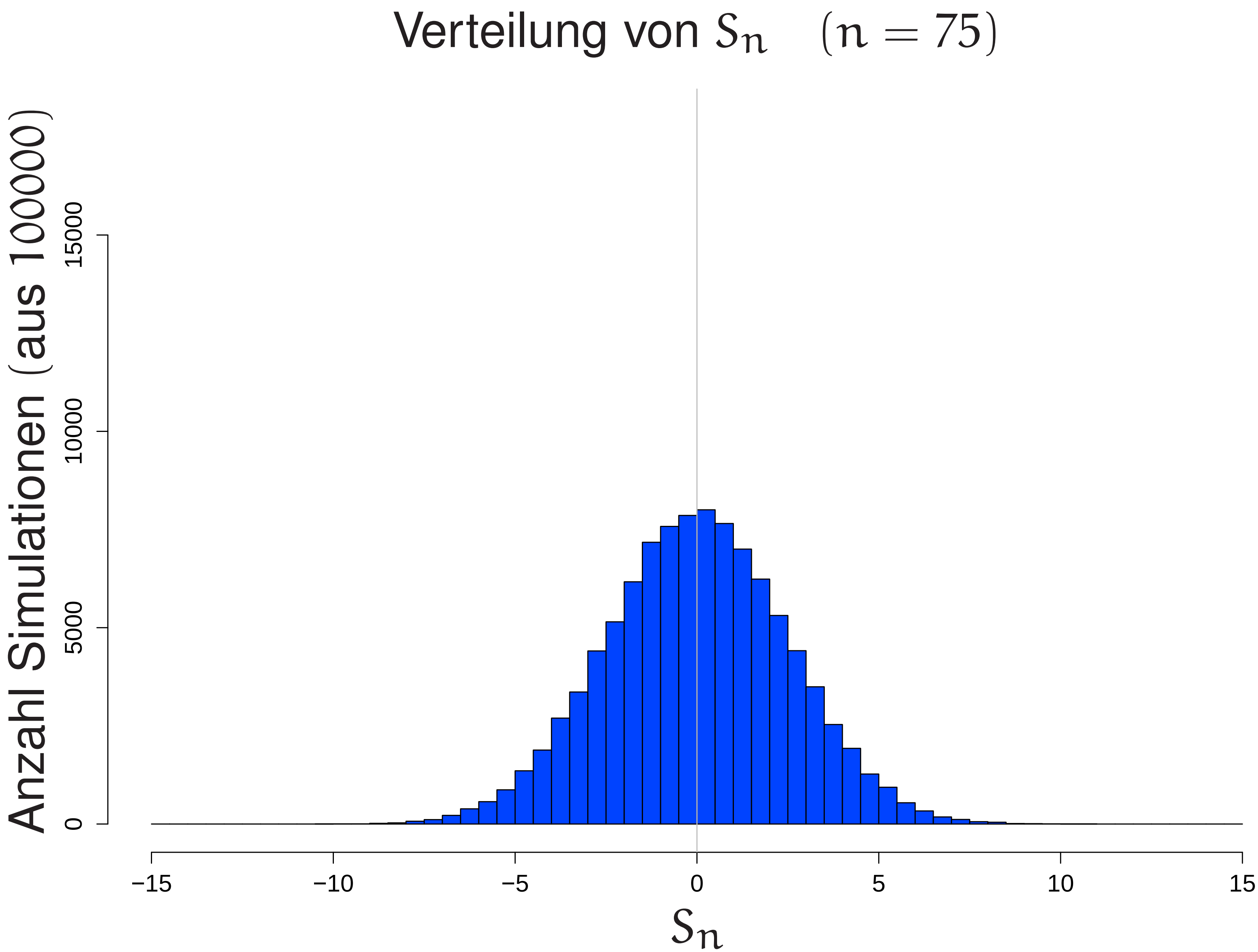
Verteilung von S_n $(n = 60)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n



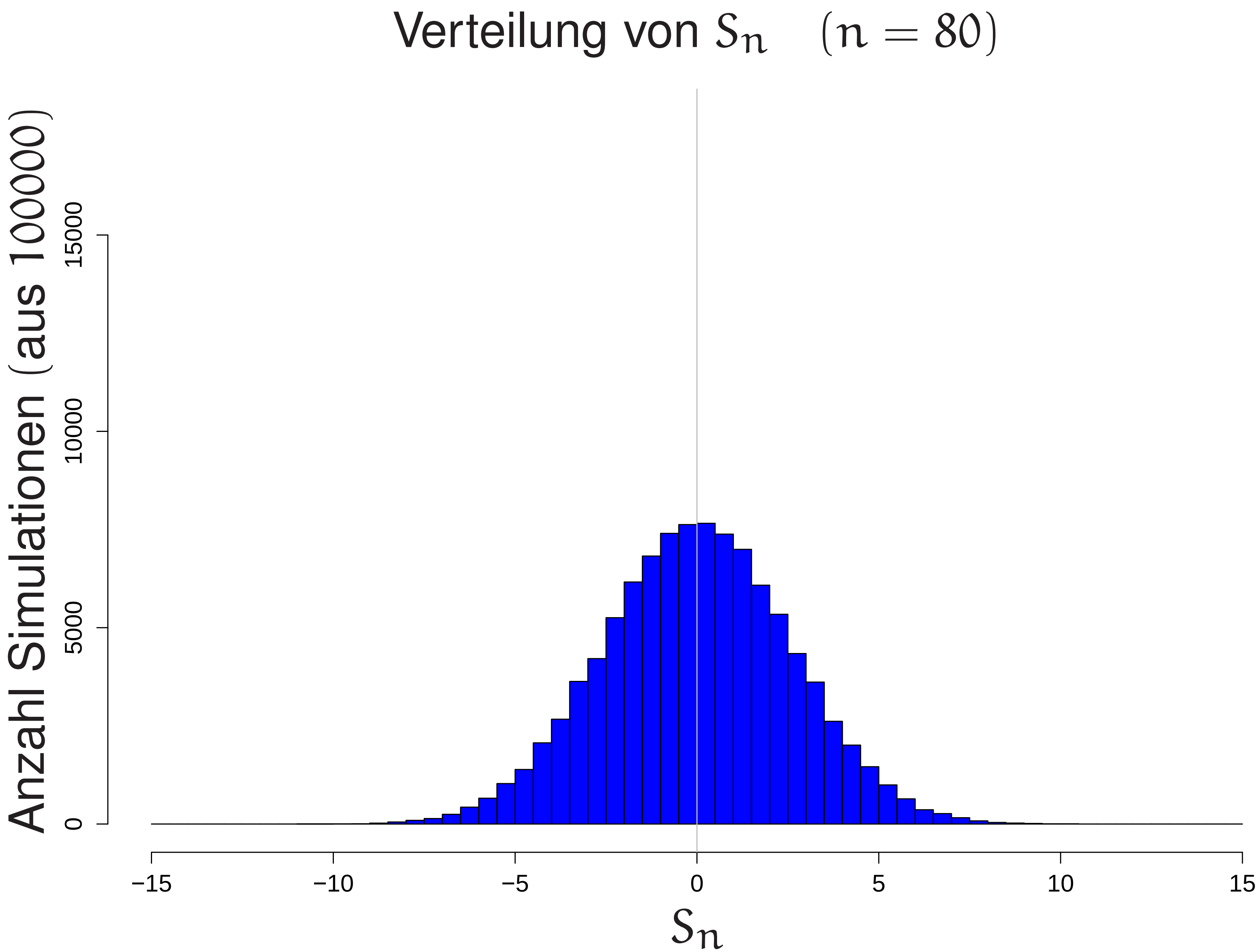
Verteilung von S_n $(n = 65)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n



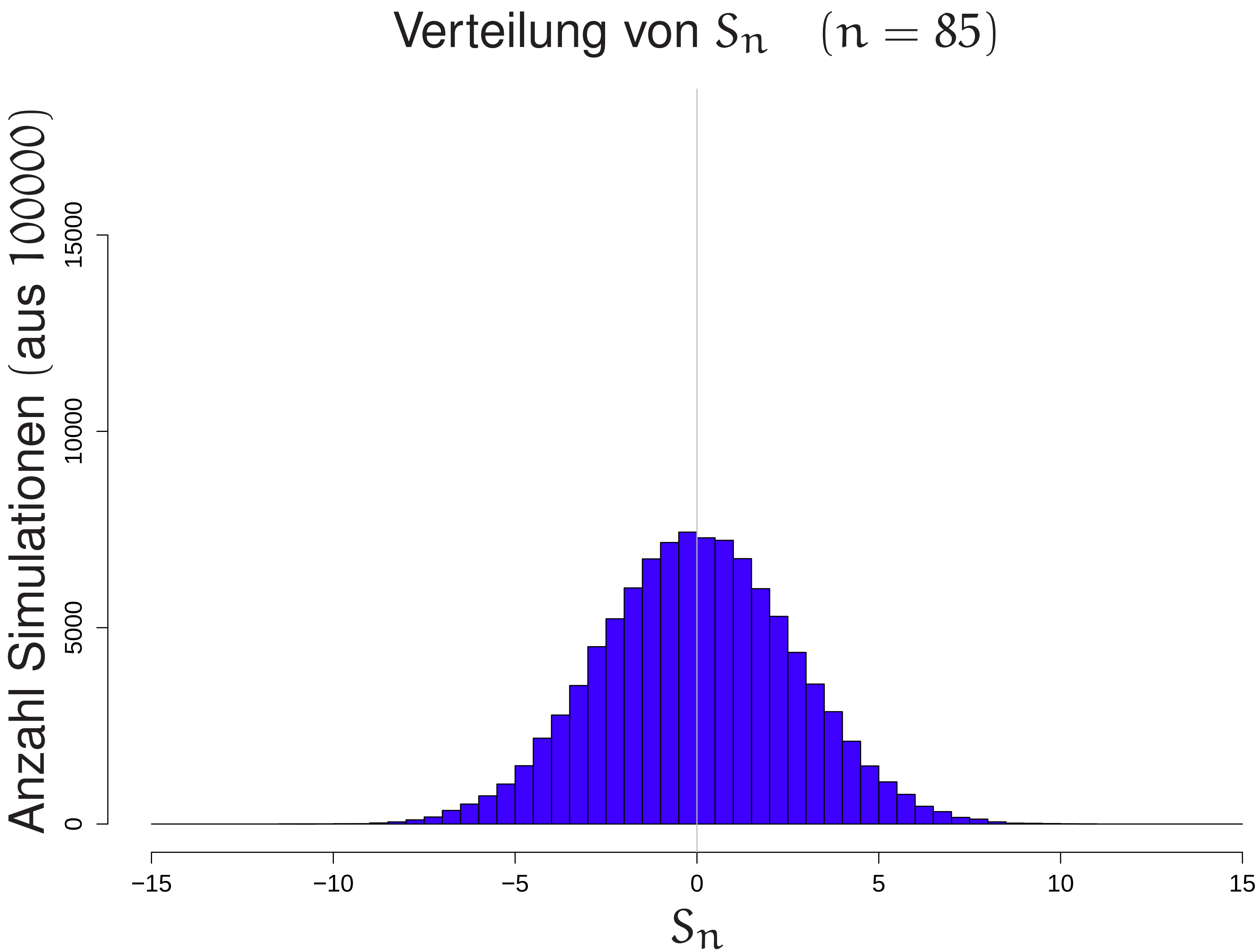
Verteilung von S_n $(n = 70)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n

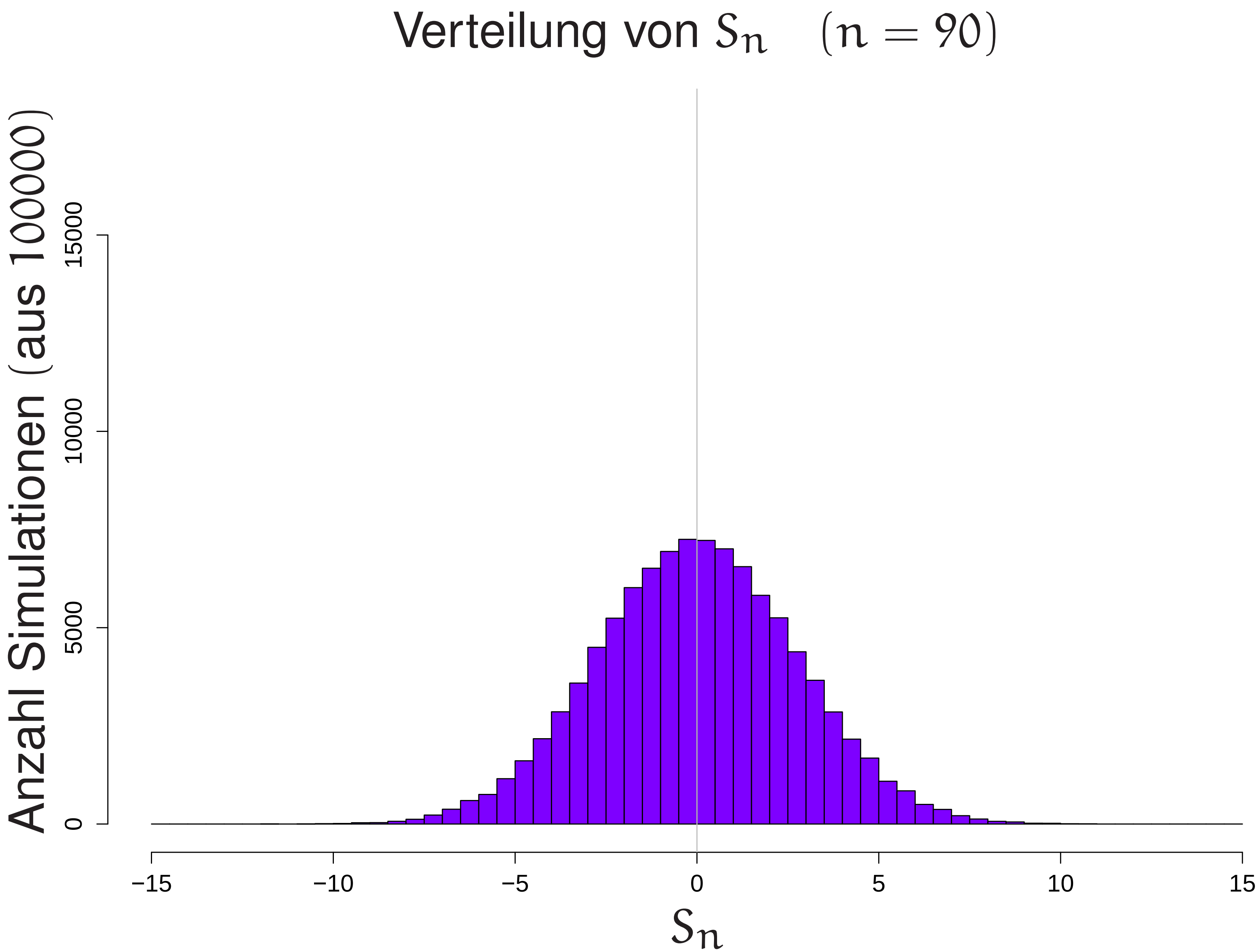


Verteilung von S_n $(n = 75)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n

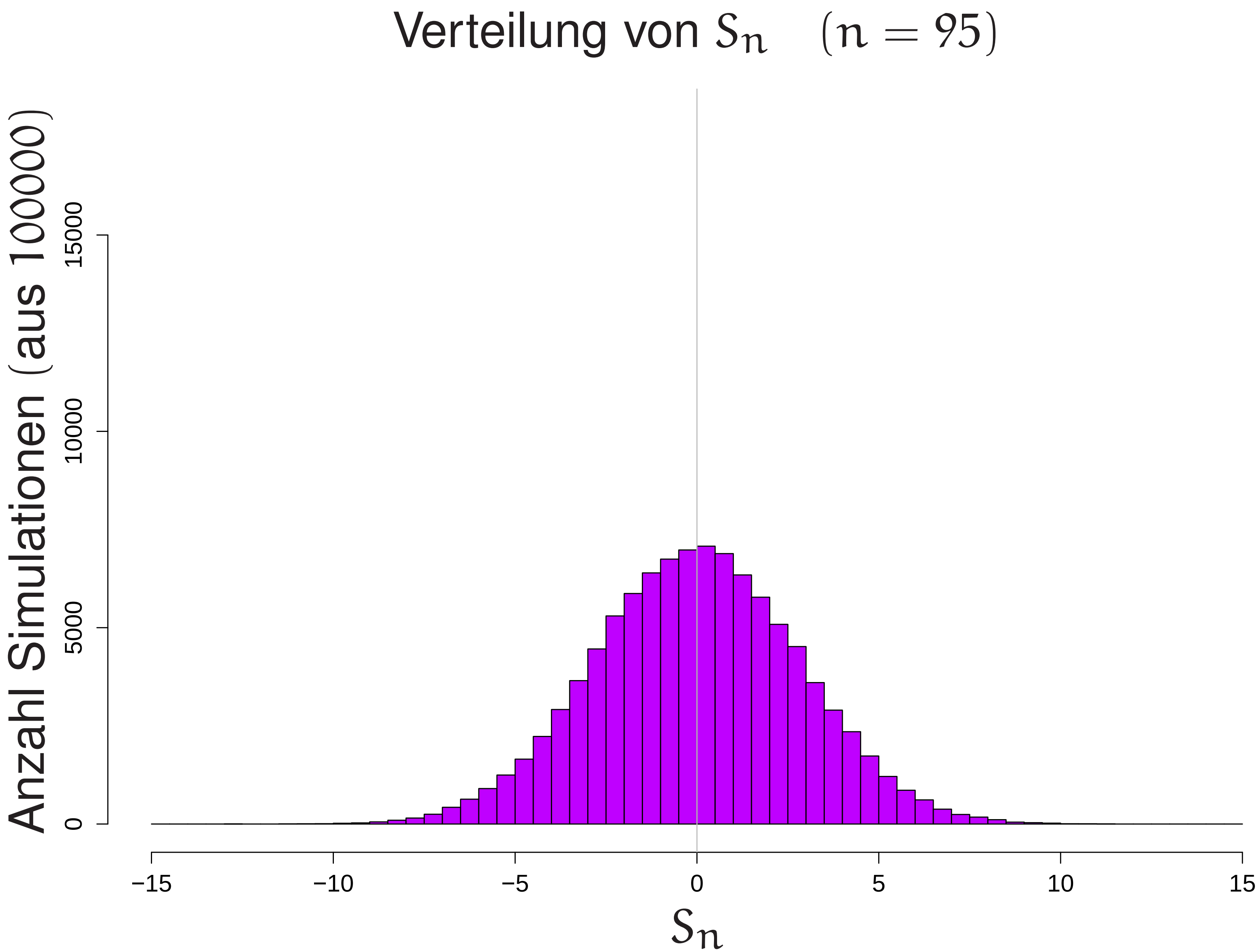


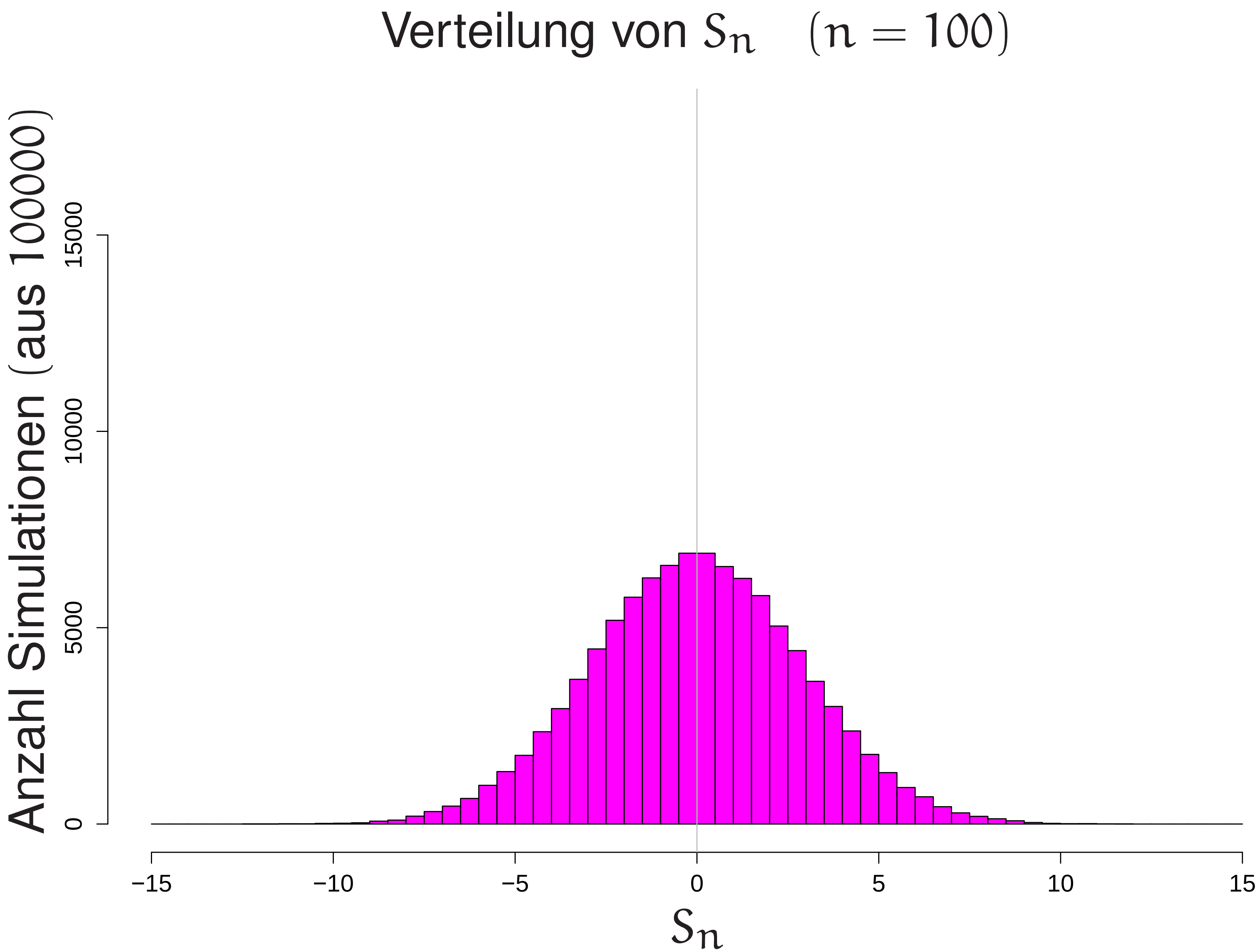
Verteilung von S_n $(n = 80)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n


Verteilung von S_n $(n = 85)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n

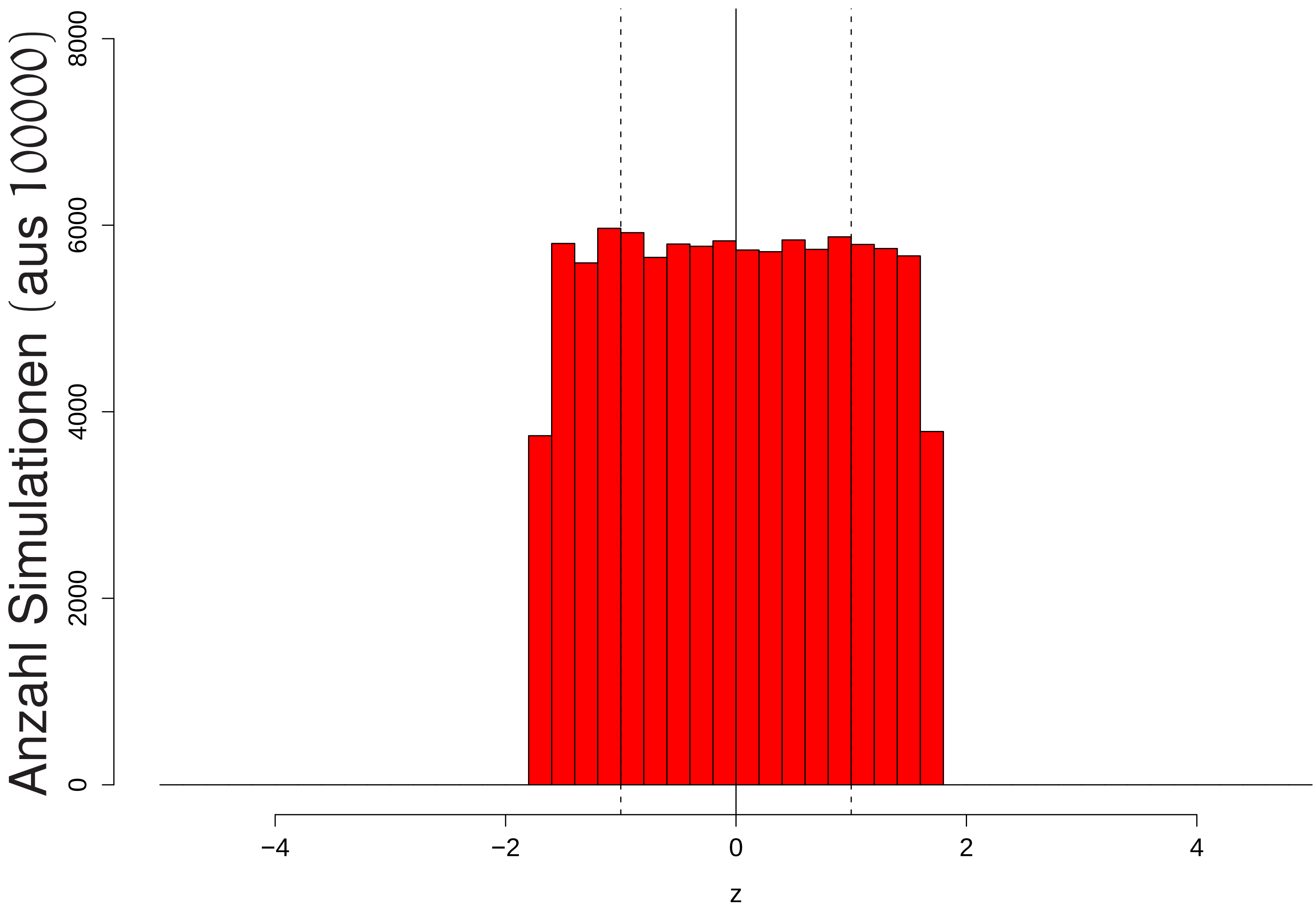


Verteilung von S_n $(n = 90)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n

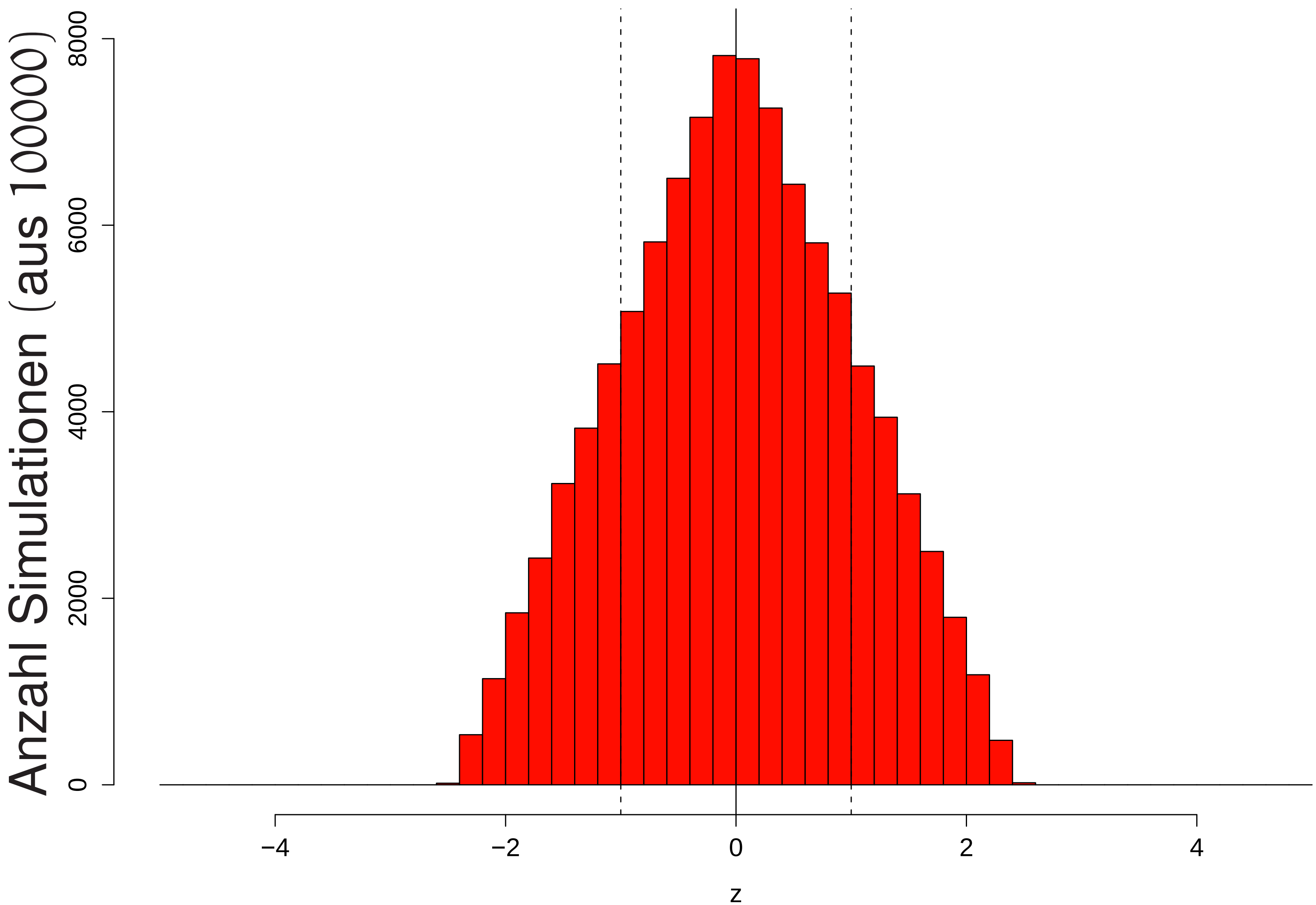

Verteilung von S_n $(n = 95)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n

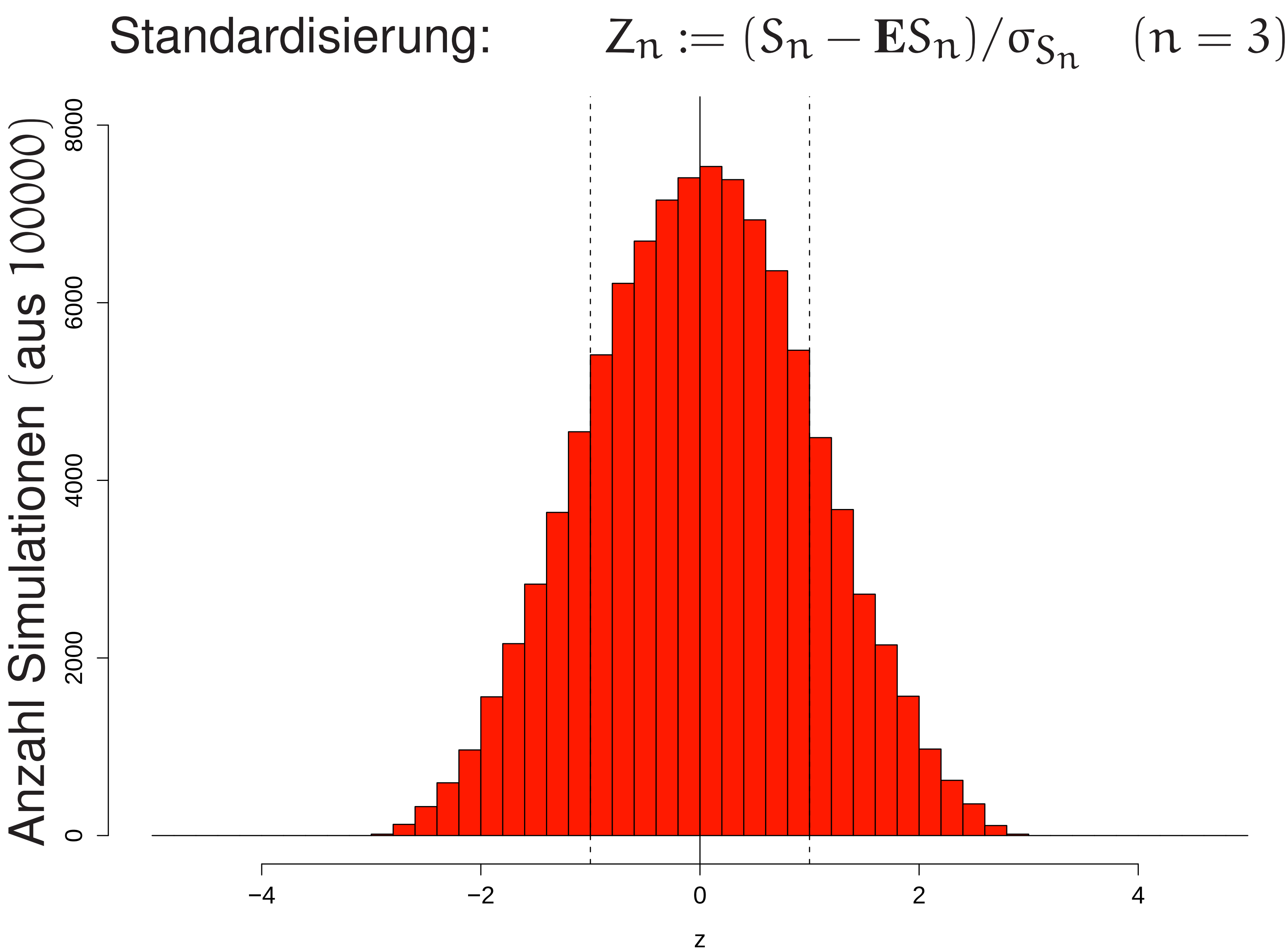

Verteilung von S_n $(n = 100)$
Anzahl Simulationen (aus 100000)
0
5000
10000
15000
−15
−10
−5
0
5
10
15
S_n

Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 1)$

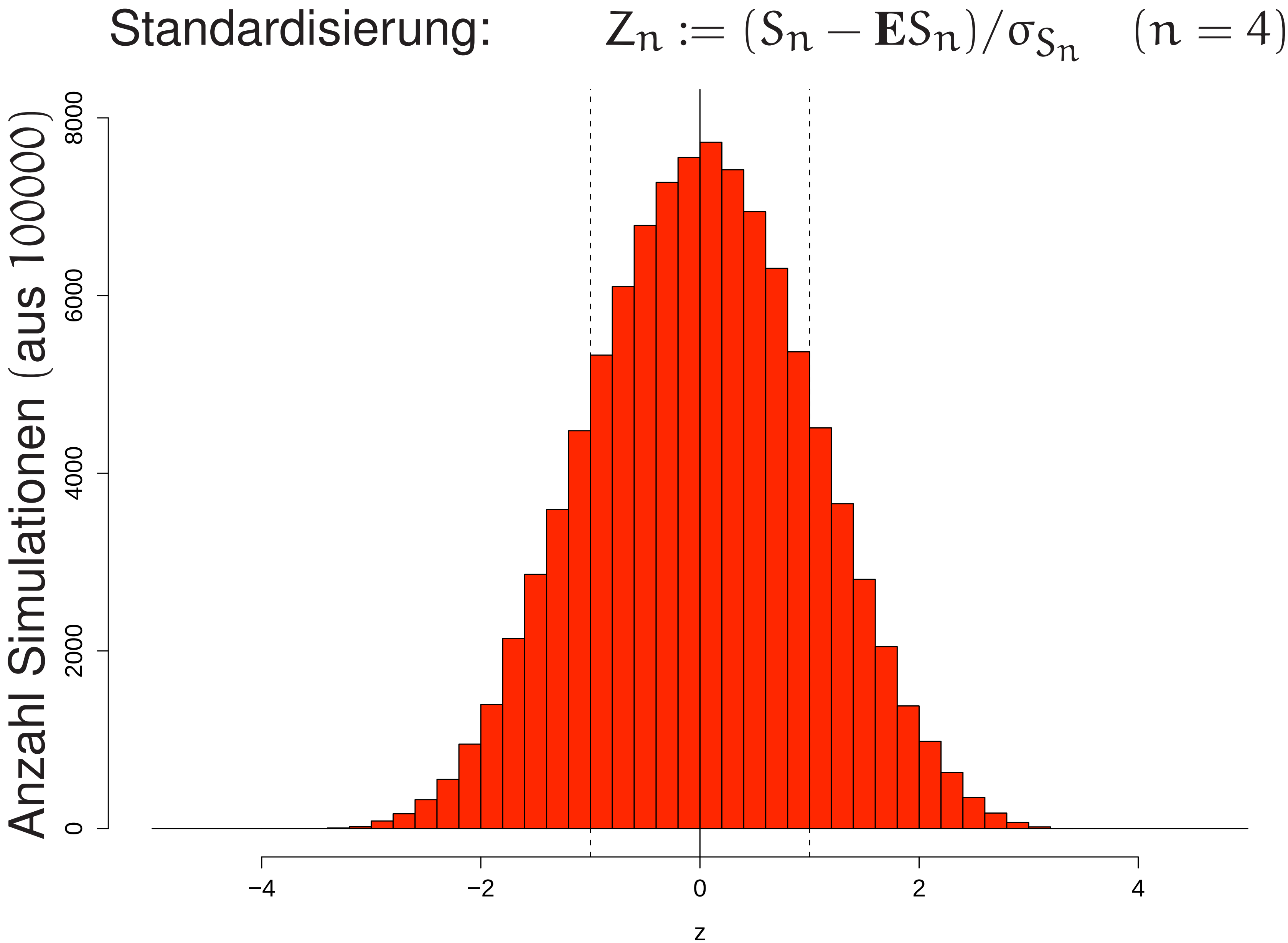


Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 2)$



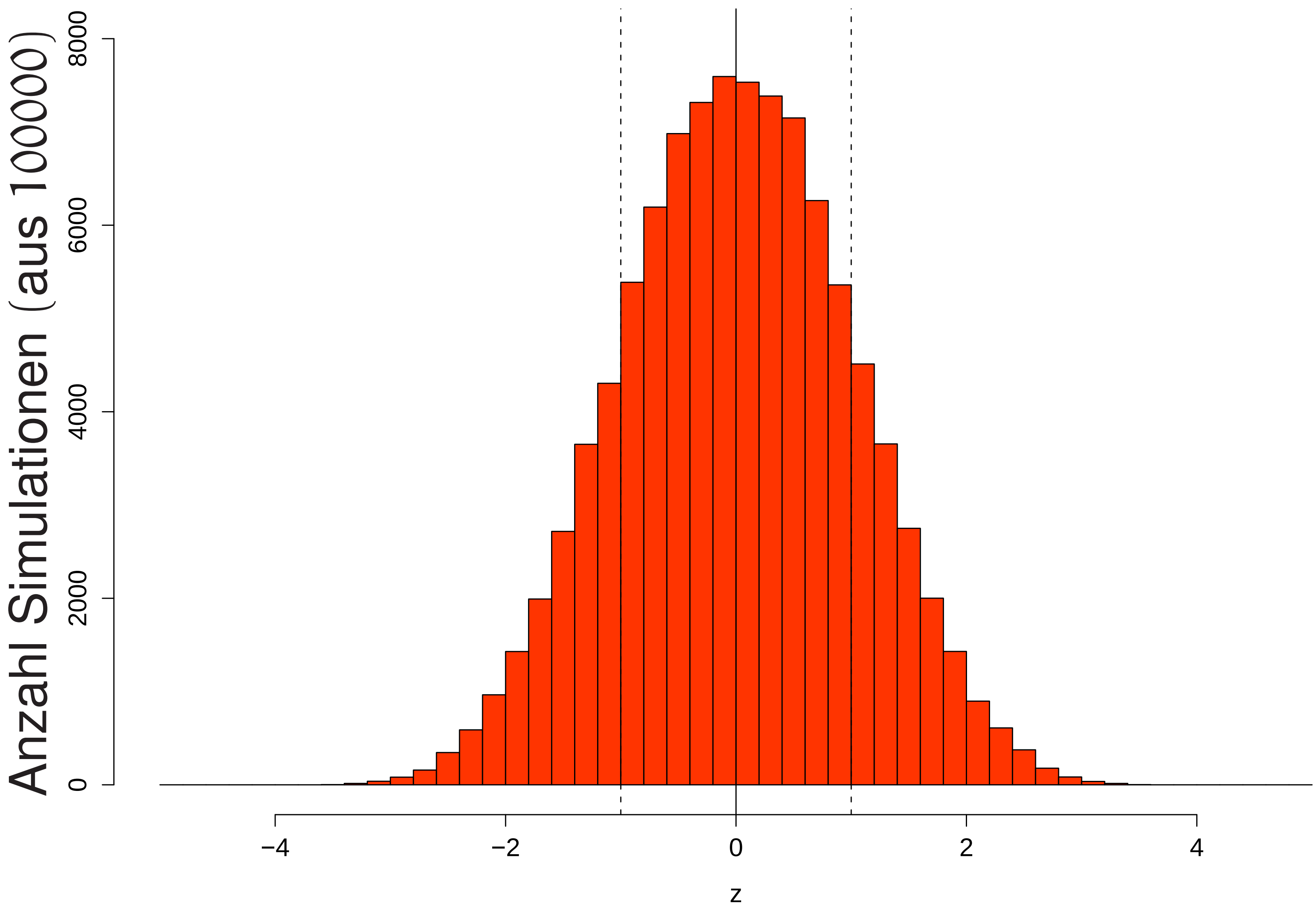


Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 3)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

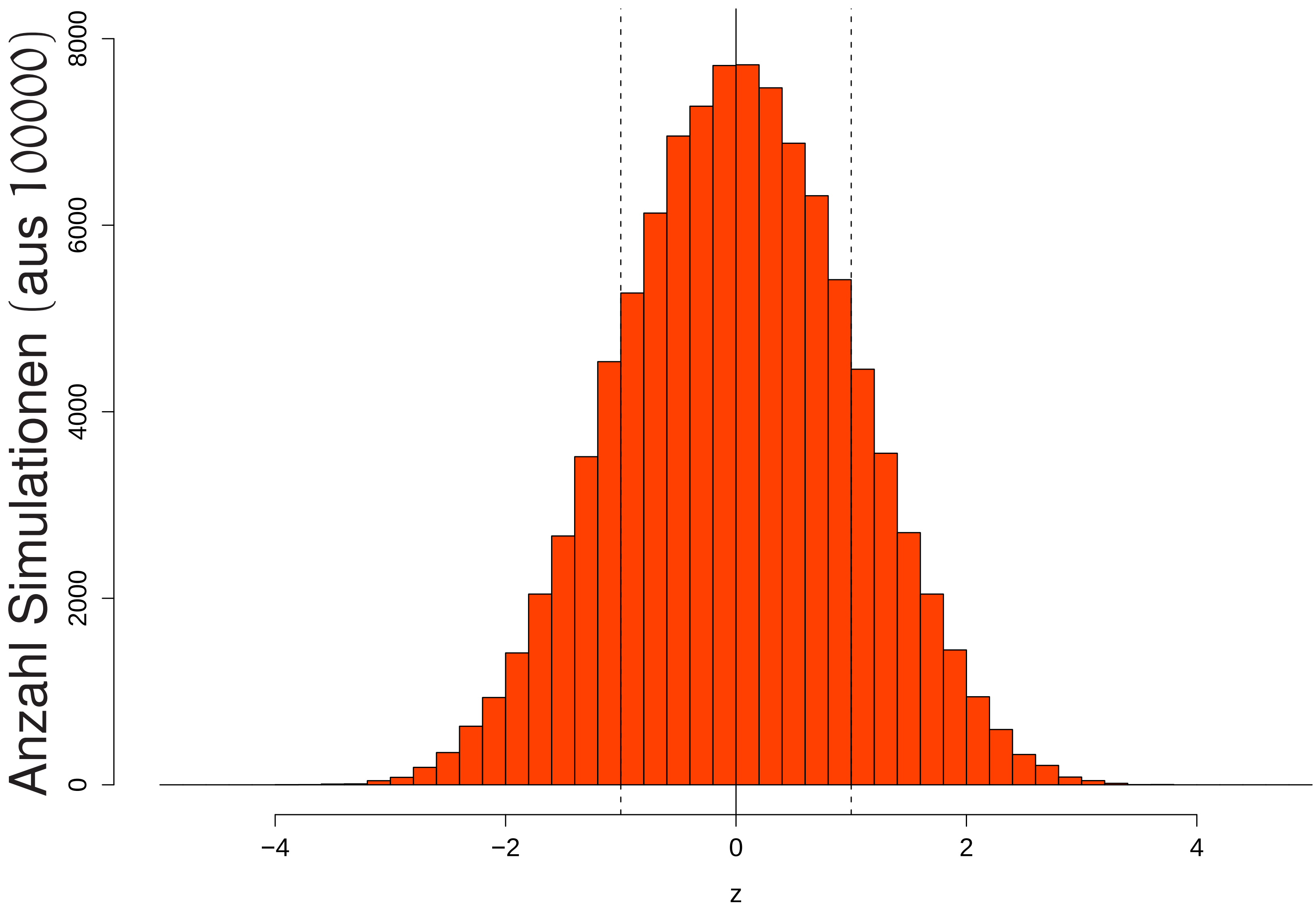


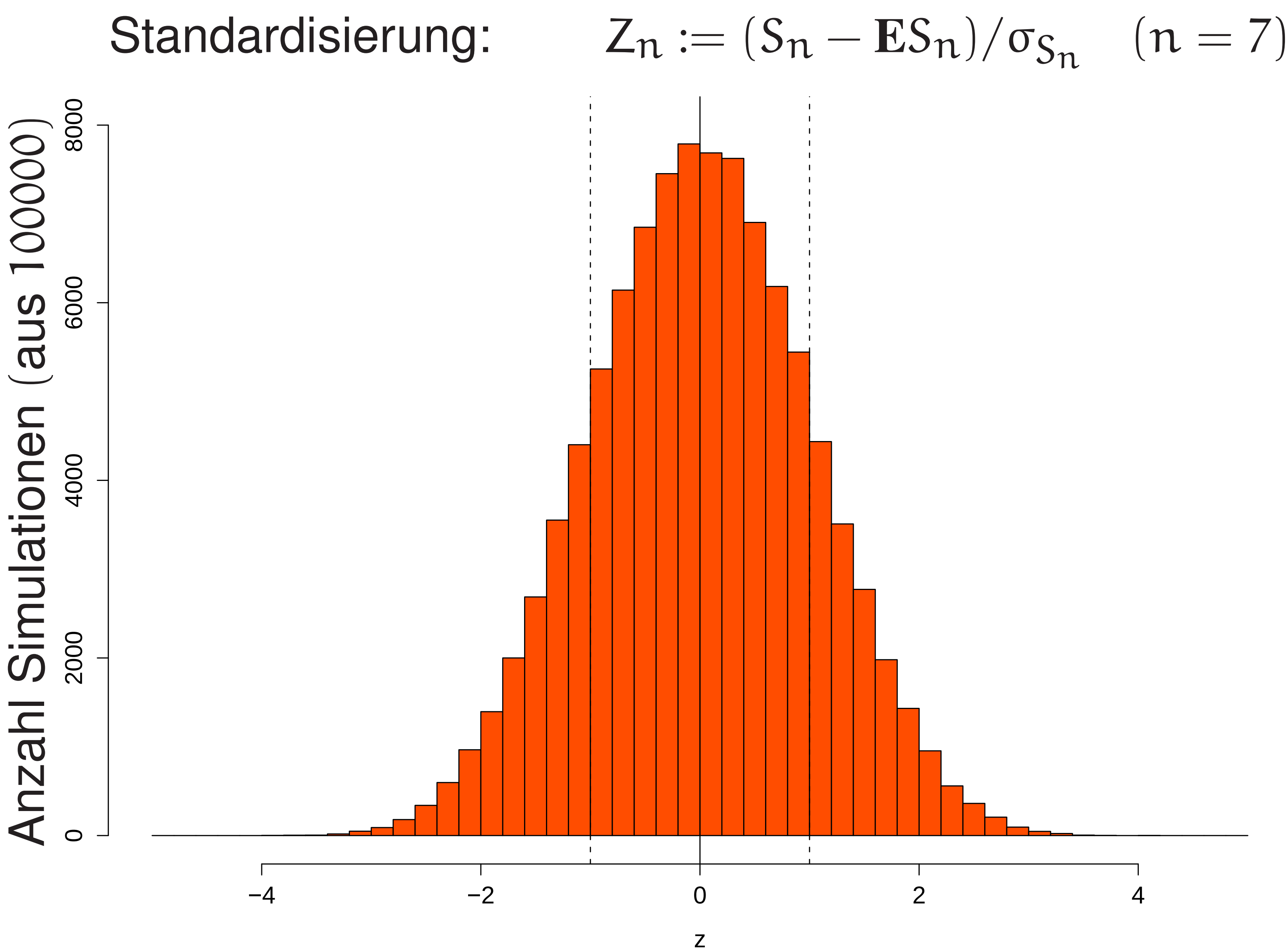
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 4)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

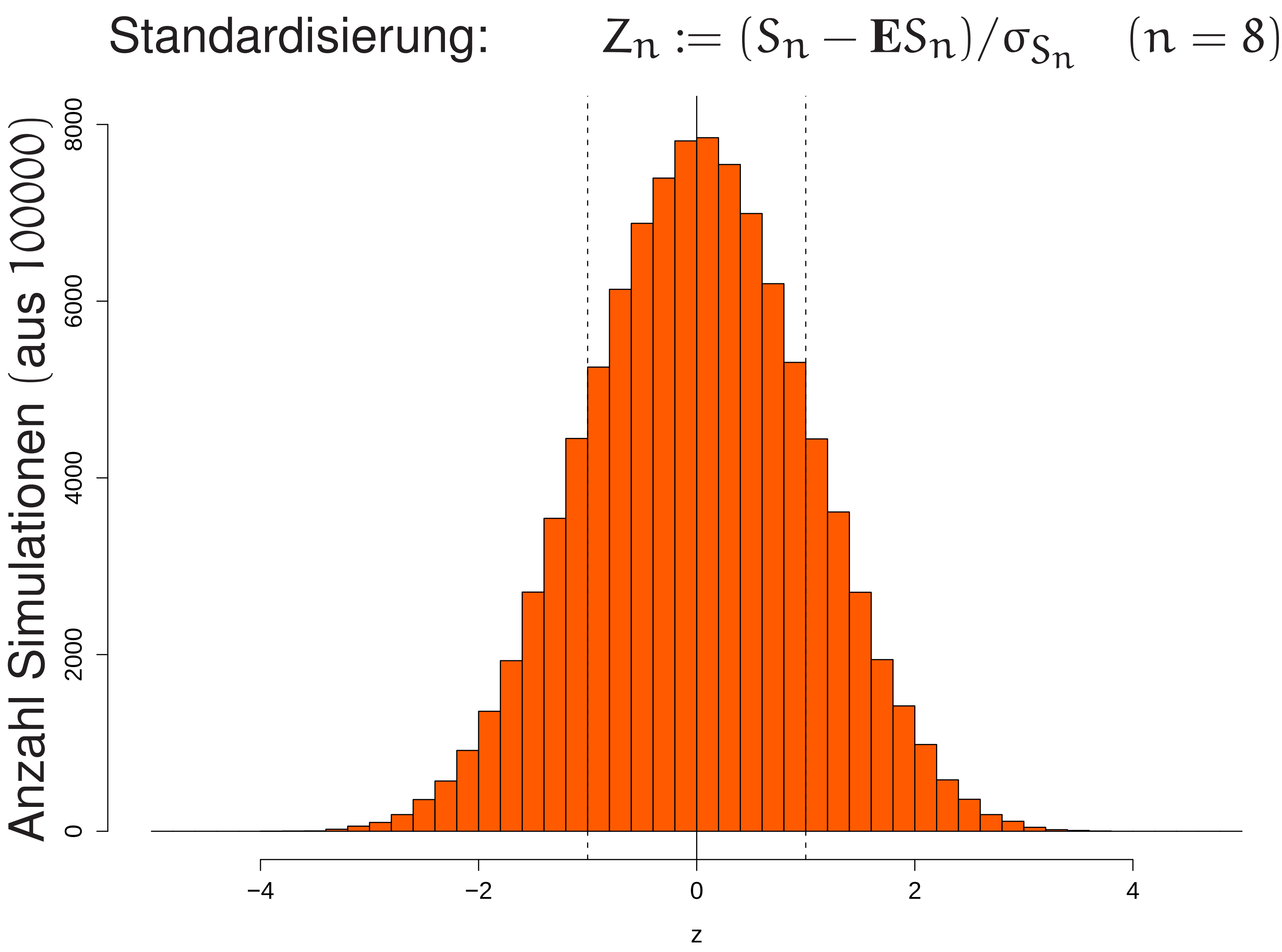
Standardisierung: $\quad Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 5)$

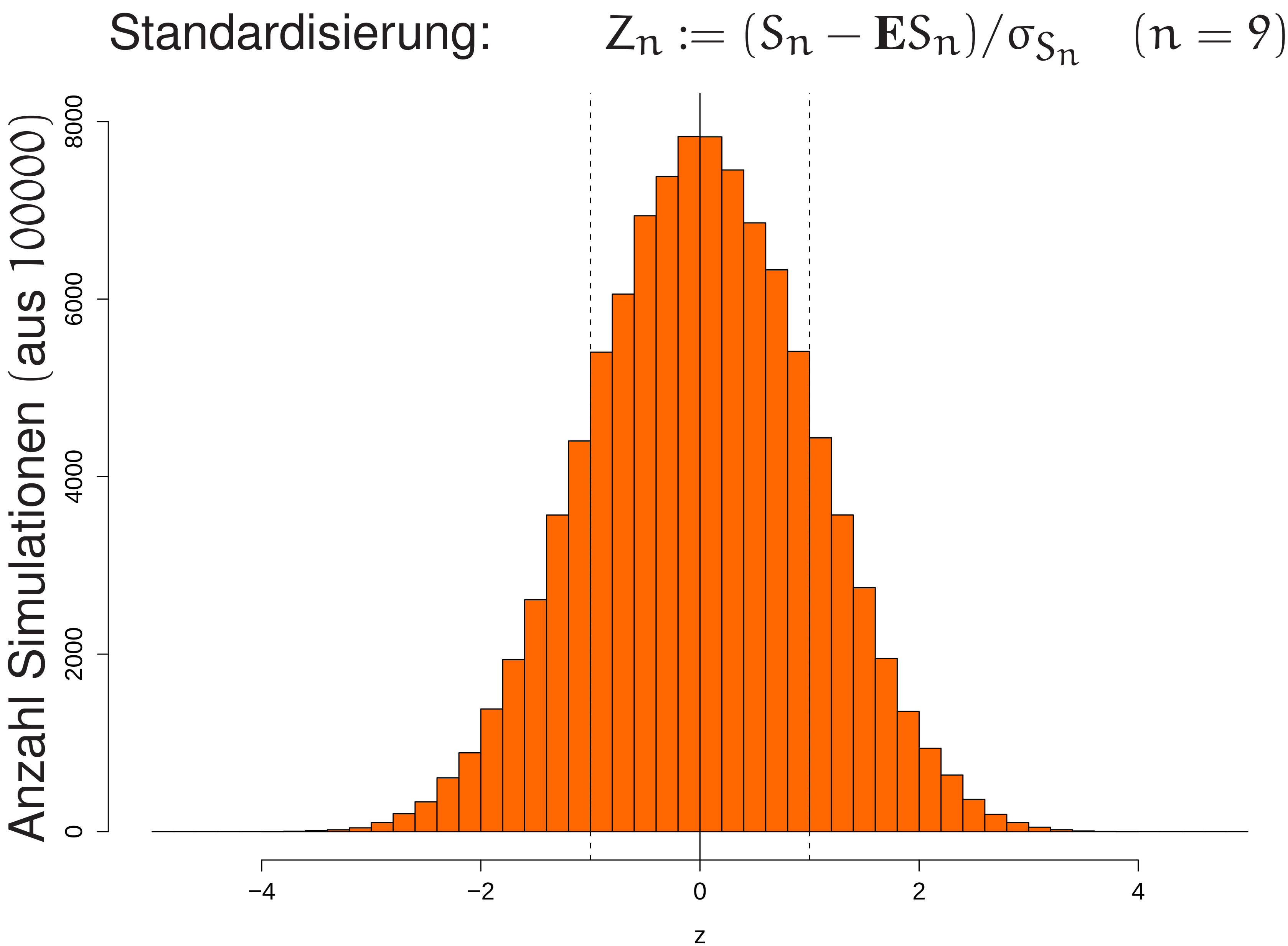


Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 6)$

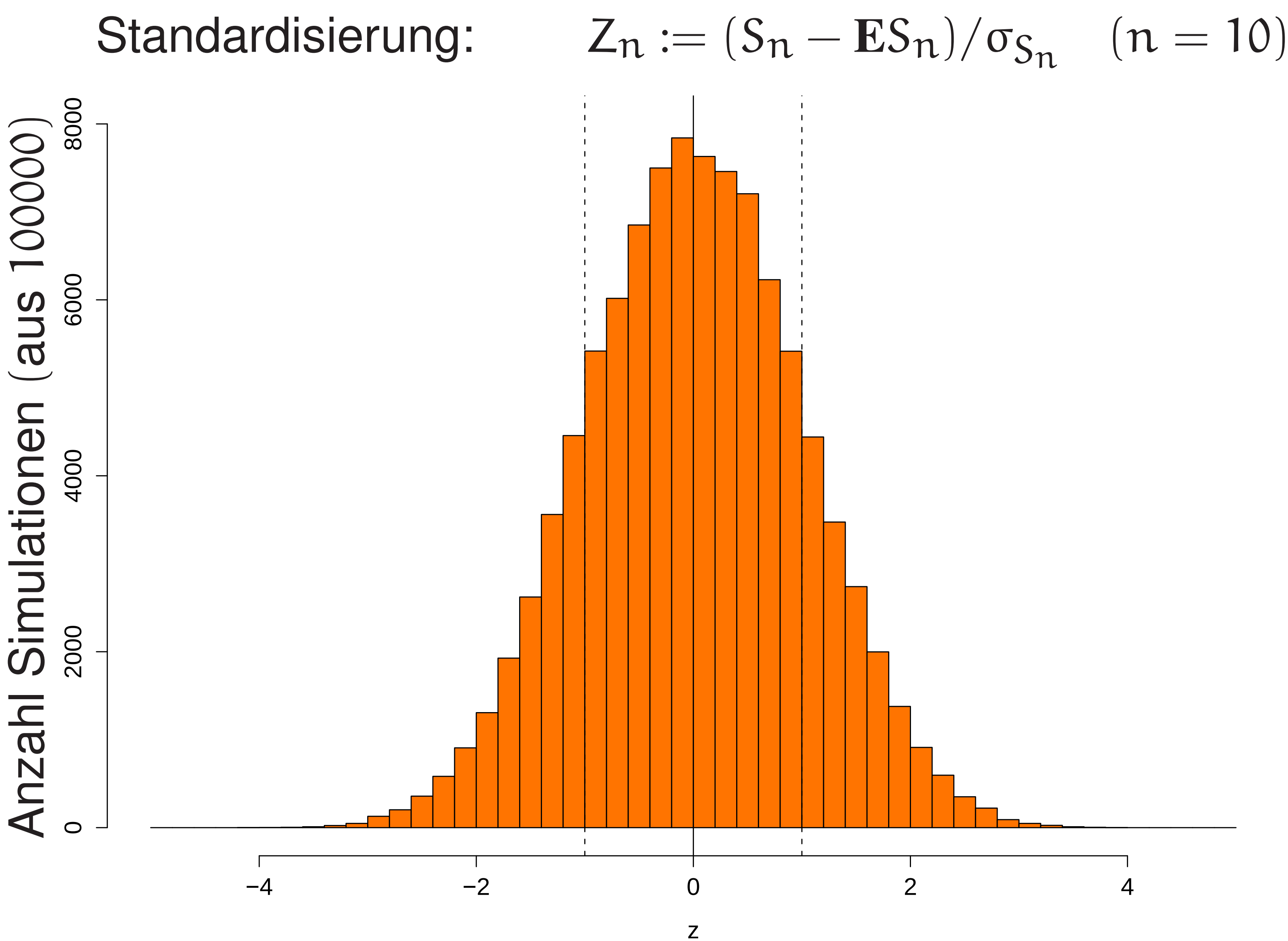



Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 7)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

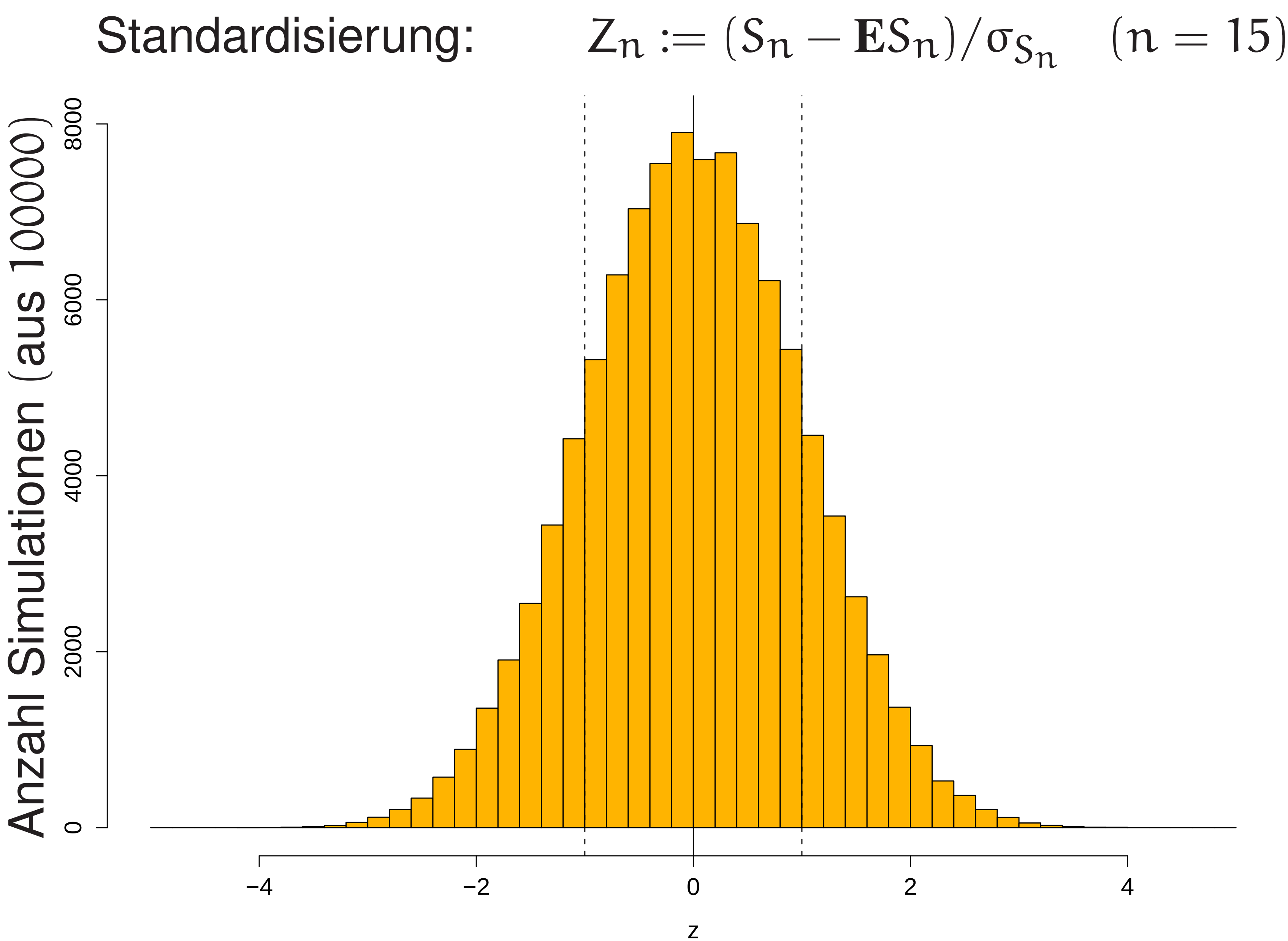

Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 8)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

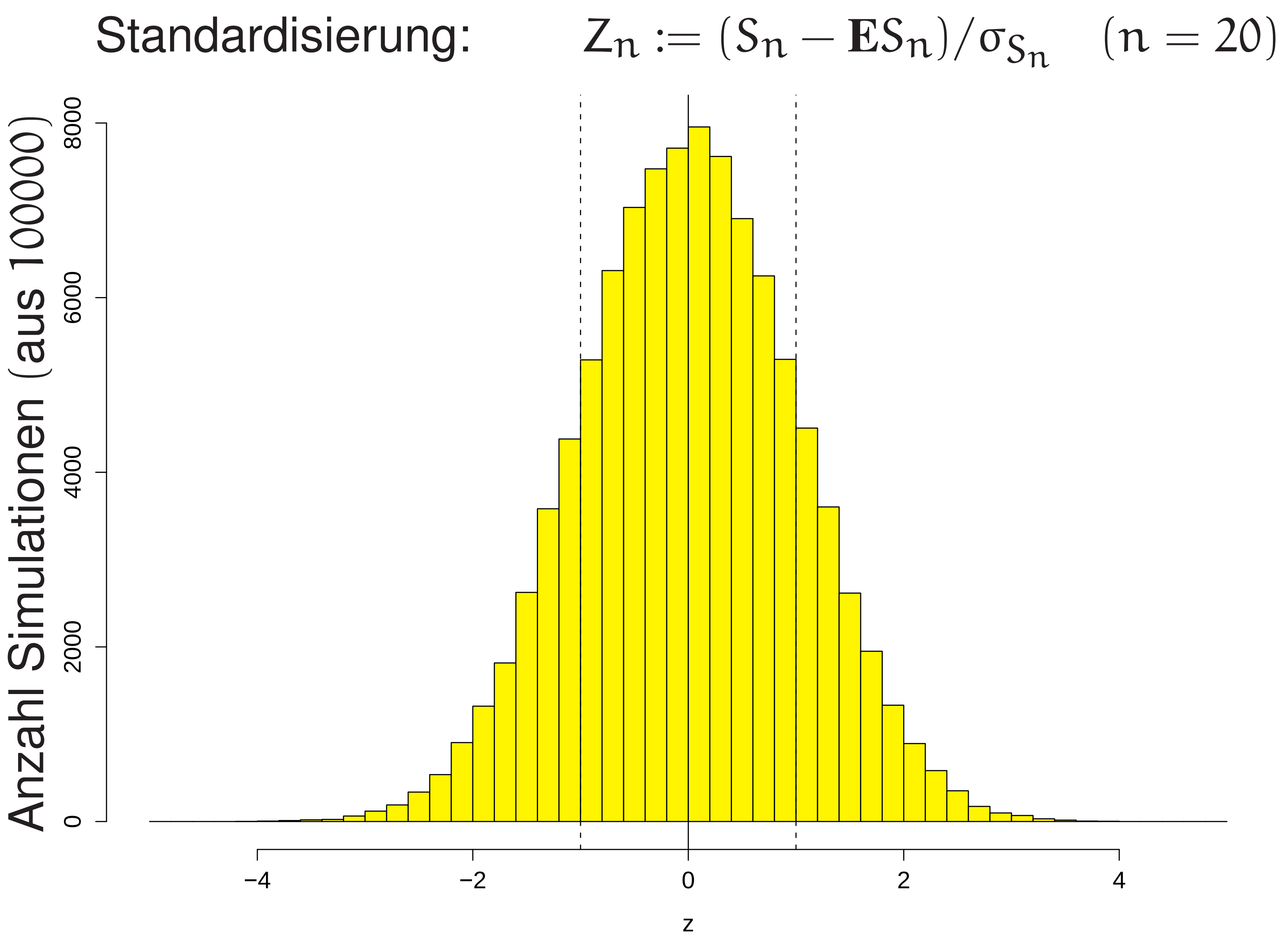


Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 9)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

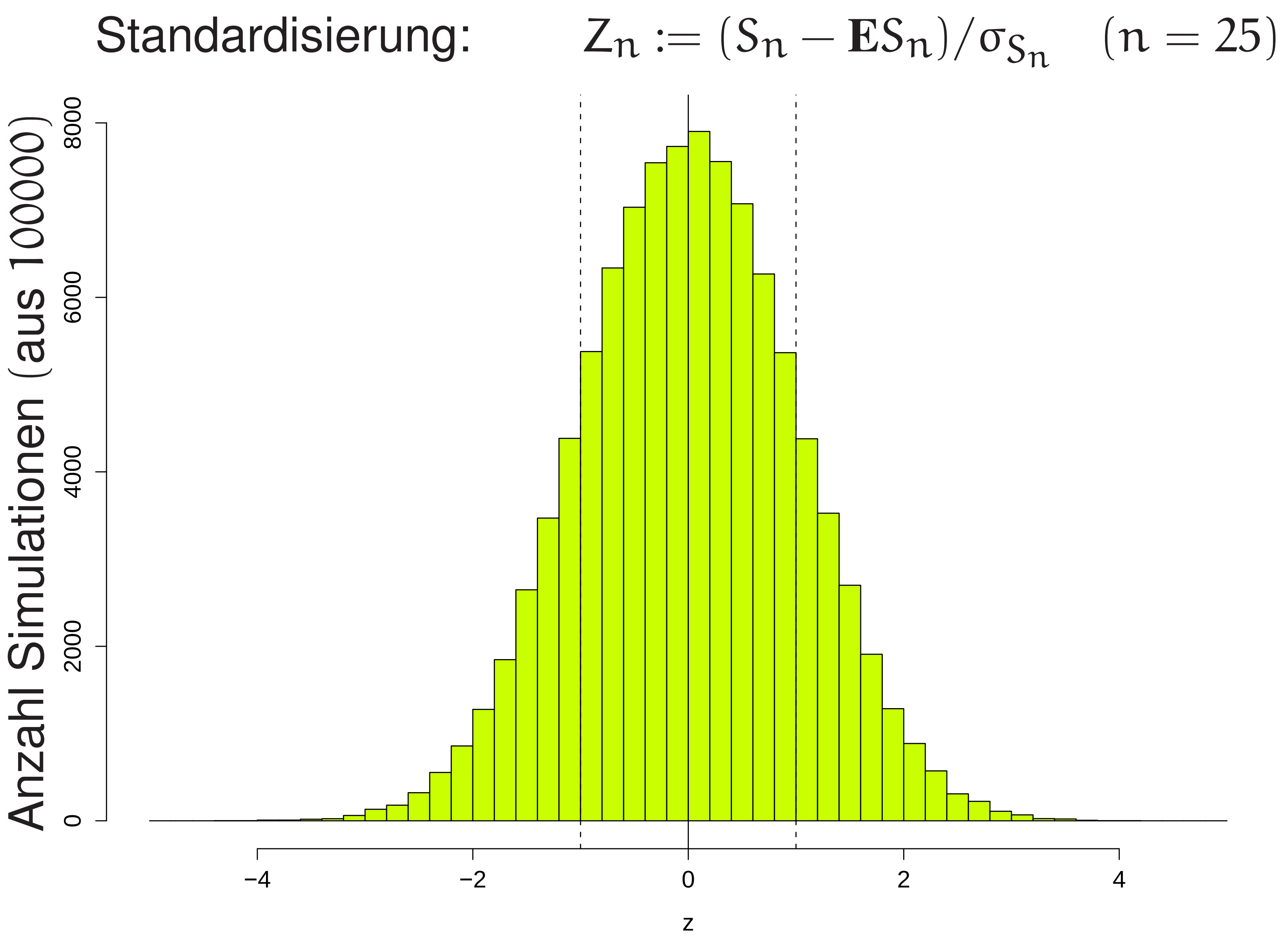


Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 10)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

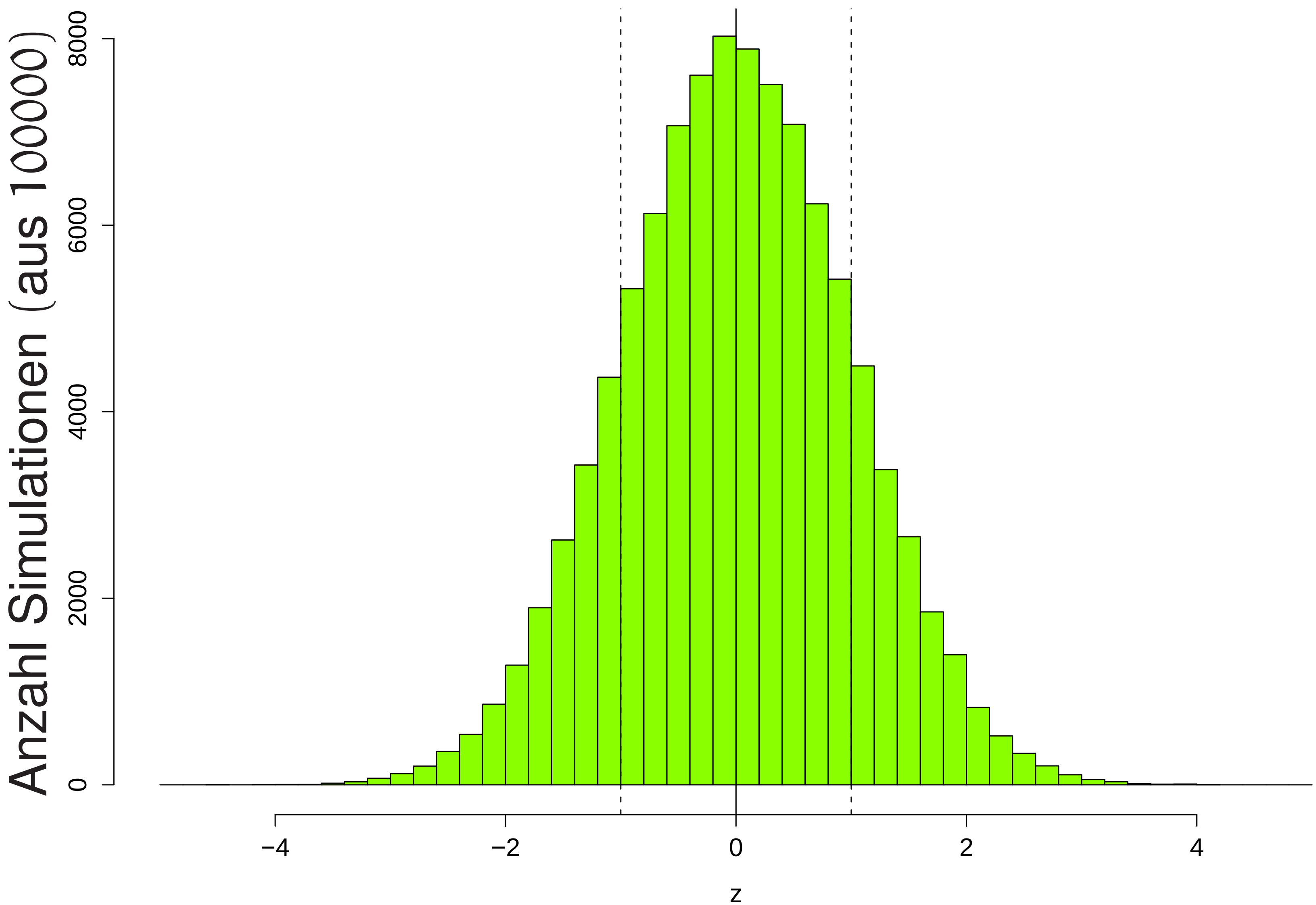

Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 15)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

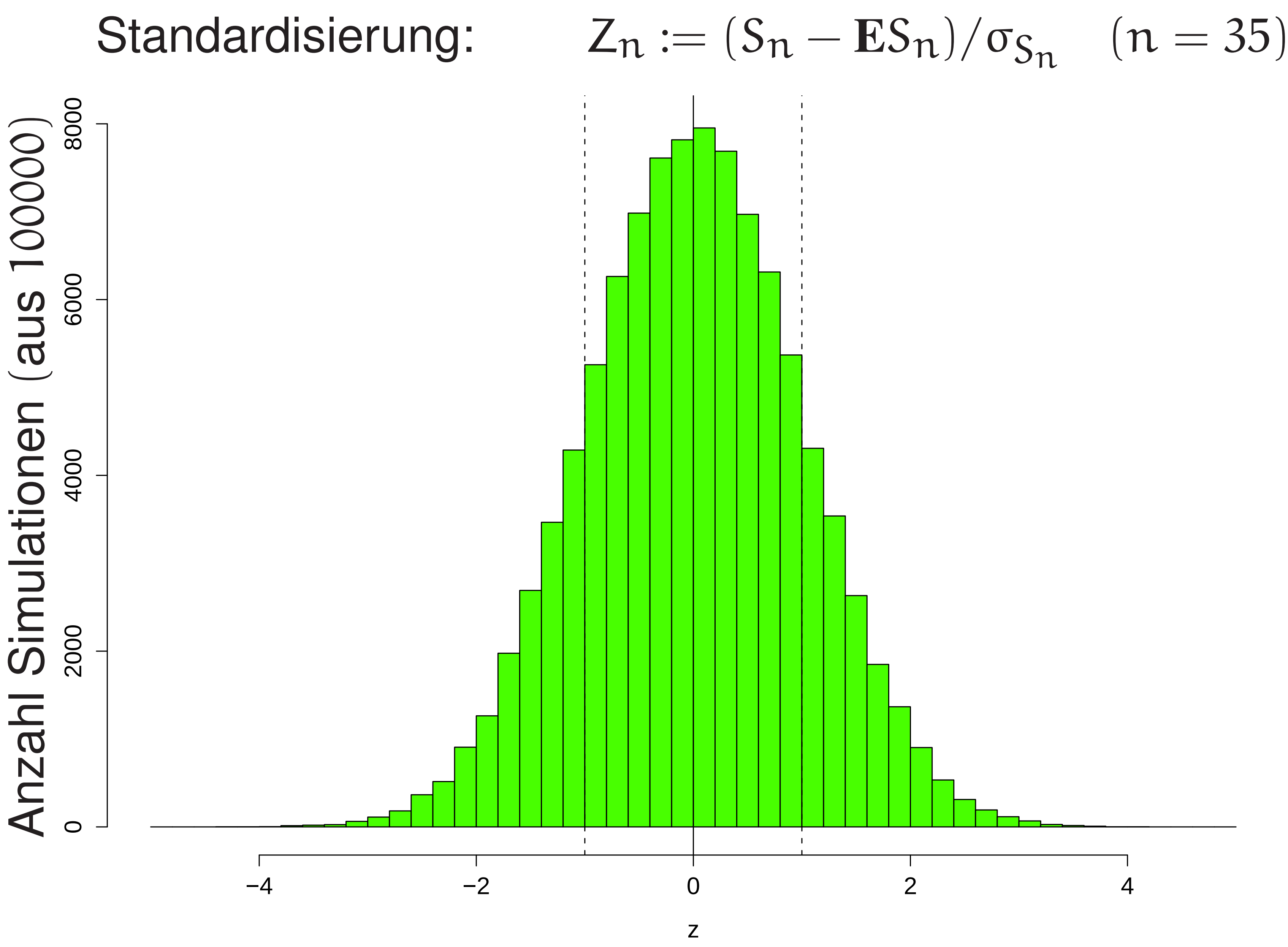


Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 20)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

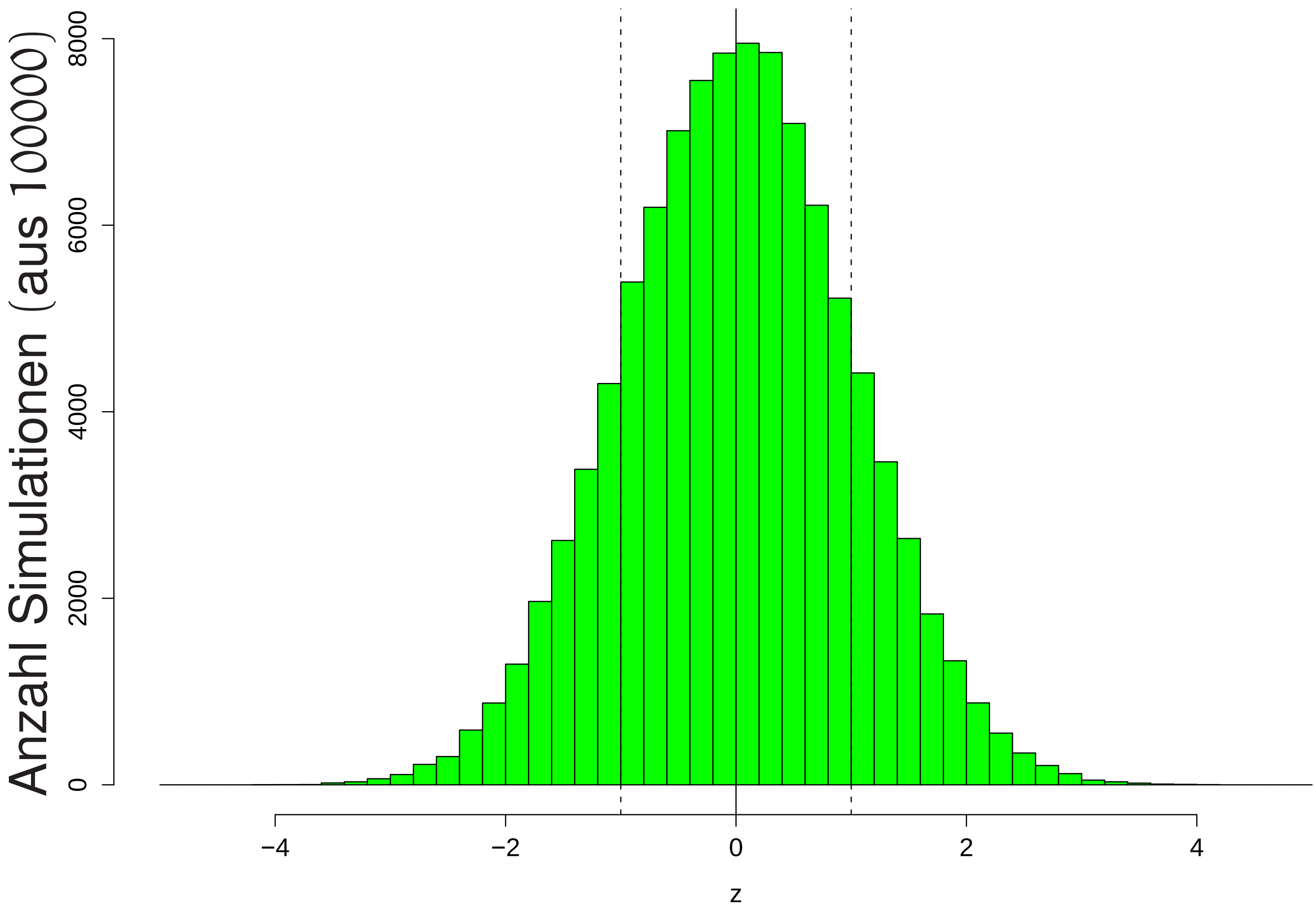

Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 25)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

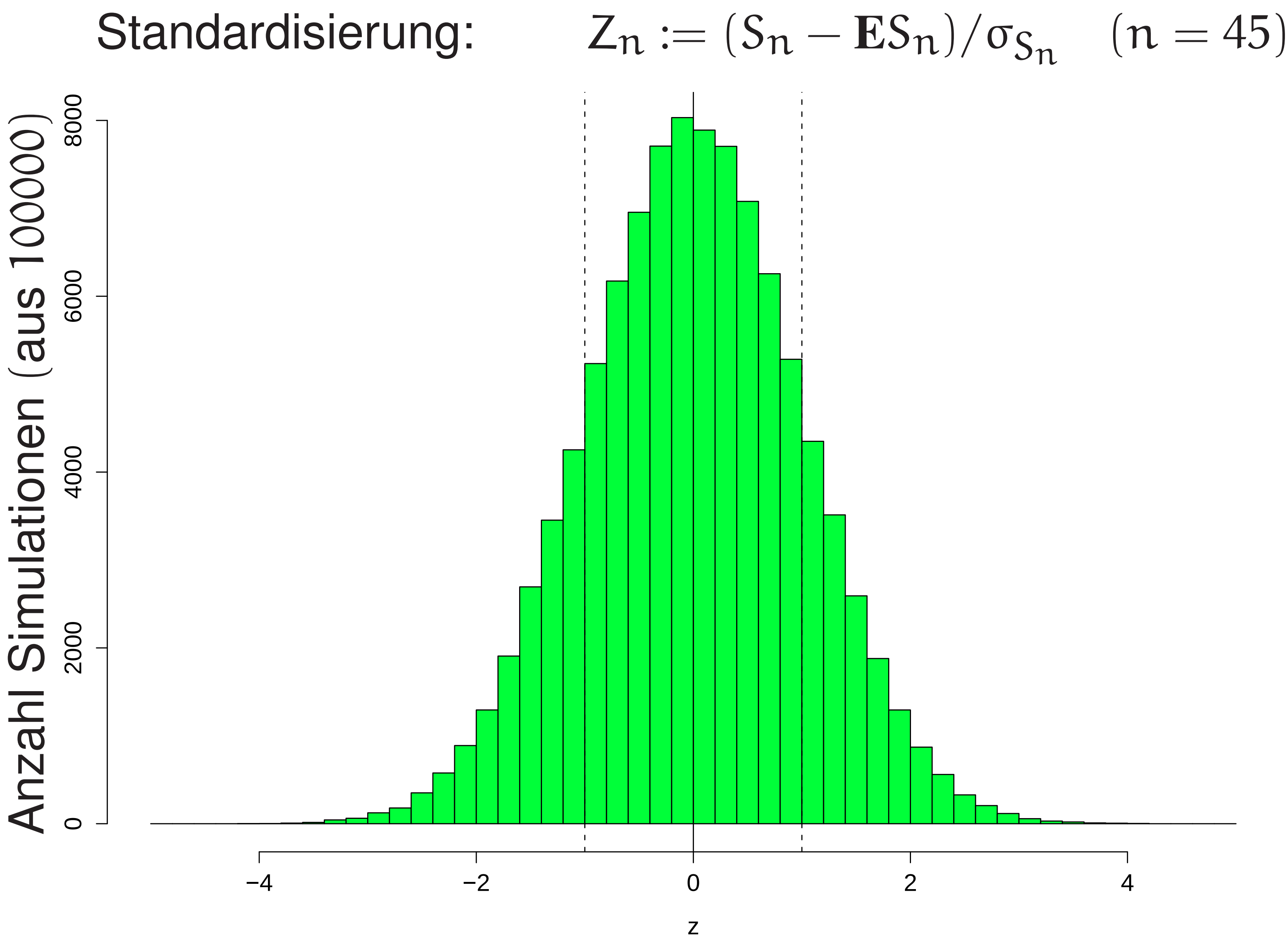
Standardisierung: $\quad Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 30)$




Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 35)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

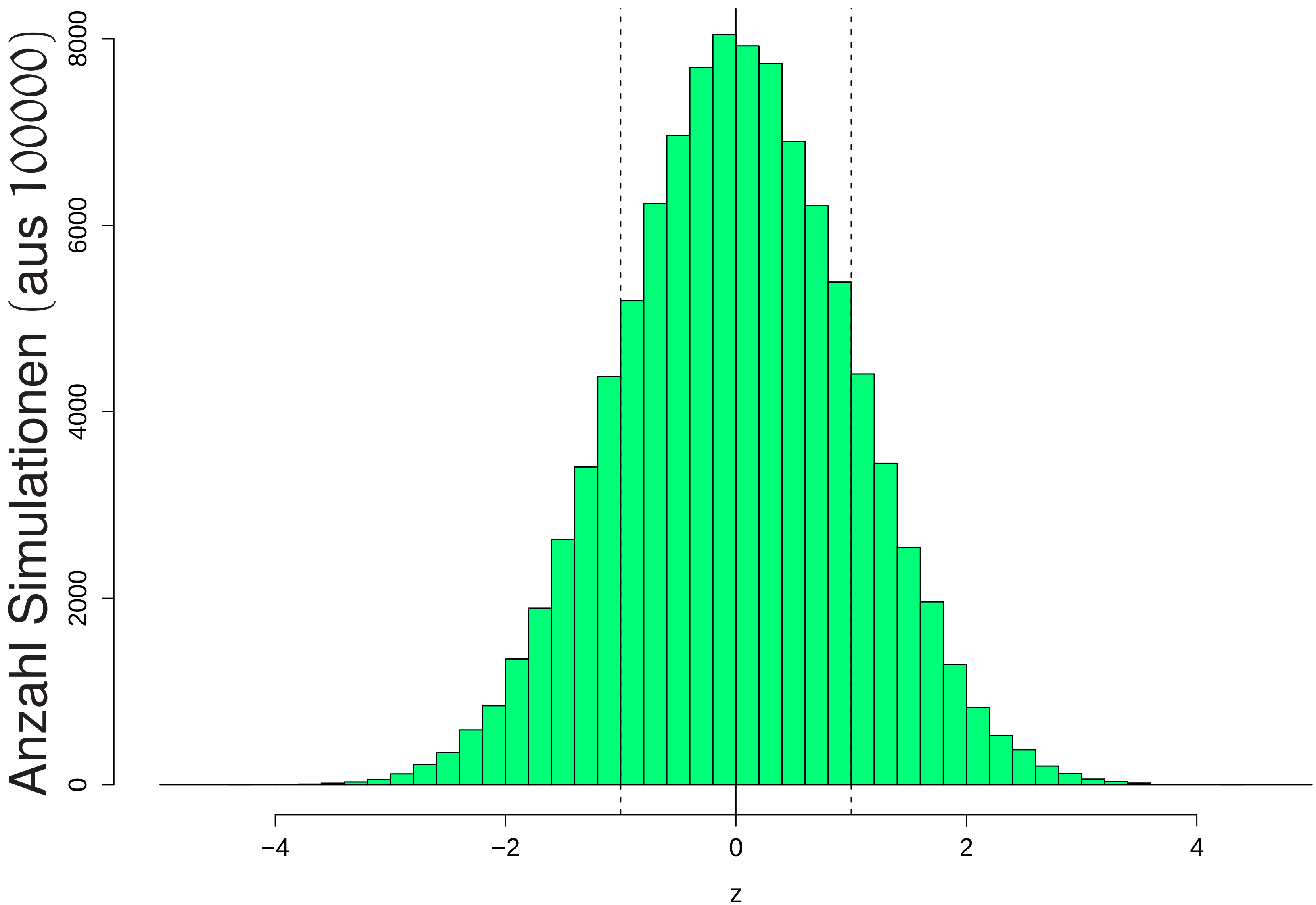
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 40)$

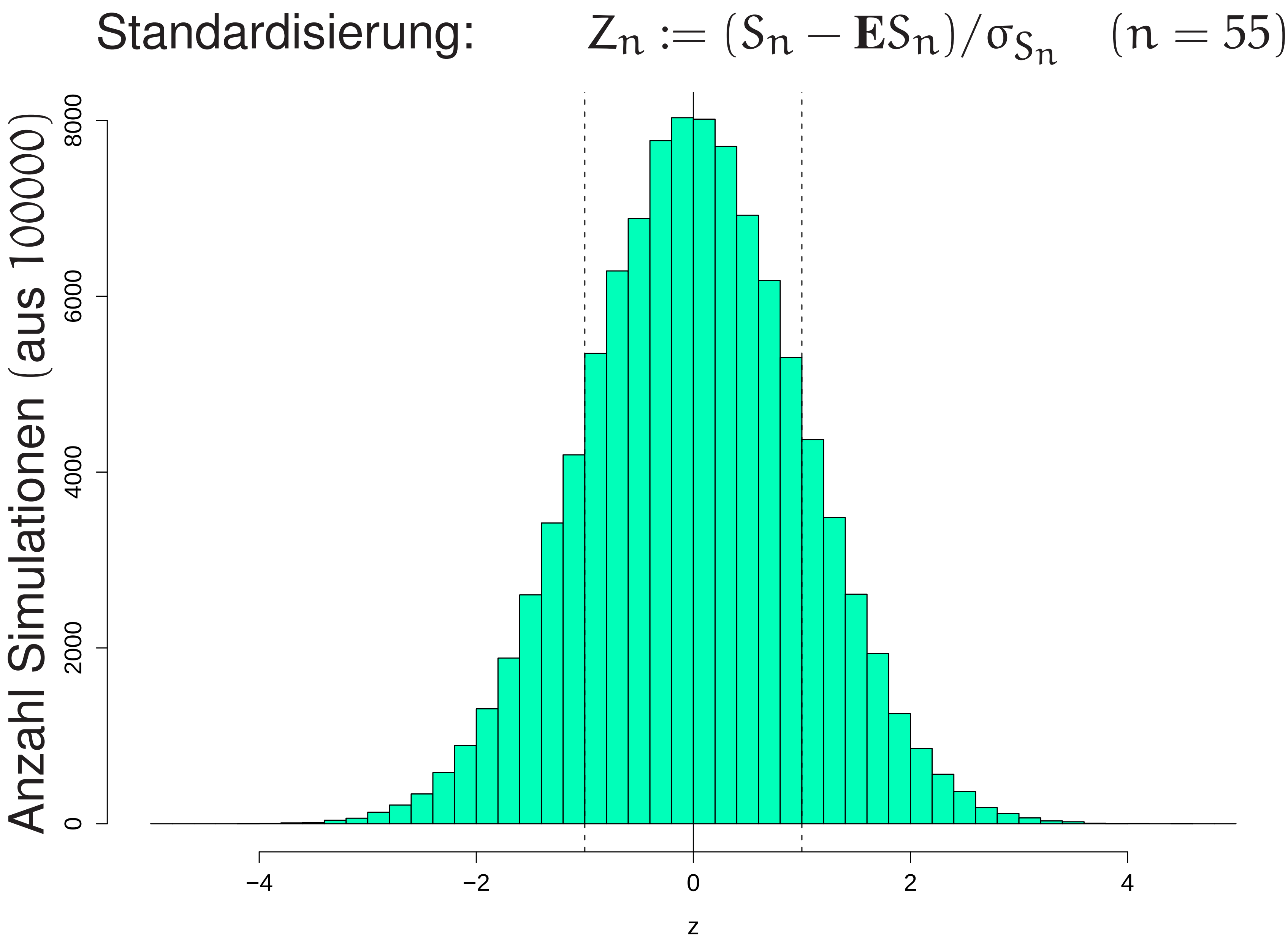


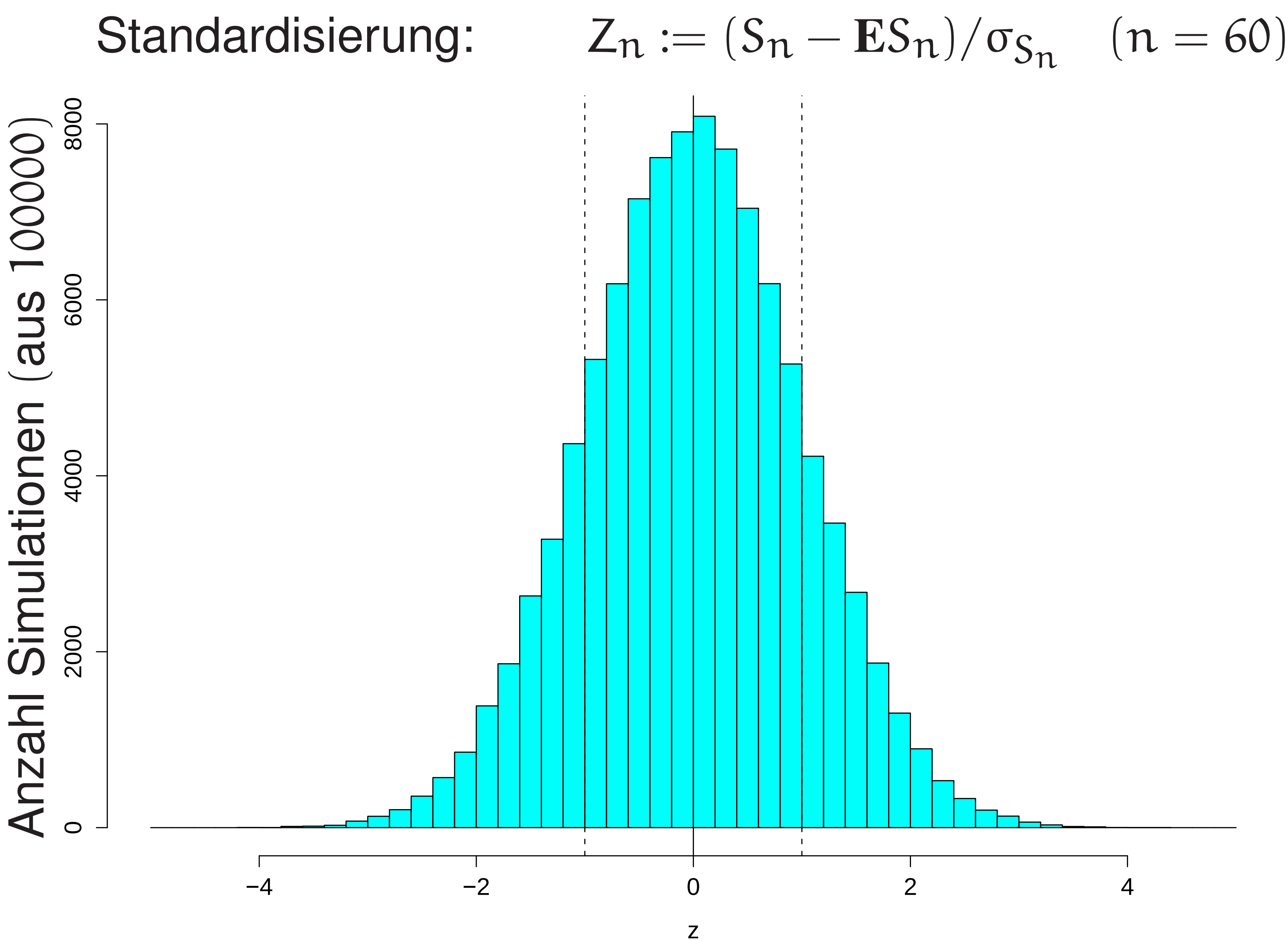


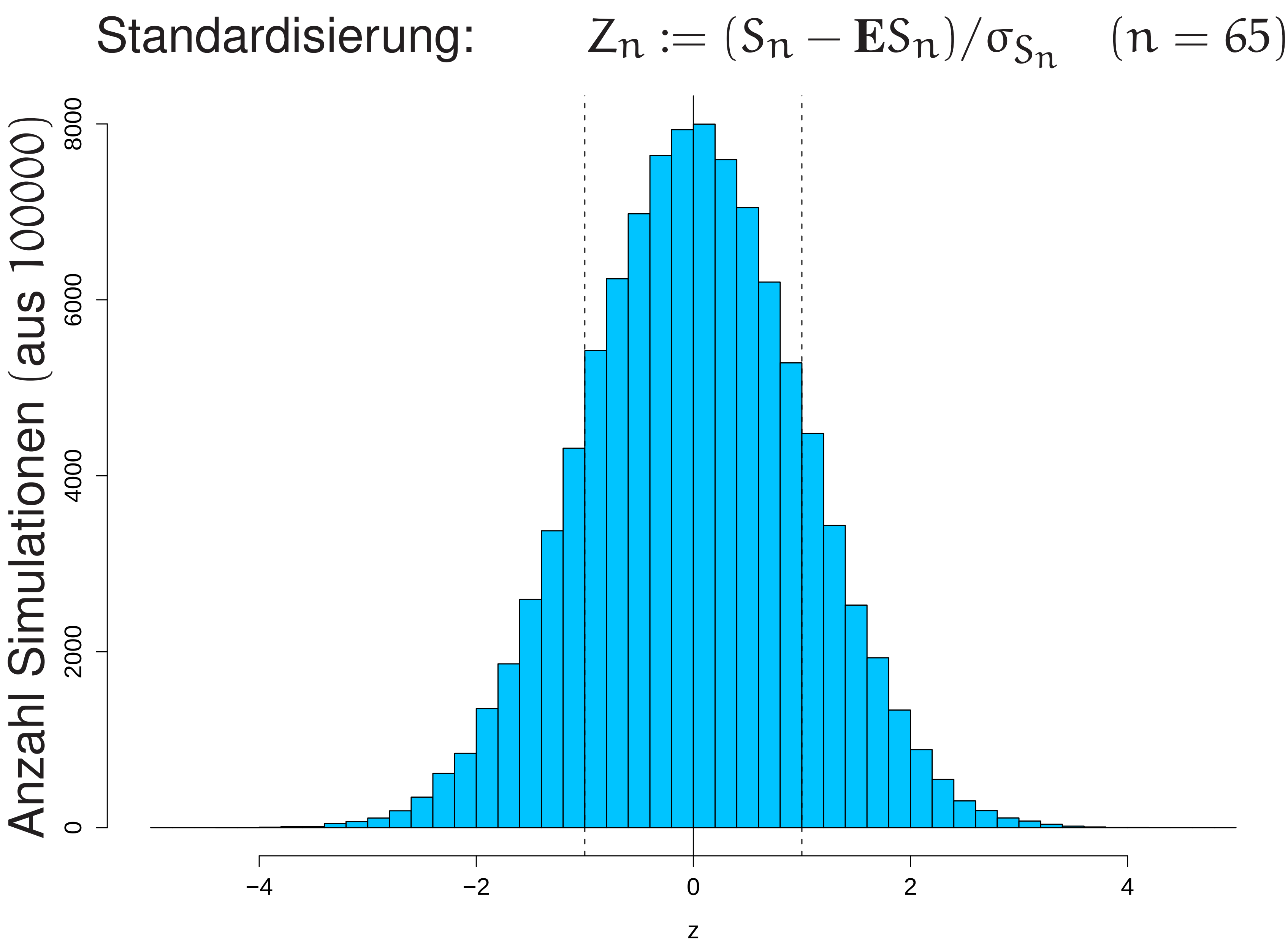
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 45)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 50)$



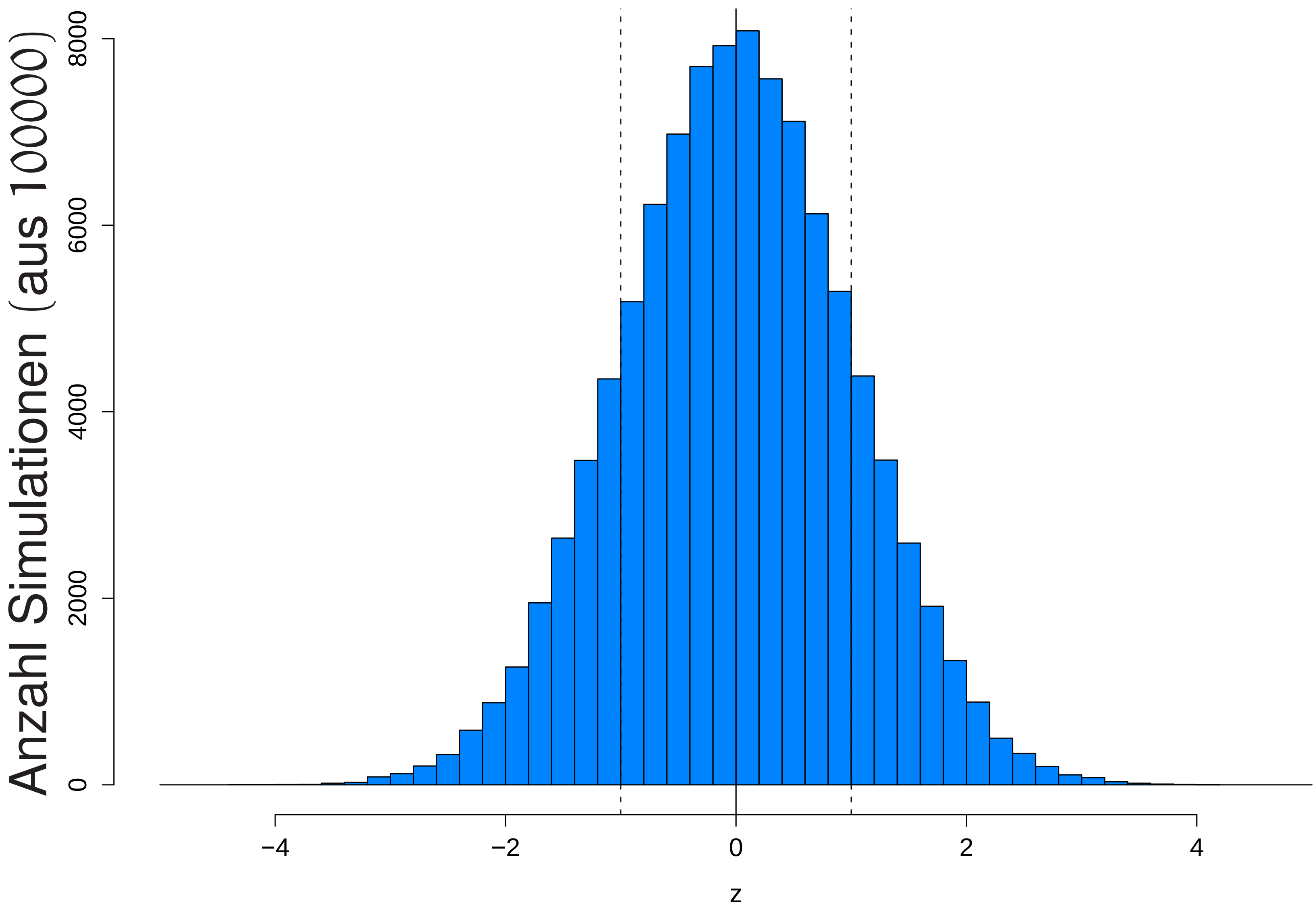

Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 55)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z


Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 60)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

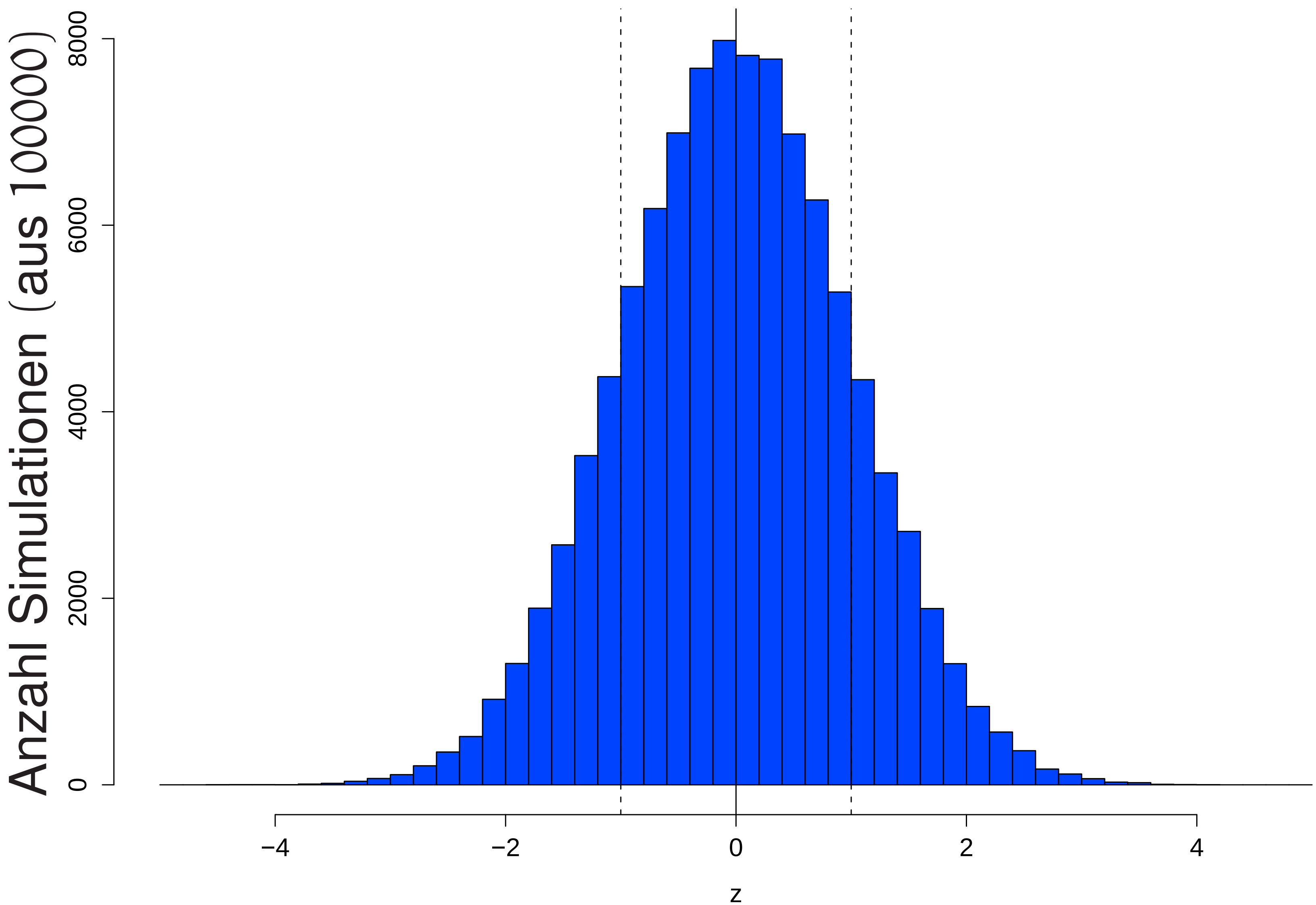


Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 65)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

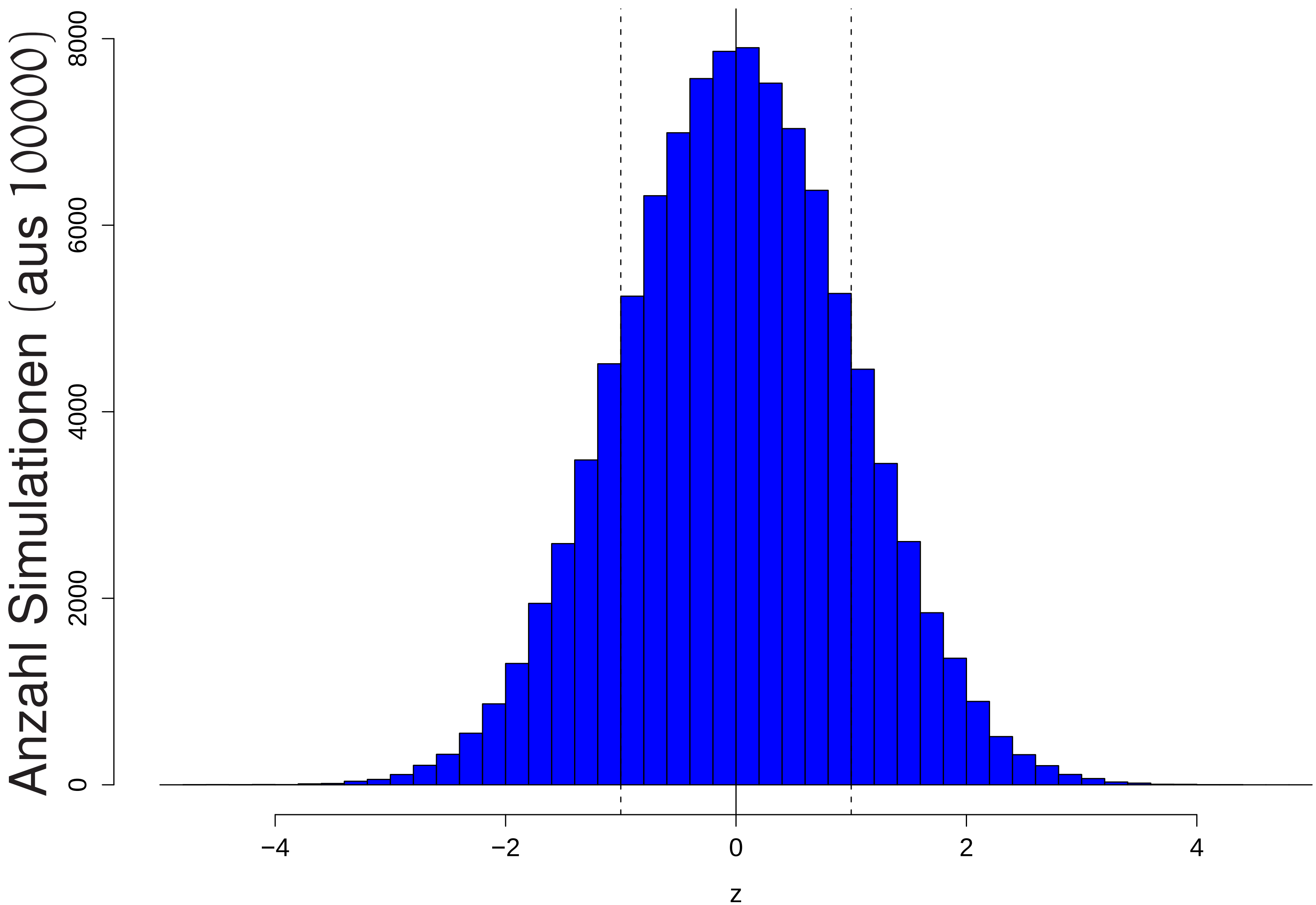
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 70)$

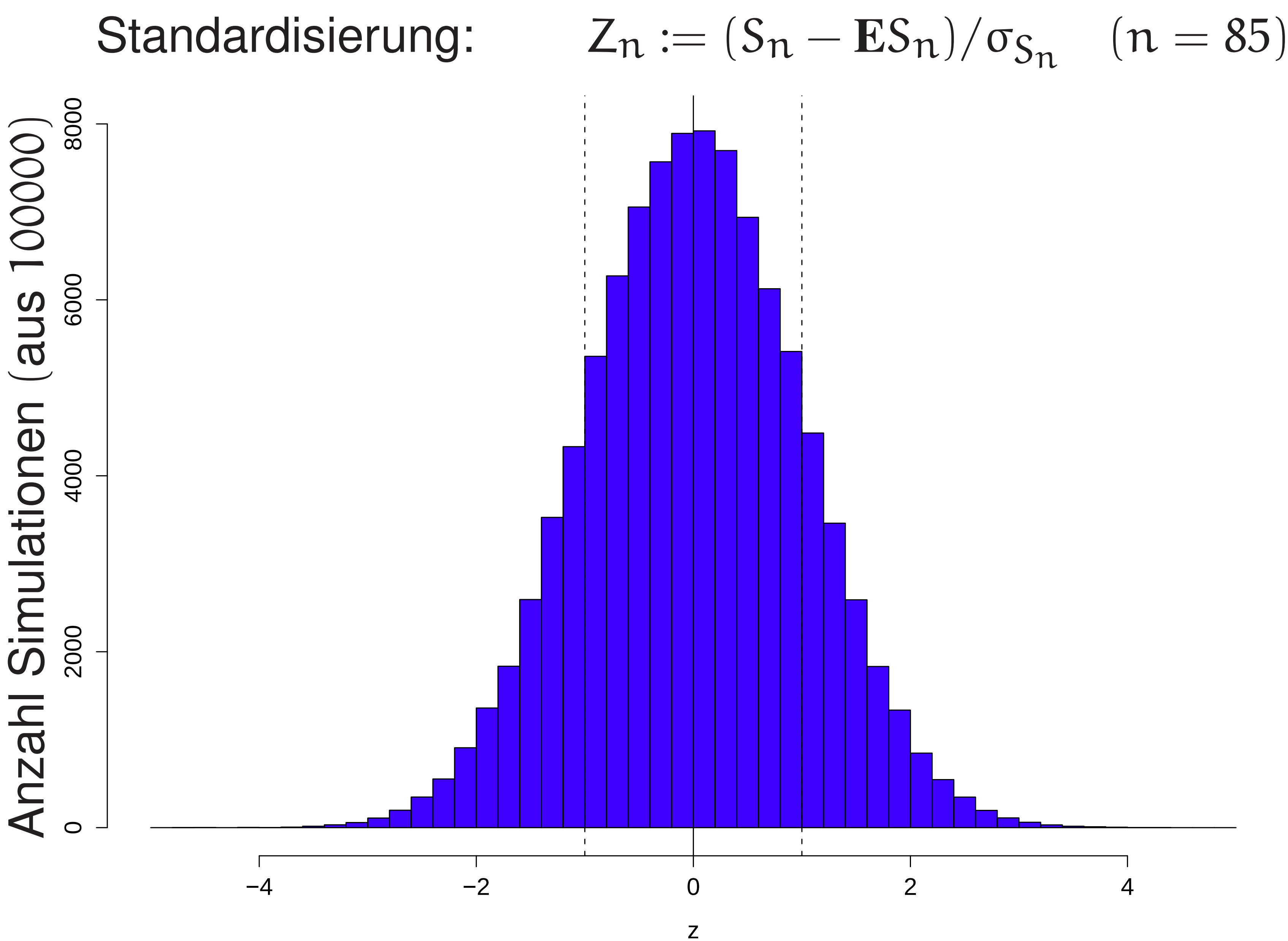


Standardisierung: $\quad Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 75)$



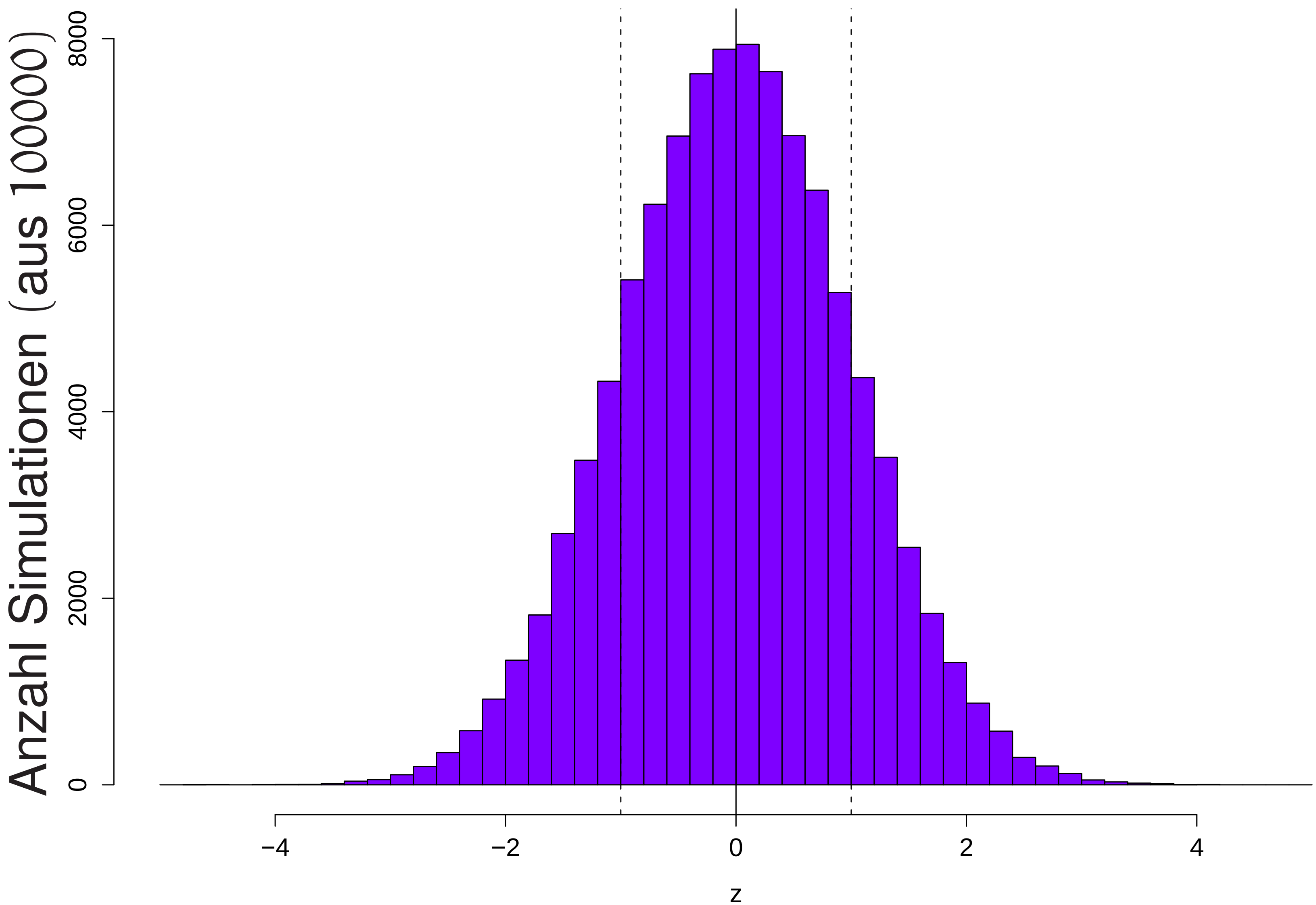
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 80)$



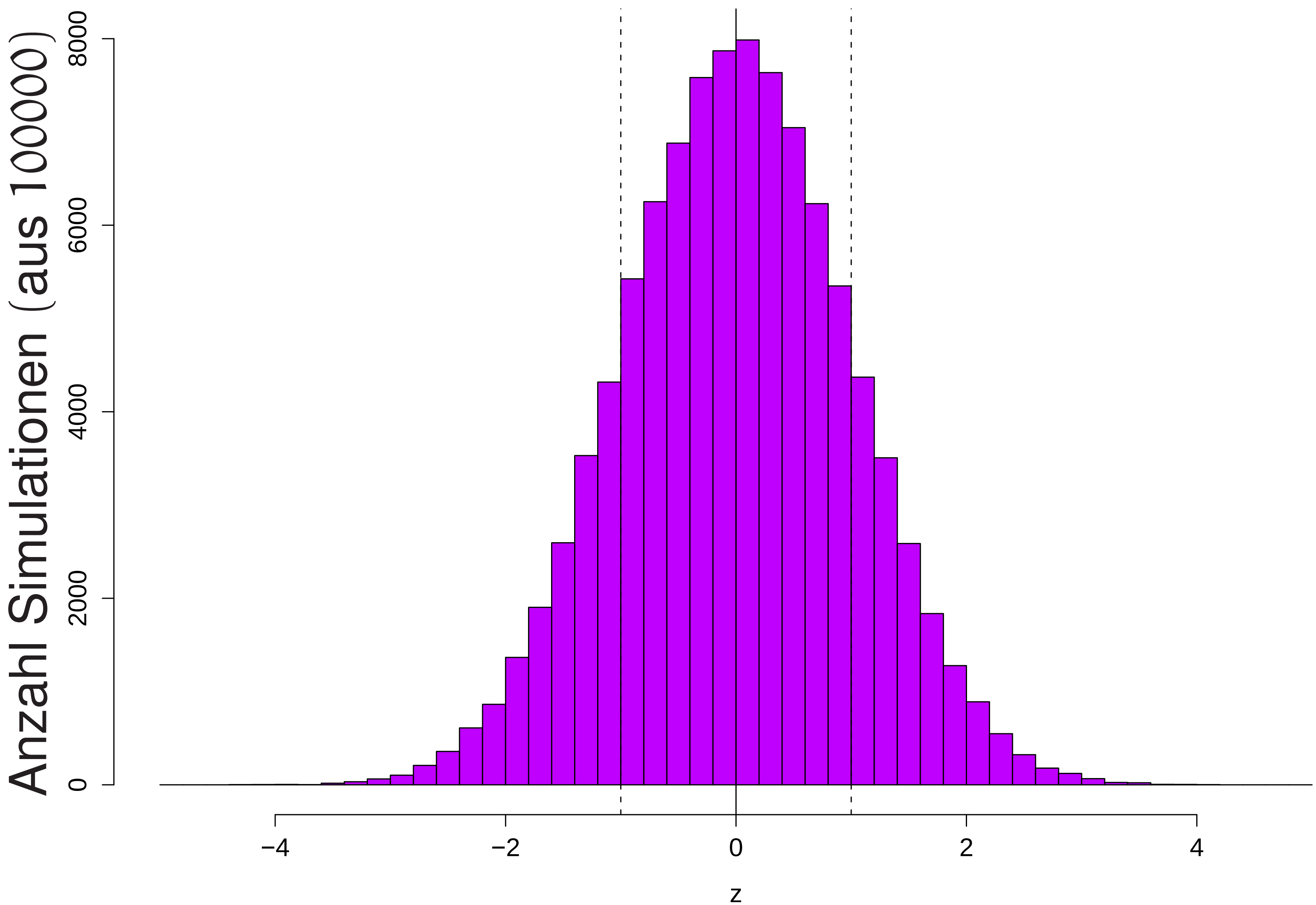


Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 85)$
Anzahl Simulationen (aus 100000)
0
2000
4000
6000
8000
−4
−2
0
2
4
z

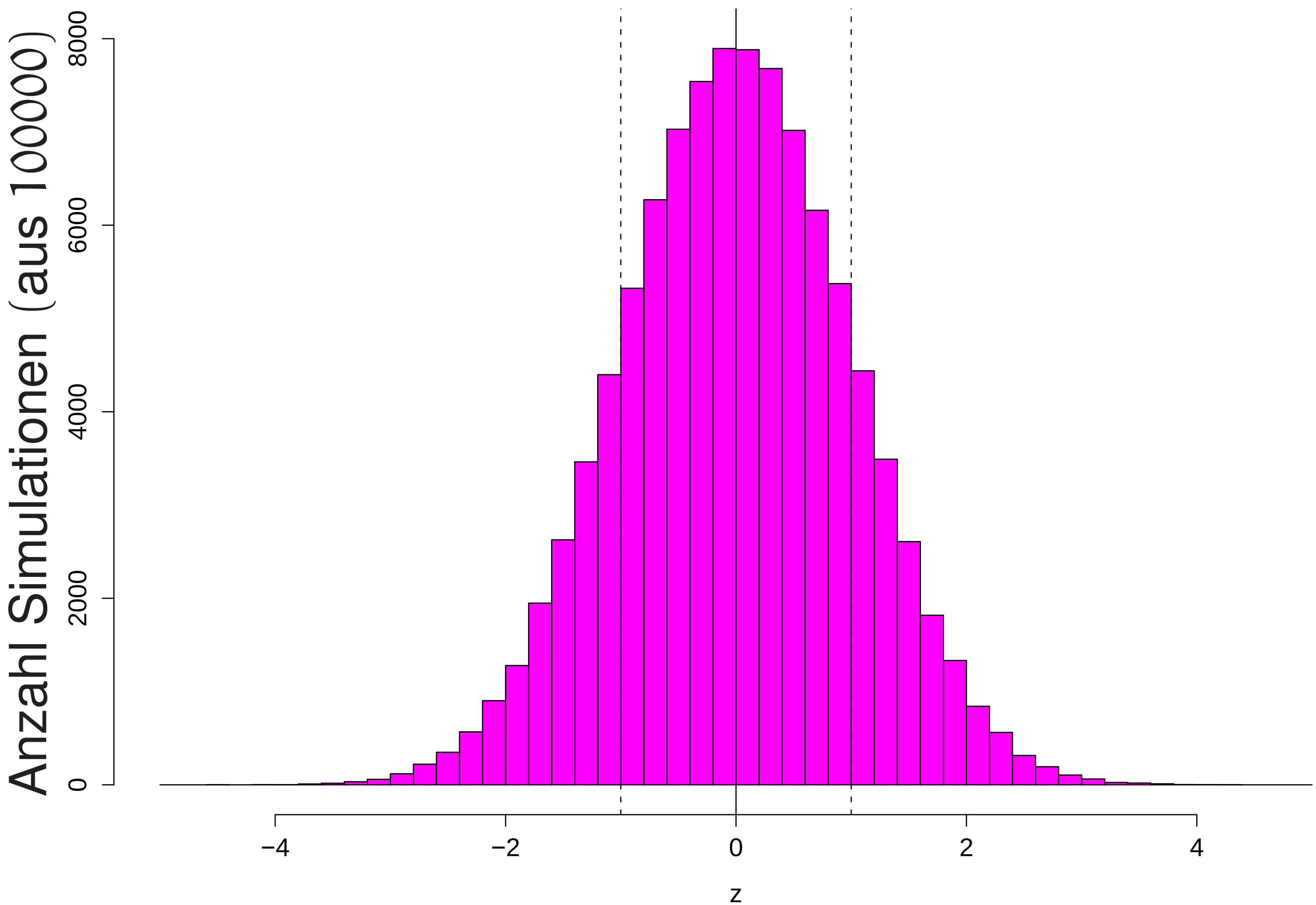
Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 90)$



Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 95)$



Standardisierung: $Z_n := (S_n - \mathbf{E}S_n)/\sigma_{S_n} \quad (n = 100)$



Die Verteilung von Z_n

scheint zu konvergieren.

Die Verteilung von Z_n

scheint zu konvergieren.

Welche Form

hat die Grenzverteilung?

Die Verteilung von

Z_{100}

ist glockenförmig.

Die Verteilung von

Z_{100}

ist glockenförmig.

Welche Glocke?

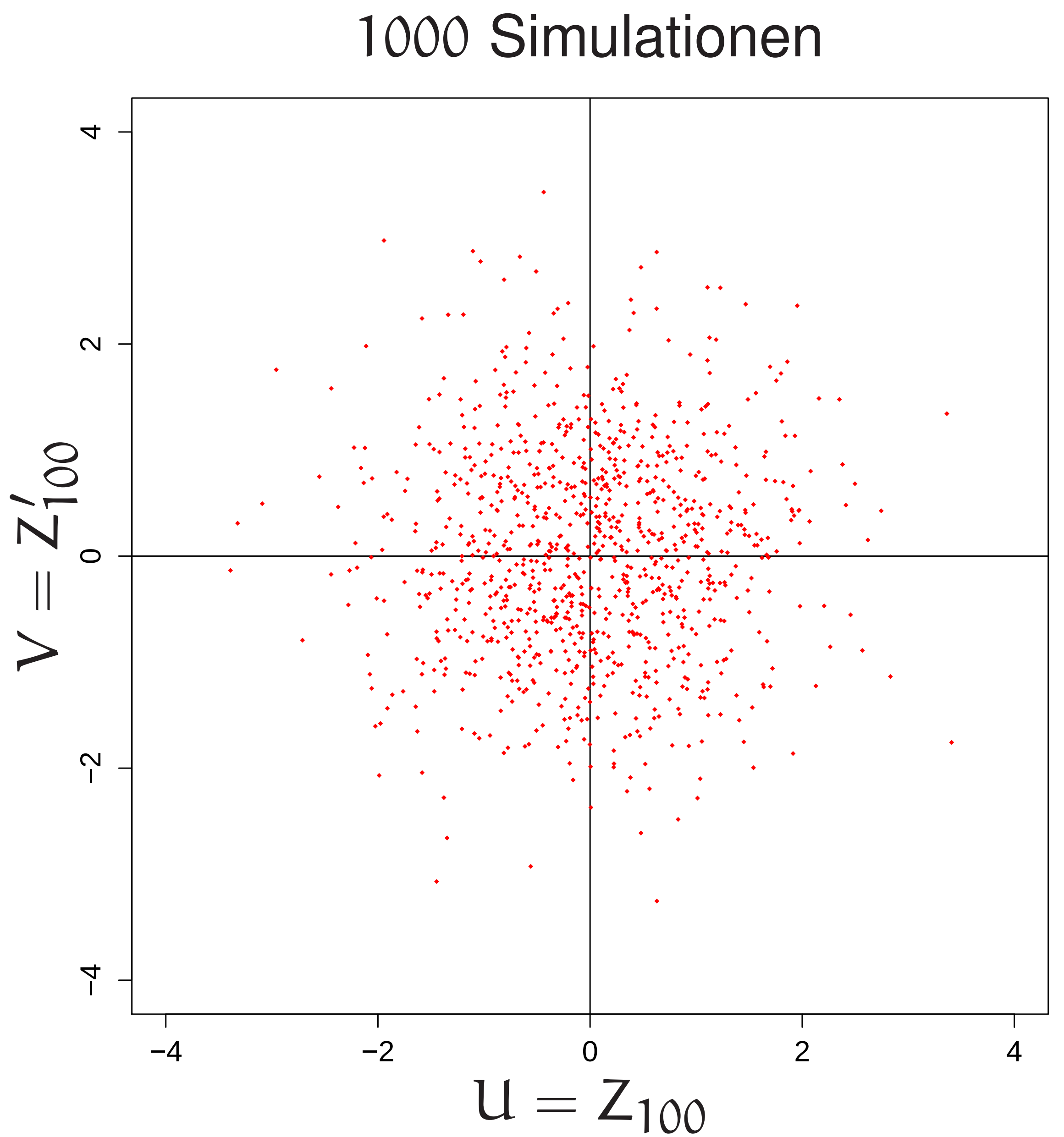
Glücklicher Einfall:

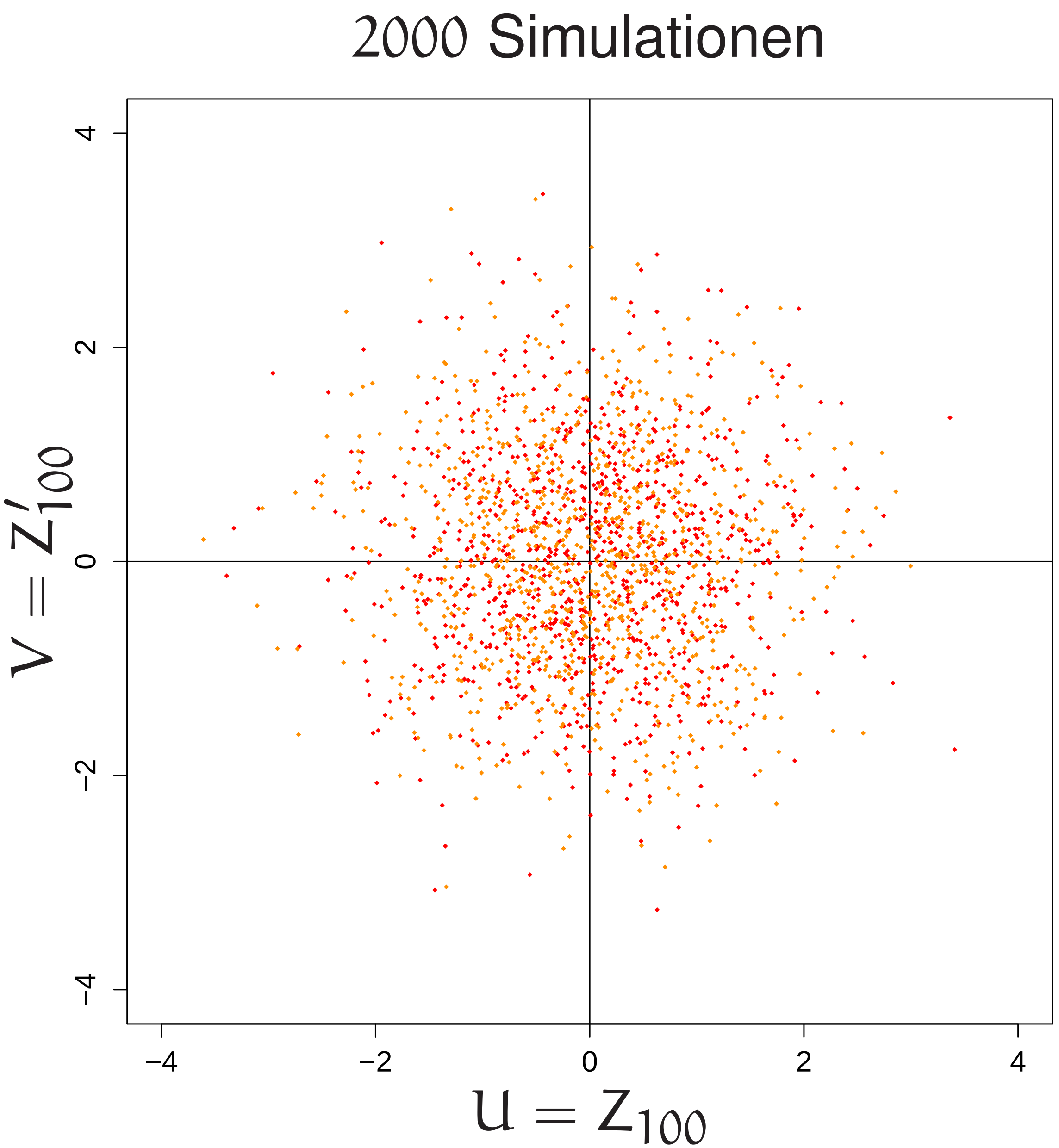
Zwei unabhängige Kopien

$$(U, V) = (Z_{100}, Z'_{100})$$

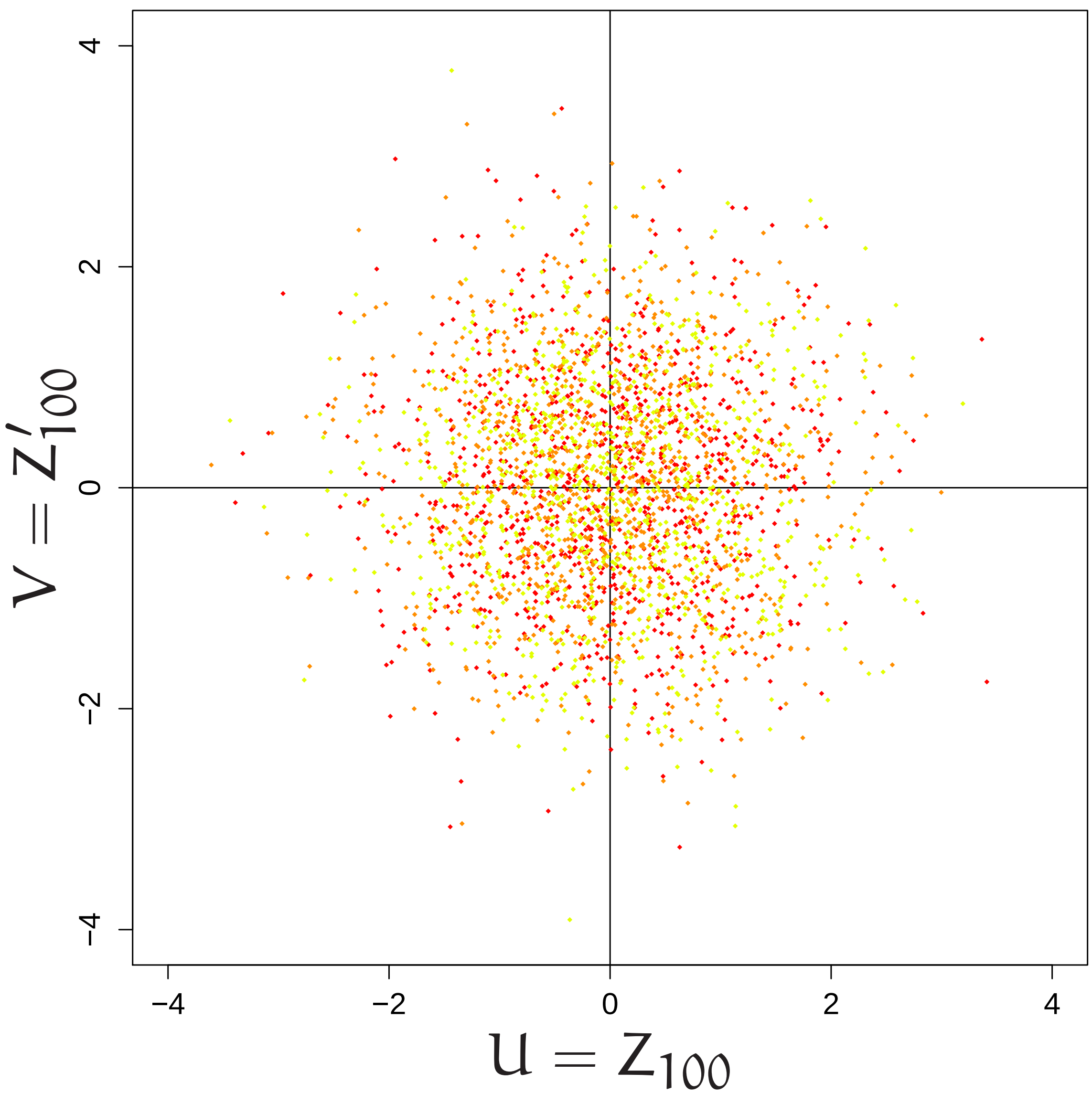
Wie sieht die gemeinsame Verteilung

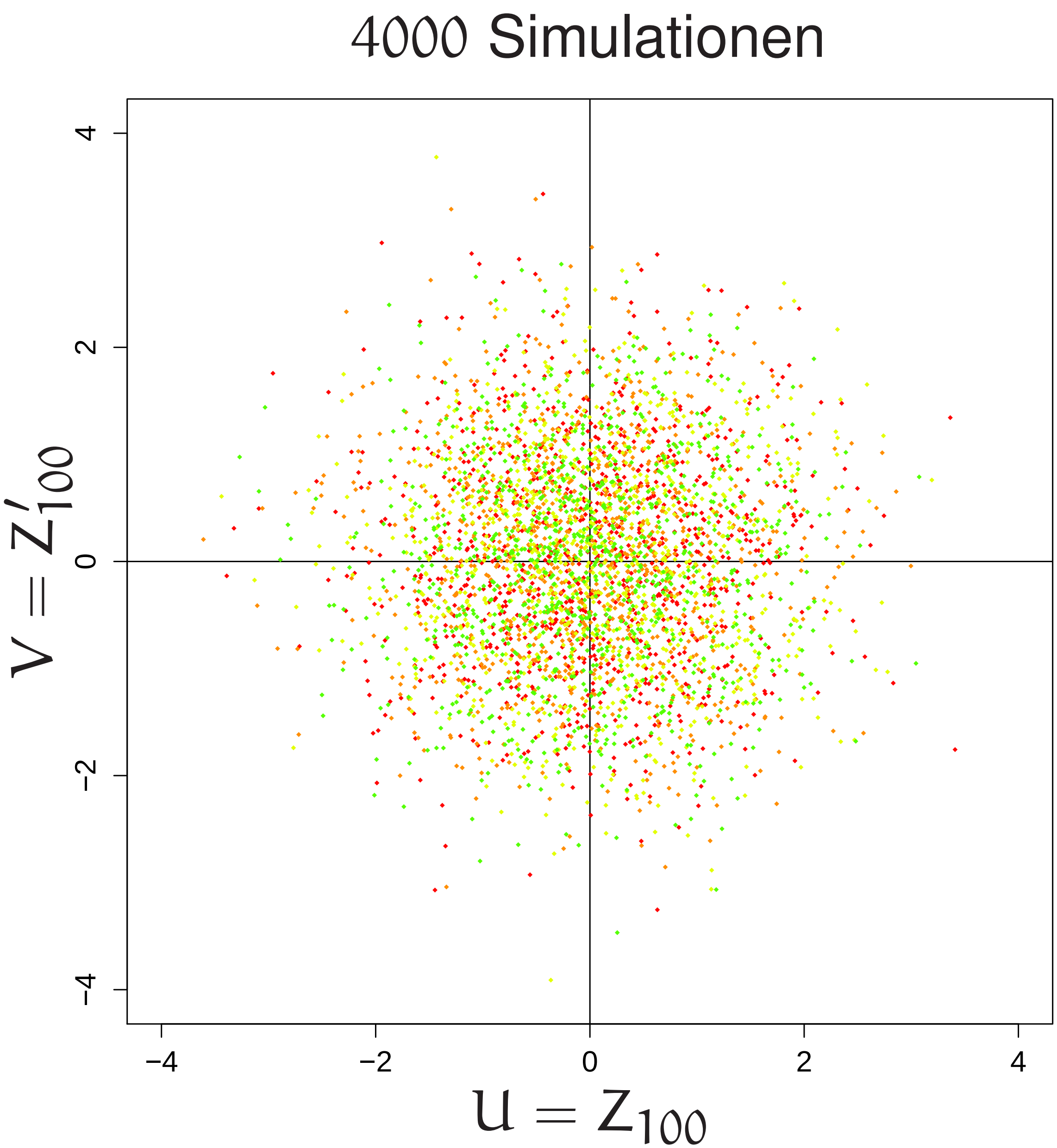
von U und V aus?


1000 Simulationen
−4
−2
0
2
4
V = Z'₁₀₀
U = Z₁₀₀

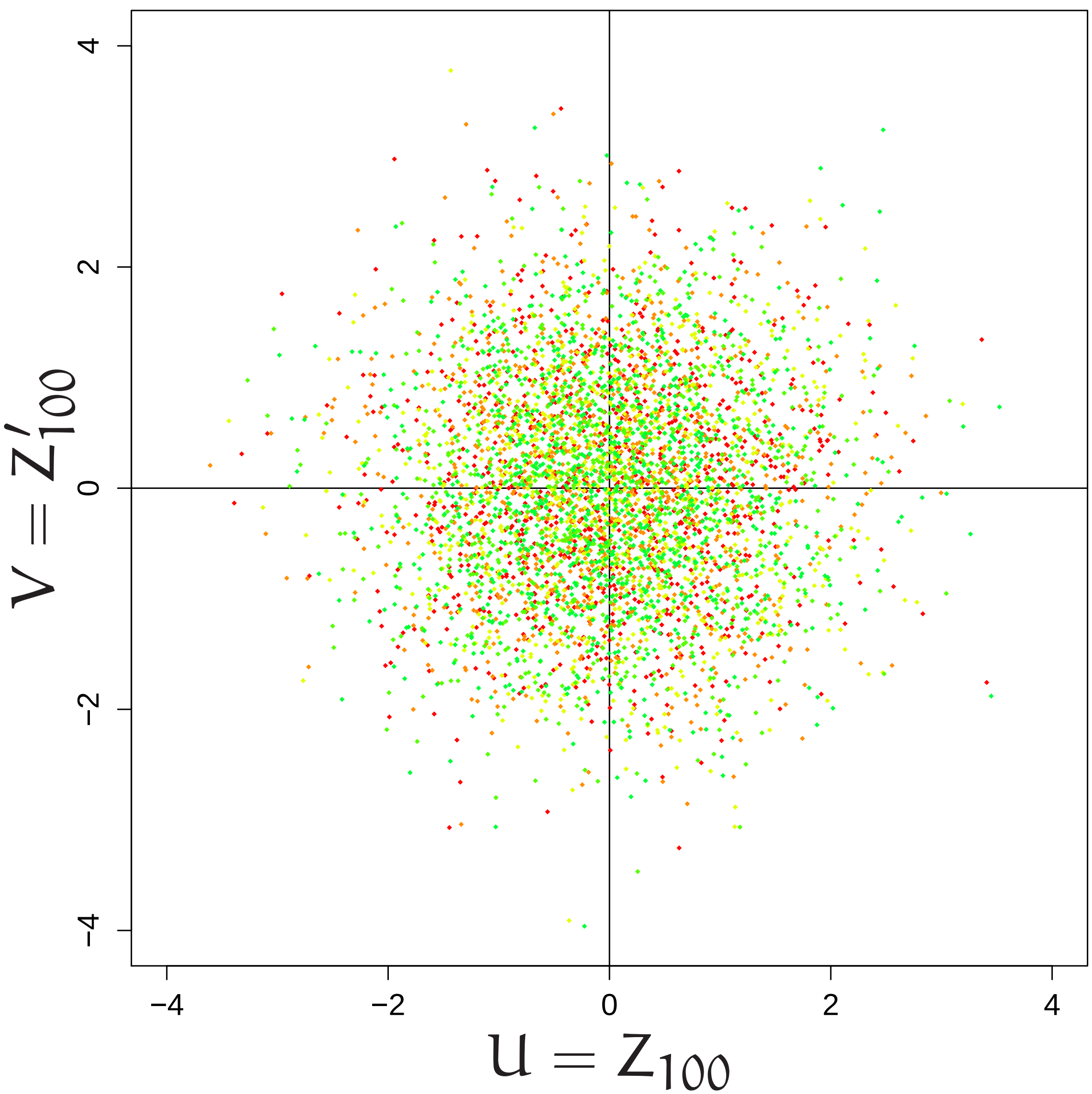

2000 Simulationen
4
2
0
−2
−4
−4
−2
0
2
4
$V = Z'_{100}$
$U = Z_{100}$

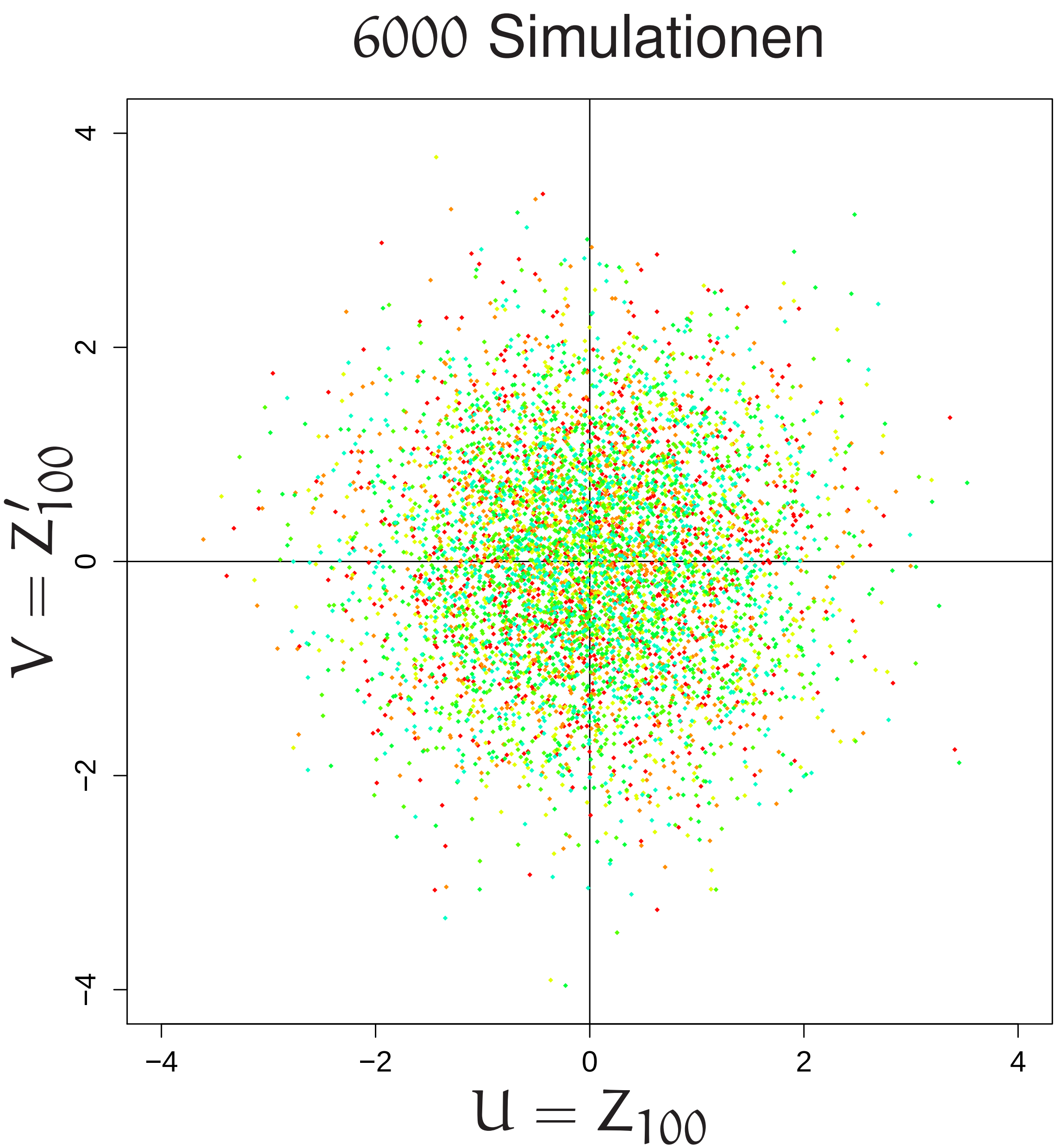
3000 Simulationen



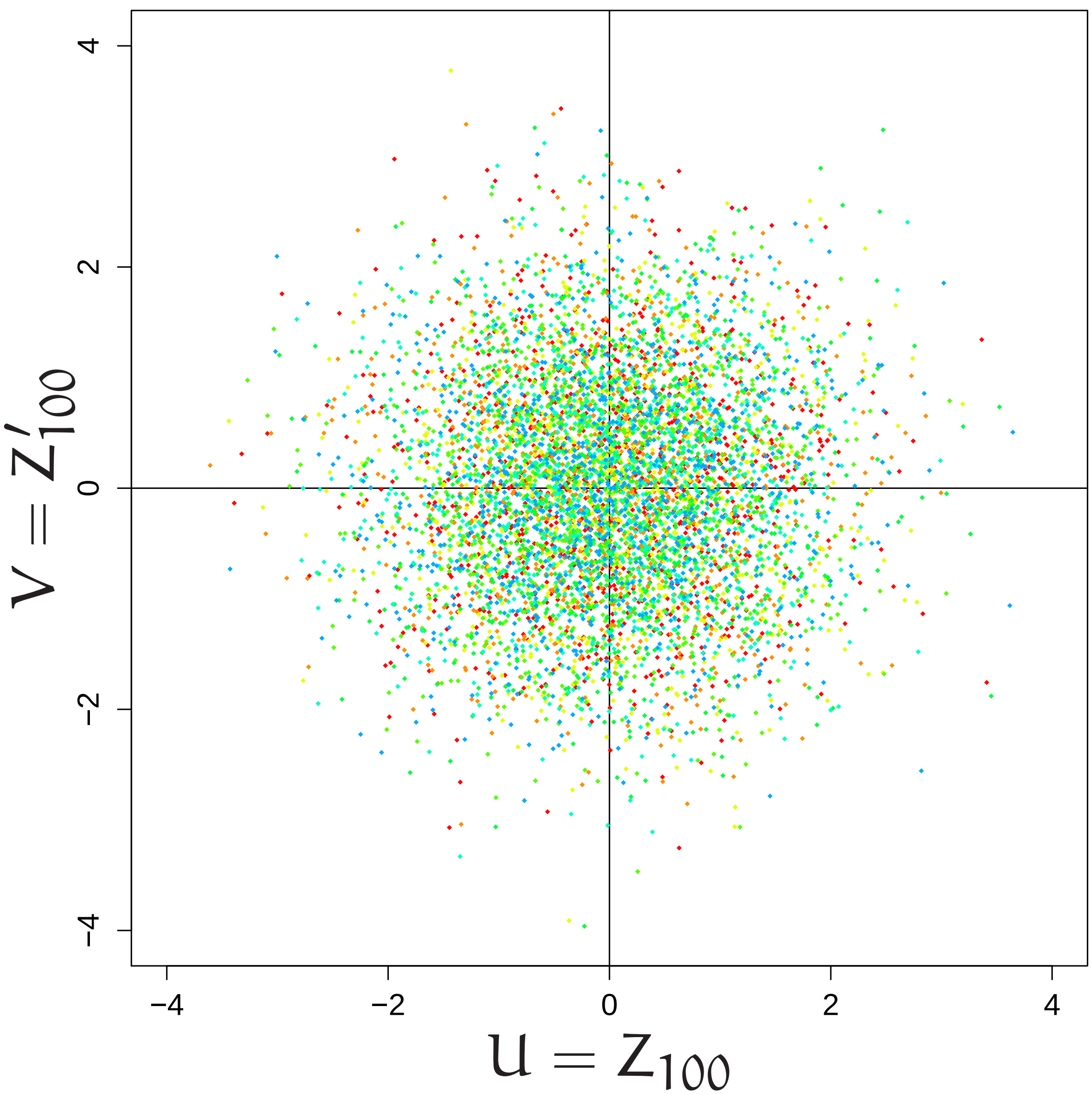

4000 Simulationen
4
2
0
−2
−4
−4
−2
0
2
4
V = Z'100
U = Z100

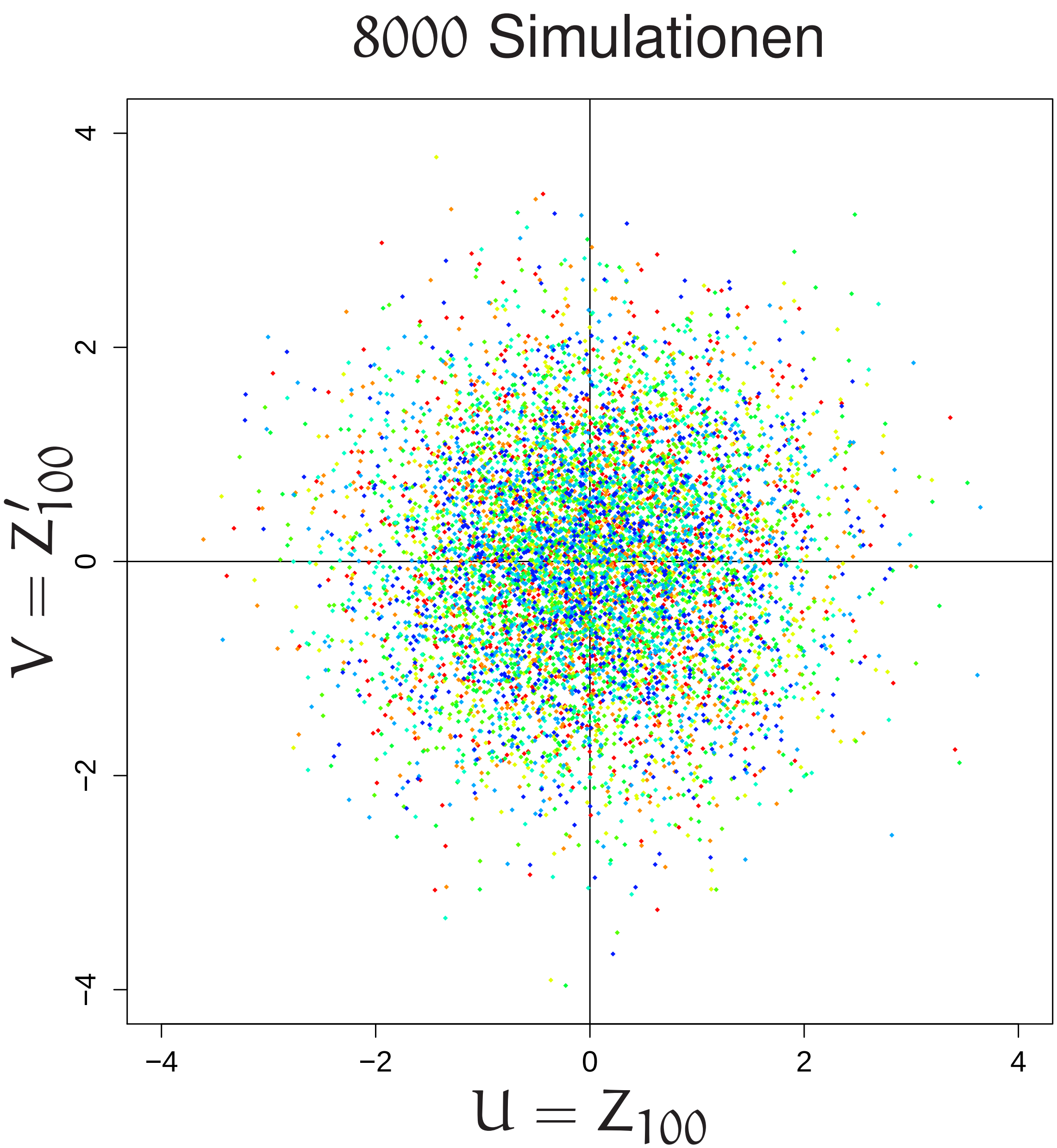
5000 Simulationen



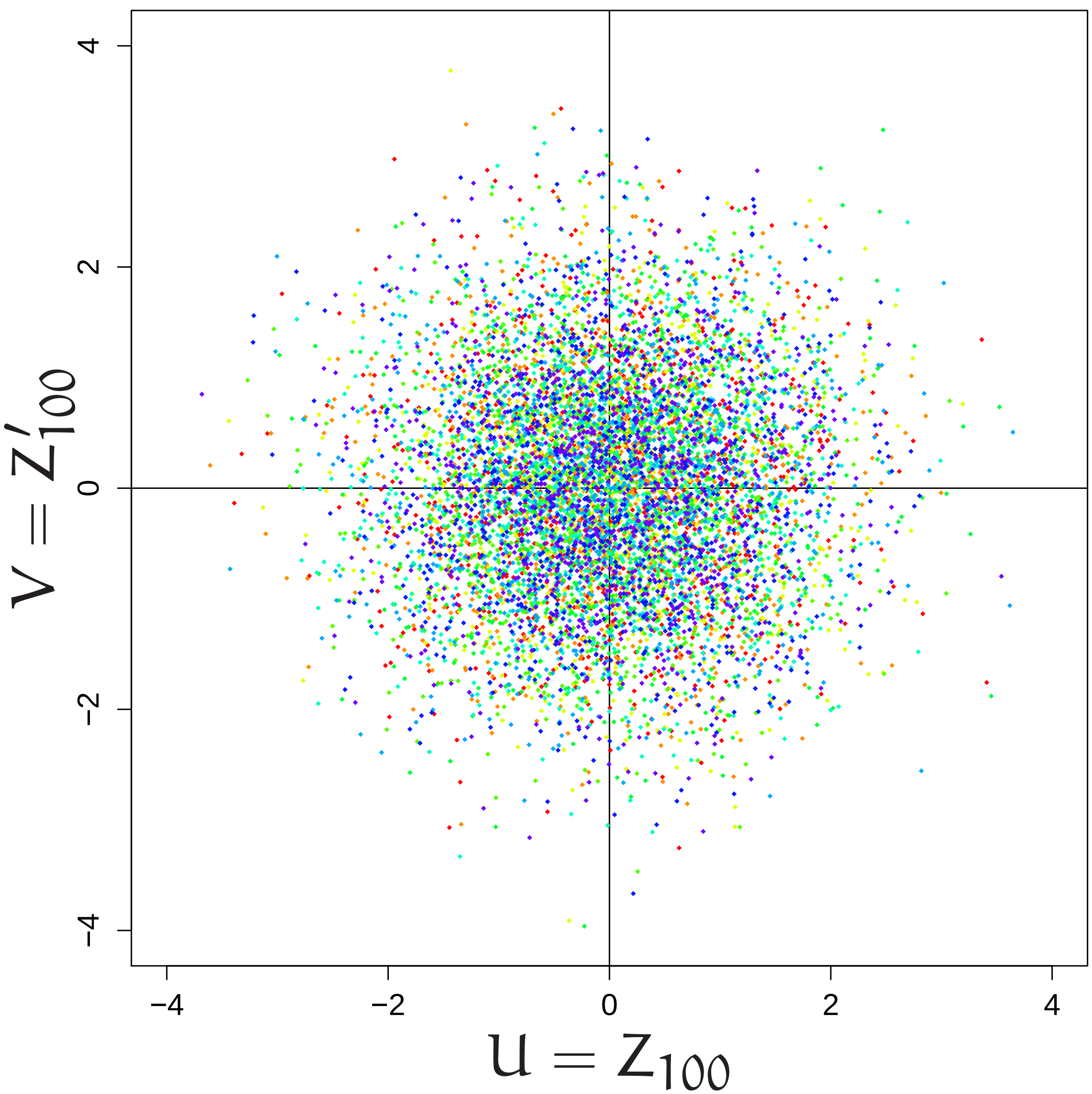

6000 Simulationen
4
2
0
−2
−4
−4
−2
0
2
4
$V = Z'_{100}$
$U = Z_{100}$

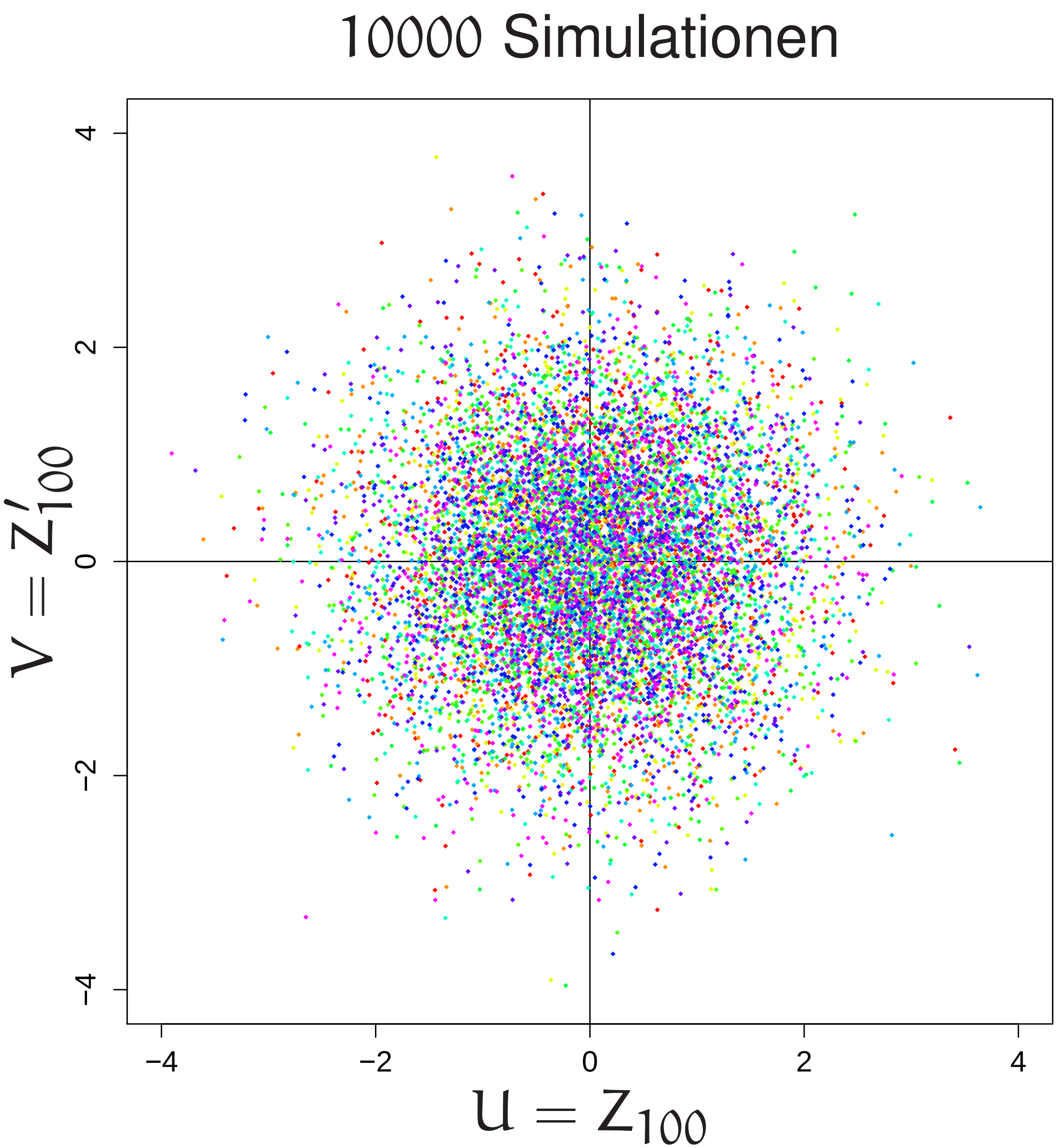
7000 Simulationen




8000 Simulationen
4
2
0
−2
−4
−4
−2
0
2
4
$V = Z'_{100}$
$U = Z_{100}$

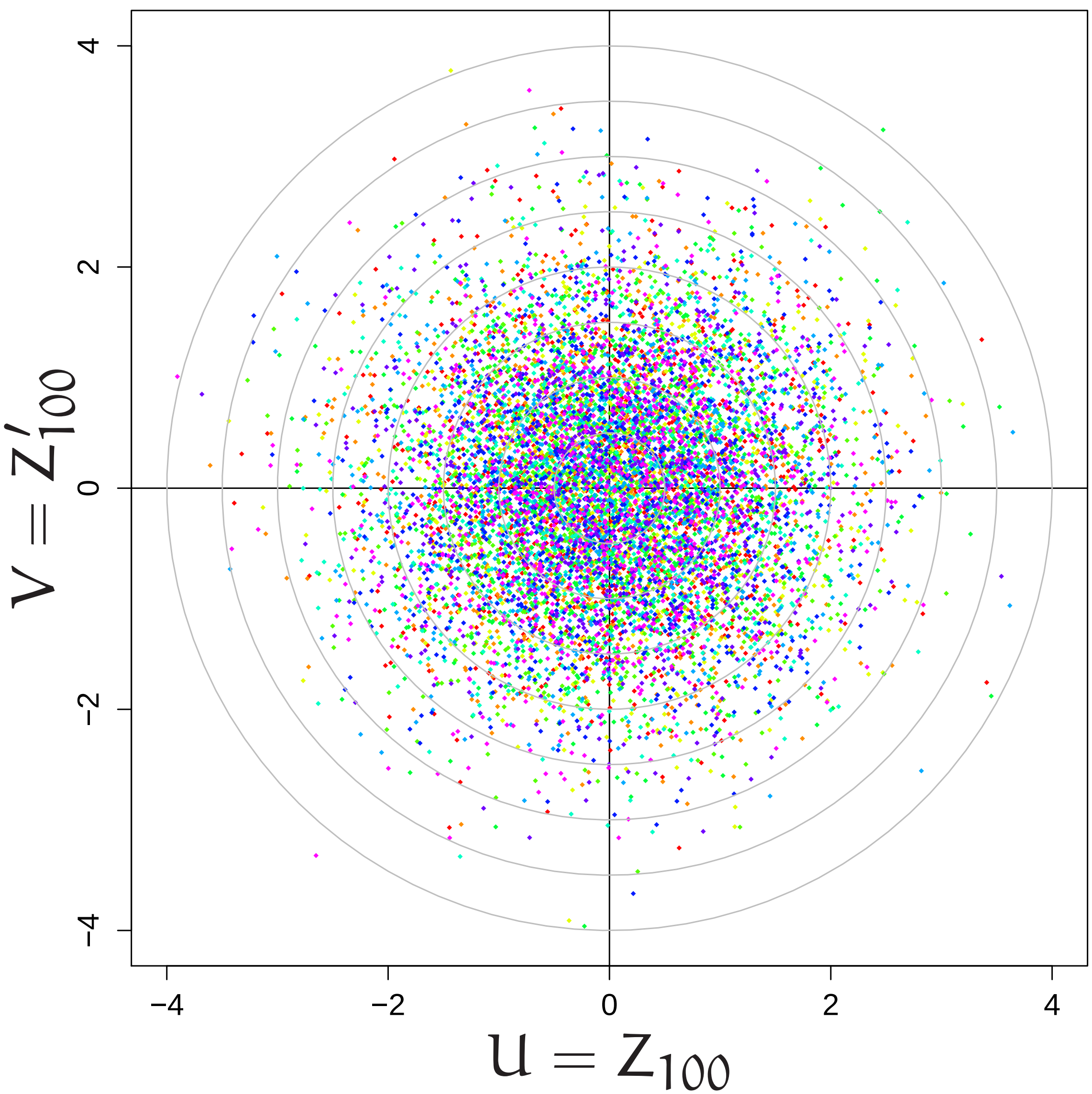
9000 Simulationen



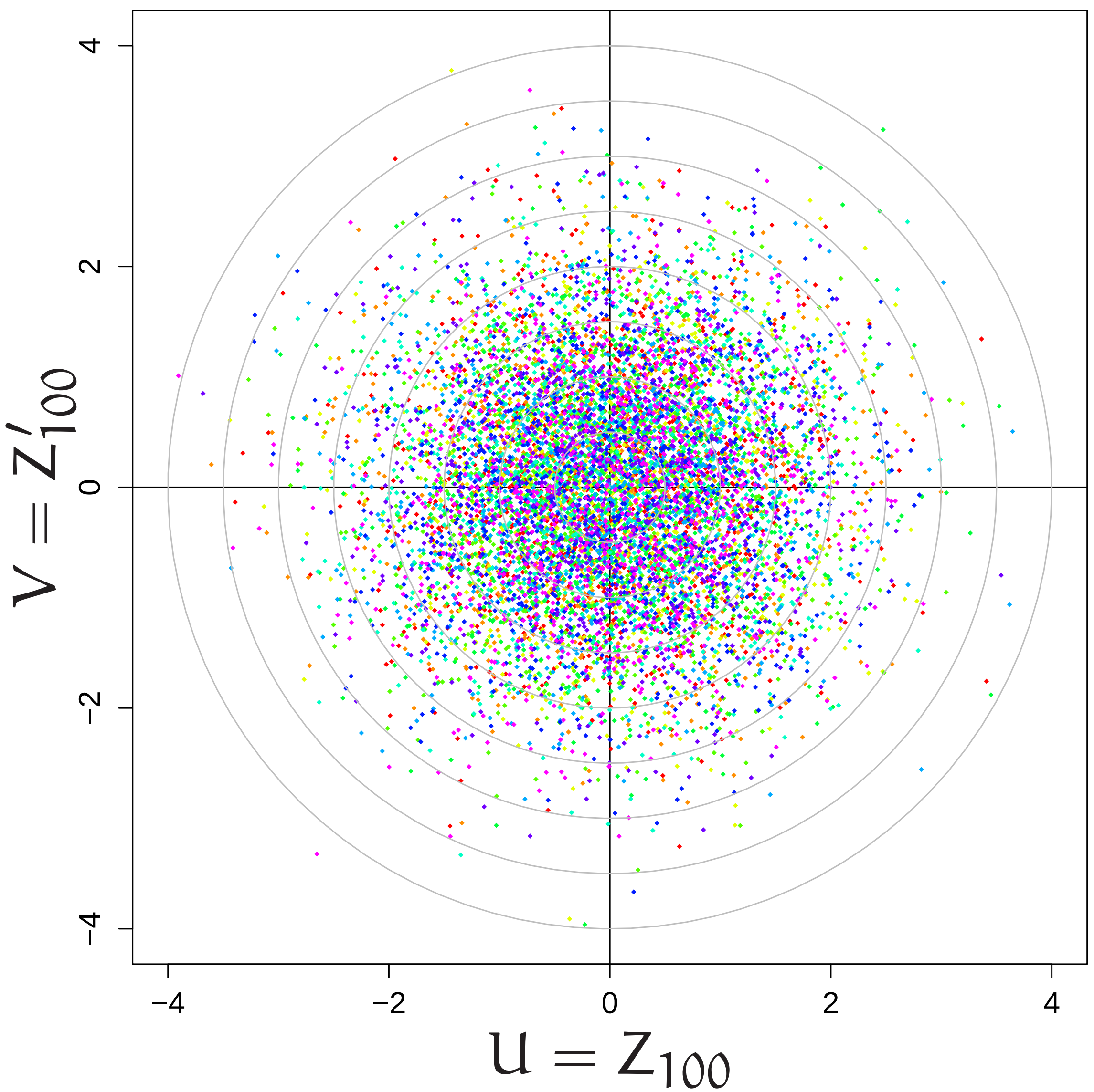


10000 Simulationen
4
2
0
−2
−4
−4
−2
0
2
4
$V = Z'_{100}$
$U = Z_{100}$

10000 Simulationen



Die Verteilung von (U, V) ist rotationssymmetrisch!



Behauptung:

Aus “U und V unabhängig und identisch verteilt’

und

“Verteilung von (U, V) rotationssymmetrisch”

folgt,

dass U und V normalverteilt sind:

$$f_U(x) = f_V(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/(2\sigma^2)}$$

Denn:

U, V unabhängig bedeutet:

$$f_{(U,V)}(a, b) = f_U(a) f_V(b)$$

$f_{(U,V)}$ *rotationssymmetrisch* heißt: es existiert ein g mit

$$f_{(U,V)}(a, b) = g(r) \qquad r := \sqrt{a^2 + b^2}$$

Denn:

U, V unabhängig bedeutet:

$$f_{(U,V)}(a, b) = f_U(a) f_V(b)$$

$f_{(U,V)}$ *rotationssymmetrisch* heißt: es existiert ein g mit

$$f_{(U,V)}(a, b) = g(r) \qquad r := \sqrt{a^2 + b^2}$$

Mit $f_U = f_V =: h$ folgt

$$g(r) = h(a)h(b)$$

$$g(r) = h(a)\,h(b), \qquad r := \sqrt{a^2 + b^2}$$

$$a := 0, b := r$$

$$g(r) = h(0) \cdot h(r)$$

$$h(a)\,h(b) = h(0)\,h(\sqrt{a^2 + b^2})$$

Eine Lösung:

e$h(x) = e^{-x^2}$

Denn

$$e^{-a^2} e^{-b^2} = 1 \cdot e^{-(a^2+b^2)}$$

$$h(a)\,h(b) = h(0)\,h(\sqrt{a^2+b^2})$$

Wie sieht die allgemeine Lösung aus?

$w(x) := h(\sqrt{|x|})$ erfüllt

$$w(a^2)w(b^2) = w(0)w(a^2+b^2)$$

$$w(u)w(v) = k_0 w(u+v), \quad u, v \geq 0$$

hat als allgemeine Lösung

$$w(x) = k_0 e^{-k_1 x}$$

$$h(x) = k_0 e^{-k_1 x^2}$$

FAZIT

Der Zentrale Grenzwertsatz

lässt sich erraten

(in konkreten Fällen,

mit etwas Glück).

Der Münzwurf passt in den Zentralen Grenzwertsatz:

Seien $X_1, X_2, \dots$ unabhängige und identisch verteilte Zufallsvariable mit endlichem Erwartungswert μ und endlicher Varianz $\sigma^2 > 0$. Dann gilt für alle $\ell < r \in \mathbb{R}$

$$\mathbf{P}\left(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \in [\ell, r]\right) \underset{n\to\infty}{\longrightarrow} \mathbf{P}(Z \in [\ell, r]).$$

Dabei ist Z standard-normalverteilt.

Ist X_i eine Bernoullifolge (mit $\mu := p$ und $\sigma^2 := pq$), so ergibt sich der alte Satz von de Moivre und Laplace.

Hier ist noch einmal die (im ZGS präzisierte) Botschaft der Stunde:

Summen (und Mittelwerte) von vielen unabhängigen,
identisch verteilten ZV mit endlicher Varianz
sind annähernd normalverteilt.

Diese Aussage bleibt übrigens auch
unter schwächeren Bedingungen bestehen,
sowohl was die Unabhängigkeit,
als auch was die identische Verteiltheit betrifft.

Eine Botschaft zum Mitnehmen ins Leben:

“Die Summe von vielen kleinen, annähernd unabhängigen Zufallsvariablen

ist annähernd normalverteilt.”