Vorlesung 6b

Die Normalverteilung

Dichten auf $\mathbb{R}^n$

und

Unabhängigkeit bei Dichten

1. Binomialverteilungen
mit großem Erwartungswert und großer Varianz

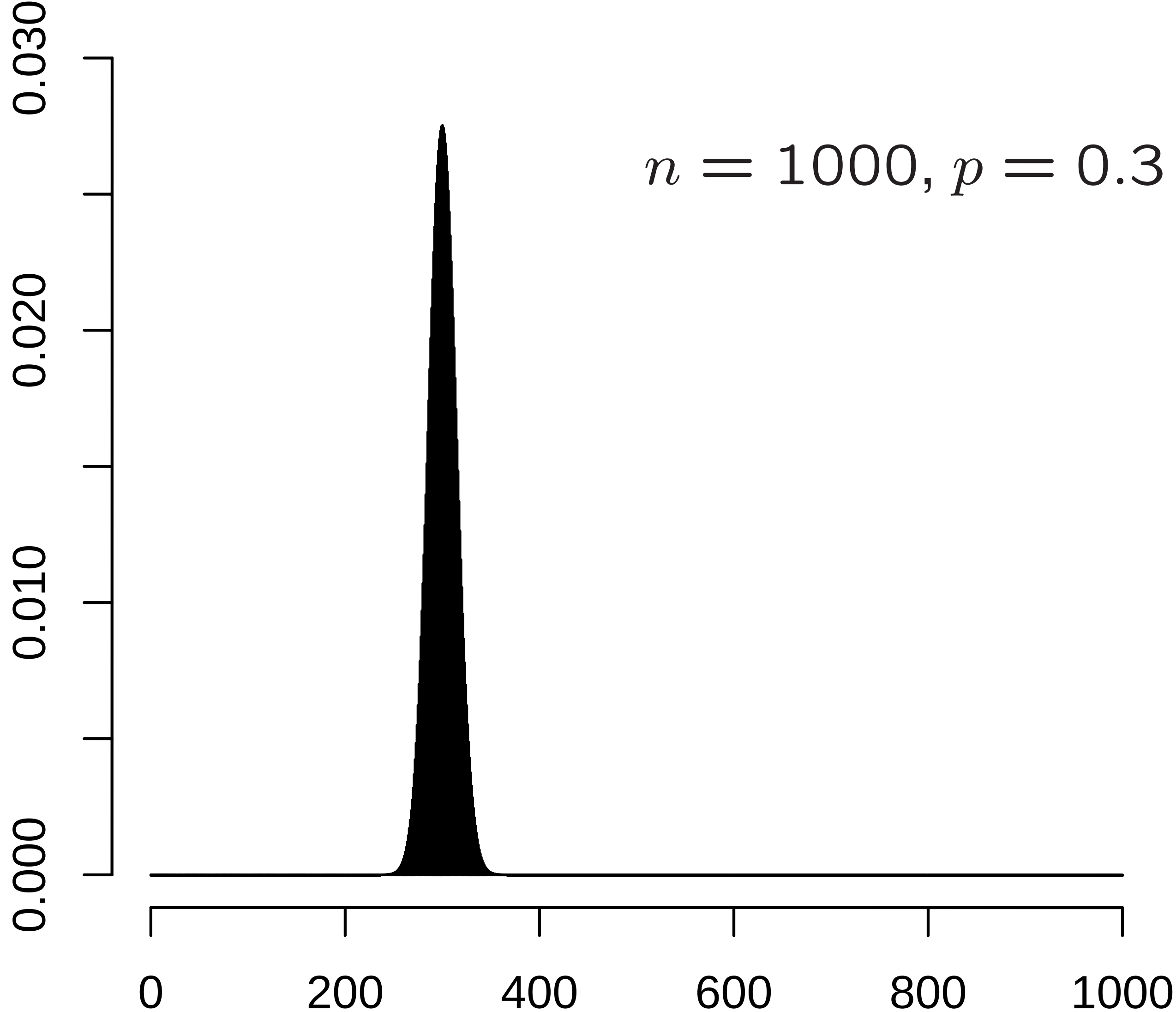
Wie sieht die Bin$(n, 1/2)$-Verteilung

für großes n aus,

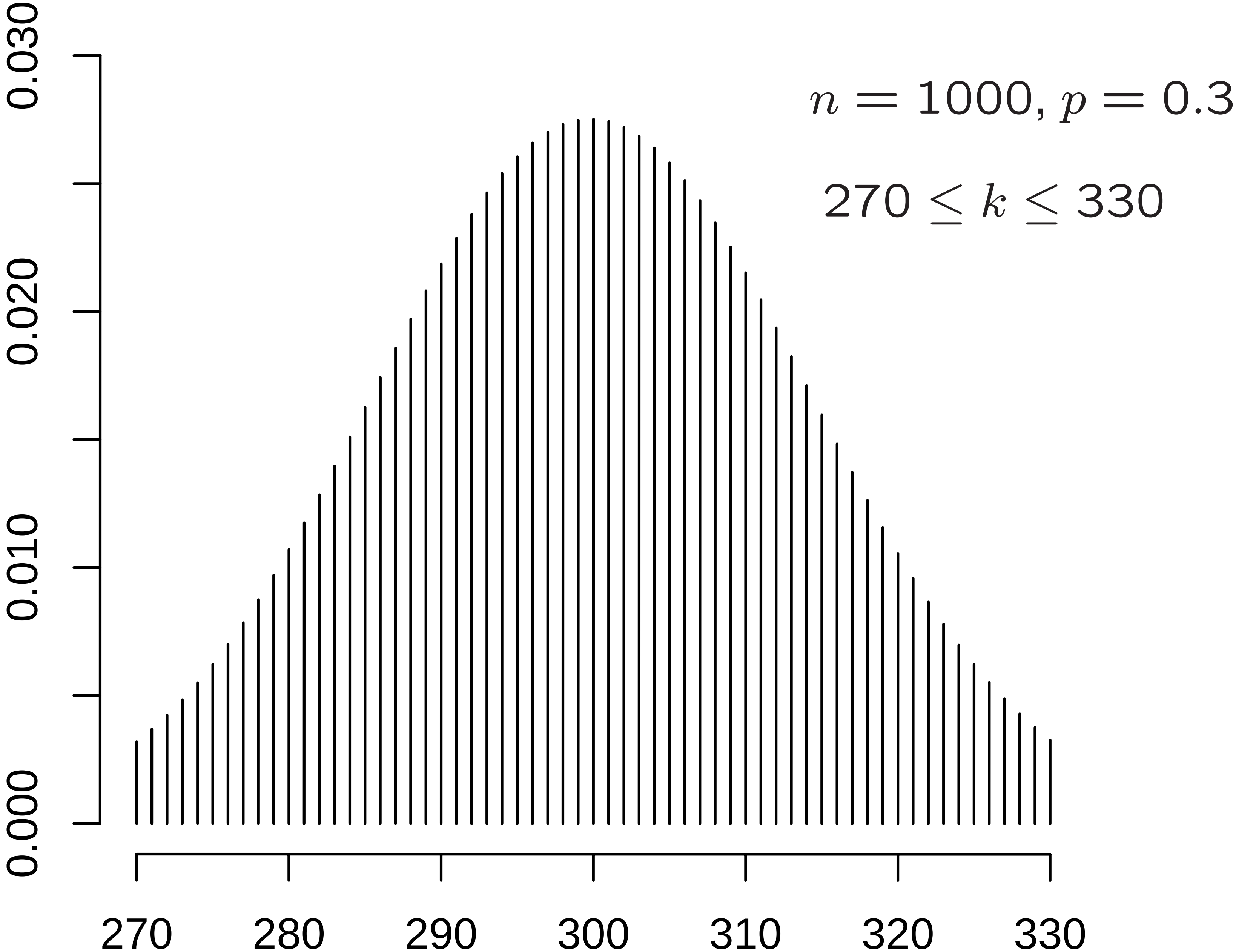
oder allgemeiner

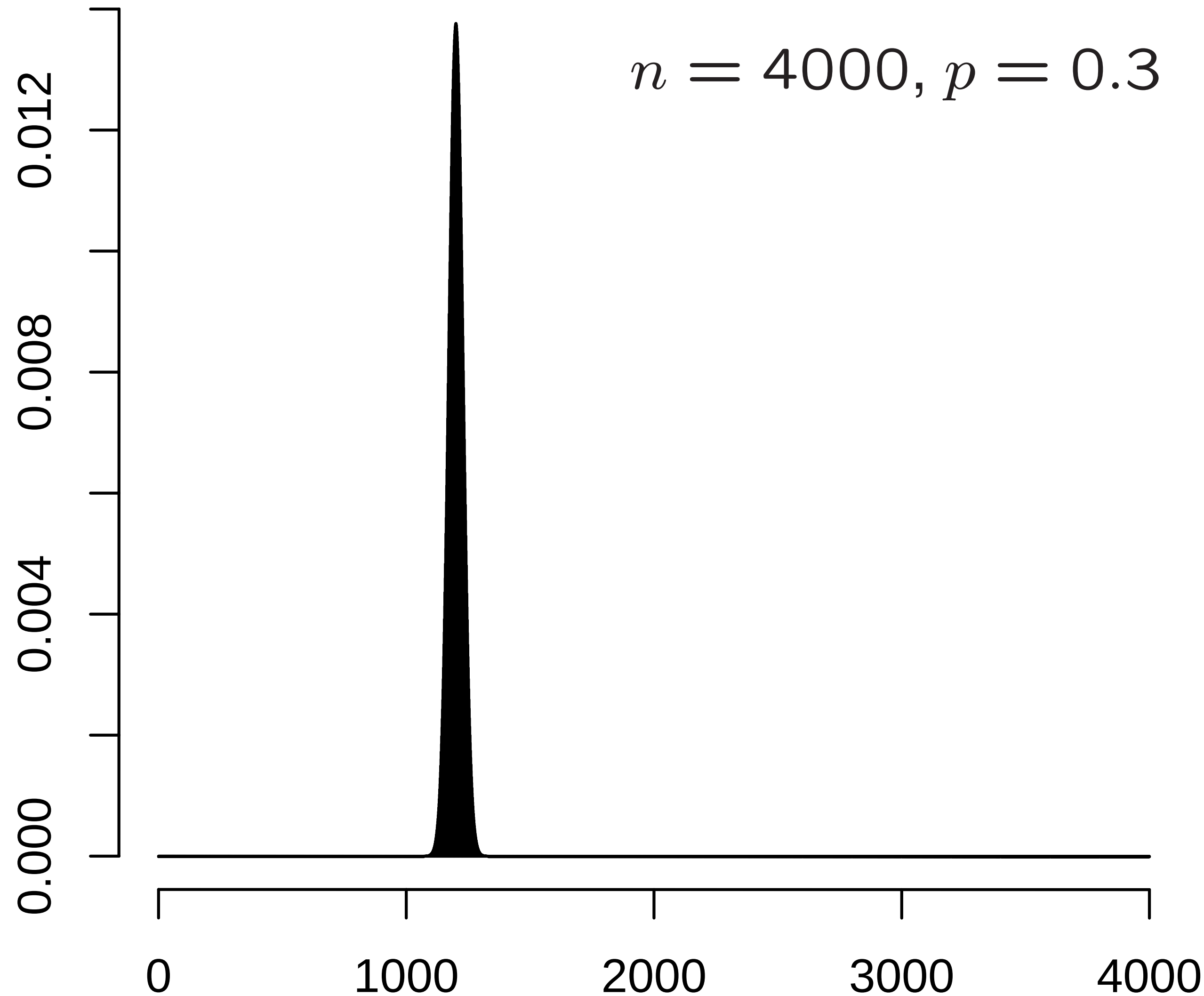
die Bin(n, p)-Verteilung

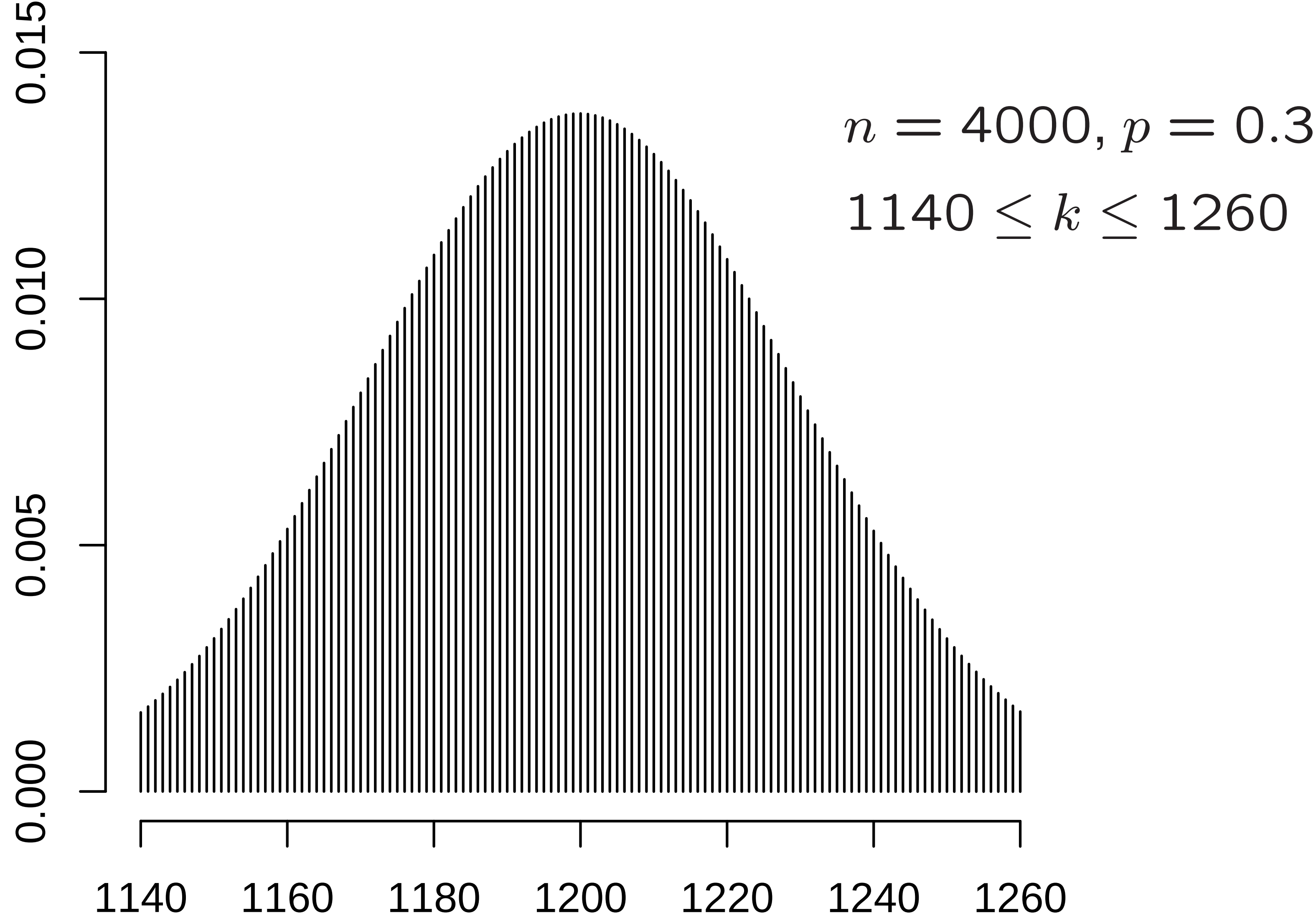
mit großem n und großem npq ?



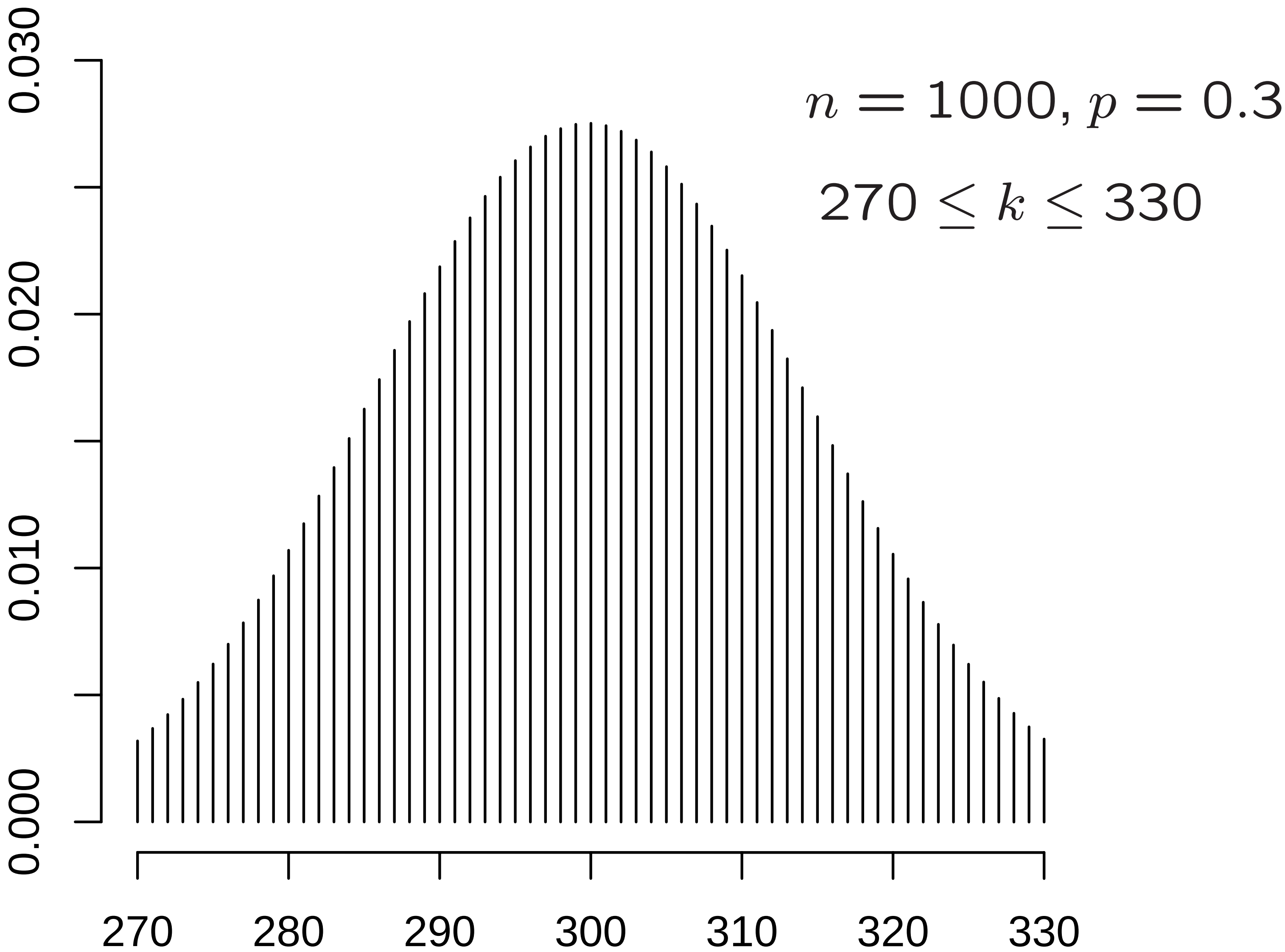
$n = 1000, p = 0.3$
0.000
0.010
0.020
0.030
0
200
400
600
800
1000


$n = 1000, p = 0.3$
$270 \leq k \leq 330$
0.000
0.010
0.020
0.030
270
280
290
300
310
320
330


$n = 4000, p = 0.3$
0.000
0.004
0.008
0.012
0
1000
2000
3000
4000



$n = 4000, p = 0.3$
$1140 \leq k \leq 1260$
0.000
0.005
0.010
0.015
1140
1160
1180
1200
1220
1240
1260



$n = 1000, p = 0.3$
$270 \leq k \leq 330$
0.000
0.010
0.020
0.030
270
280
290
300
310
320
330

Binomialverteilungen mit großem n und großer Varianz npq

sehen "glockenförmig" aus,

wenn man sie geeignet "ins Bild holt".

Approximation der Binomialgewichte

mit “Stirling & Taylor” (vgl Buch S. 27):

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2} \Big(\frac{k-np}{\sqrt{npq}} \Big)^2 \Big) .$$

Approximation der Binomialgewichte

mit "Stirling & Taylor" (vgl Buch S. 27):

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}\Big(\frac{k-np}{\sqrt{npq}}\Big)^2\Big)$$

$$\mu := np, \quad \sigma := \sqrt{npq}$$

Approximation der Binomialgewichte

mit “Stirling & Taylor” (vgl Buch S. 27):

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sigma} \quad \frac{1}{\sqrt{2\pi}} \exp\Big(- \frac{1}{2} \Big(\frac{k-\mu}{\sigma} \Big)^2 \Big)$$

$$\mu := np, \quad \sigma := \sqrt{npq}$$

Approximation der Binomialgewichte mit “Stirling & Taylor” (vgl Buch S. 27):

$$\binom{n}{k} p^k q^{n-k} \approx \quad \frac{1}{\sigma} \quad \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2\right)$$

$$= \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right)$$

mit $\mu := np$, $\sigma := \sqrt{npq}$ und

$$\varphi(a) := \frac{1}{\sqrt{2\pi}} e^{-a^2/2}, \quad a \in \mathbb{R}.$$

$$\binom{n}{k} p^k q^{n-k} \approx \quad \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right) \qquad (*)$$

mit $\mu := np,\ \sigma := \sqrt{npq}$ und

$\varphi(a) := \frac{1}{\sqrt{2\pi}} e^{-a^2/2}, \quad a \in \mathbb{R}$.

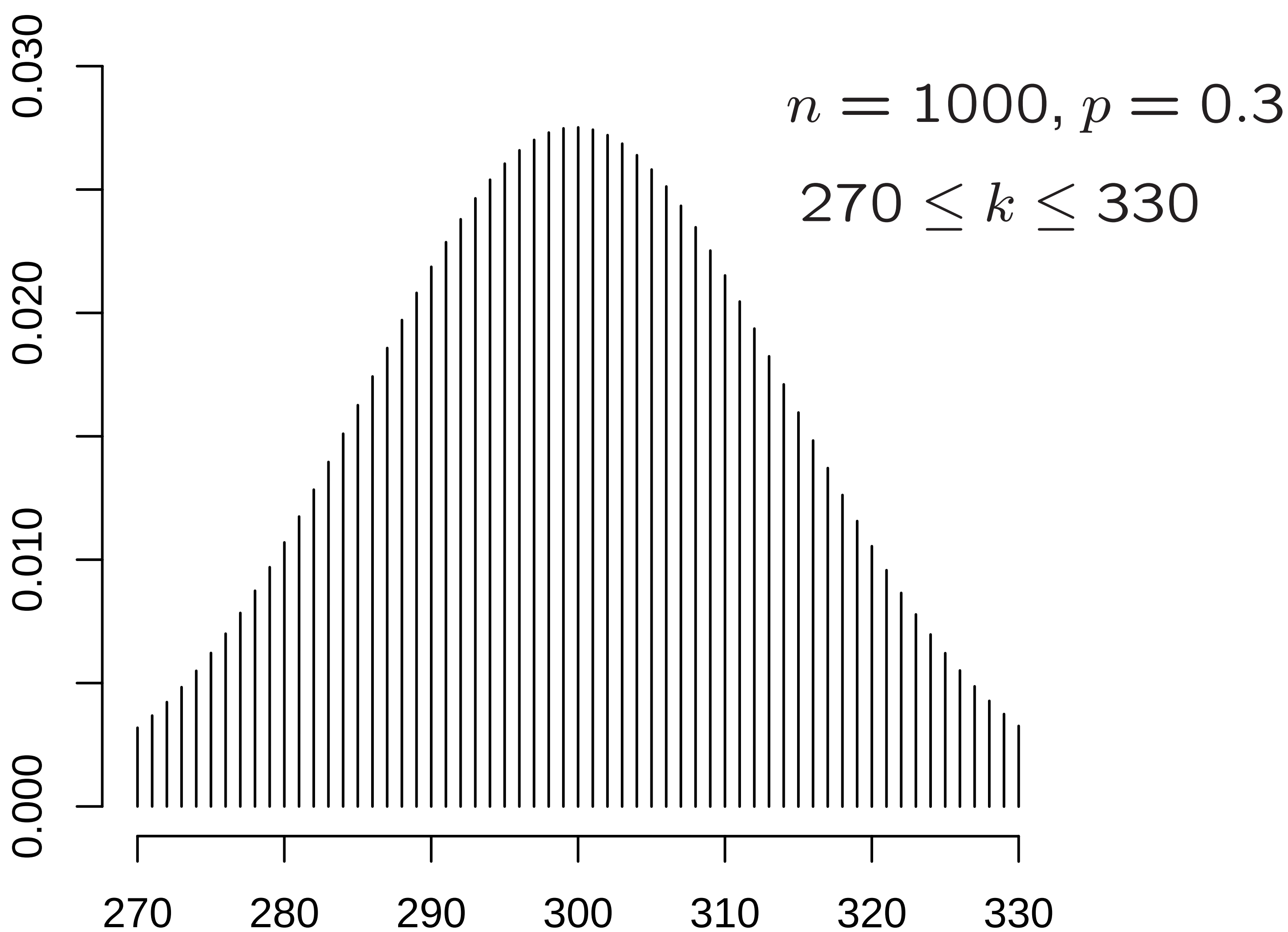
Wenn man n vervierfacht, verdoppelt sich σ.

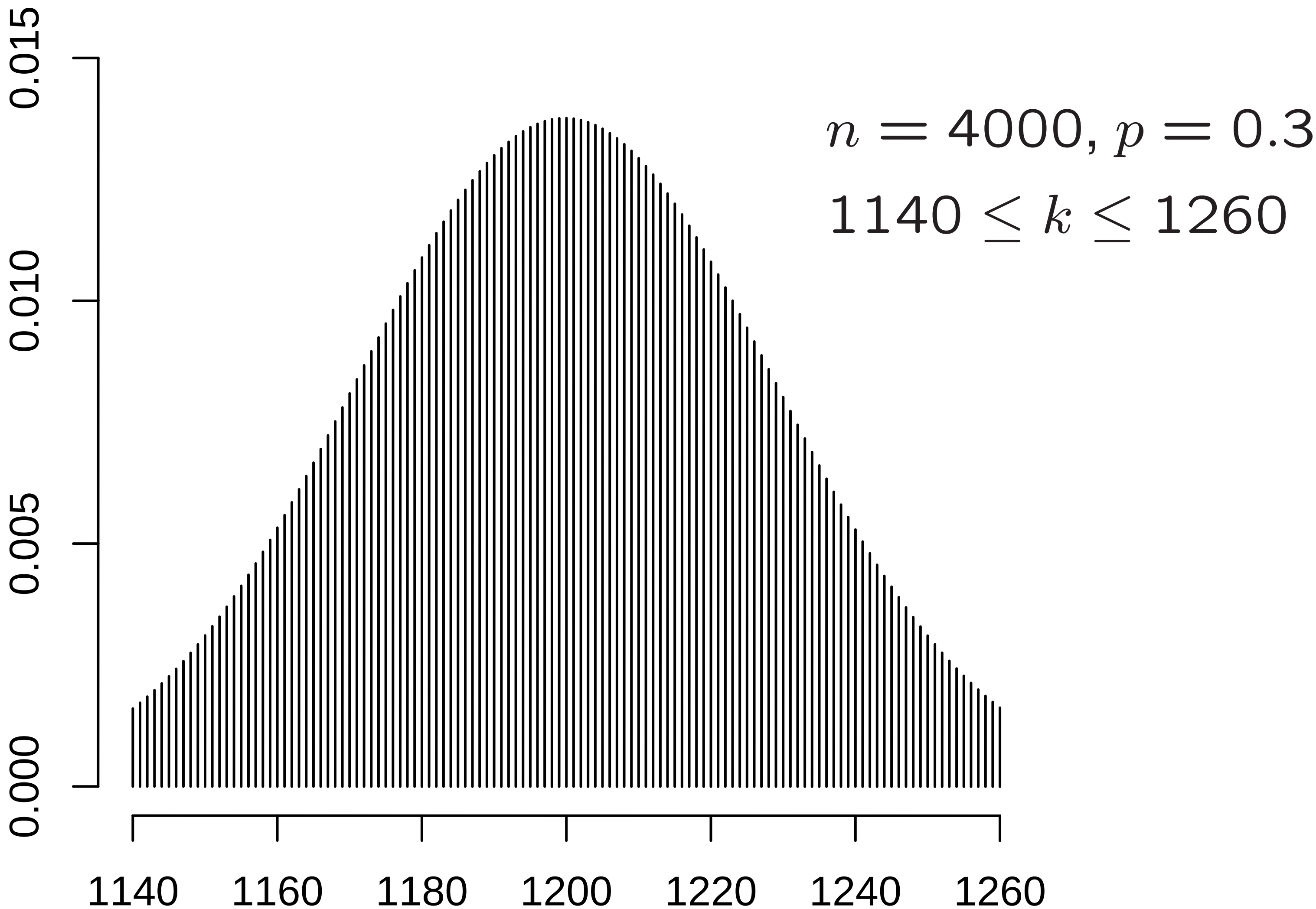
Approximativ gilt dann in der Darstellung $(*)$:

Im Bereich von einer Standardabweichung um das Zentrum μ

bringt man doppelt so viele k unter.

Das Gewicht jedes einzelnen k halbiert sich.


$n = 1000, p = 0.3$
$270 \le k \le 330$
0.000
0.010
0.020
0.030
270
280
290
300
310
320
330


$n = 4000, p = 0.3$
$1140 \leq k \leq 1260$
0.000
0.005
0.010
0.015
1140
1160
1180
1200
1220
1240
1260

2. Die Standard-Normalverteilung

Die nächste Definition beinhaltet

die wichtigste Verteilung der Stochastik: Eine $\mathbb{R}$-wertige

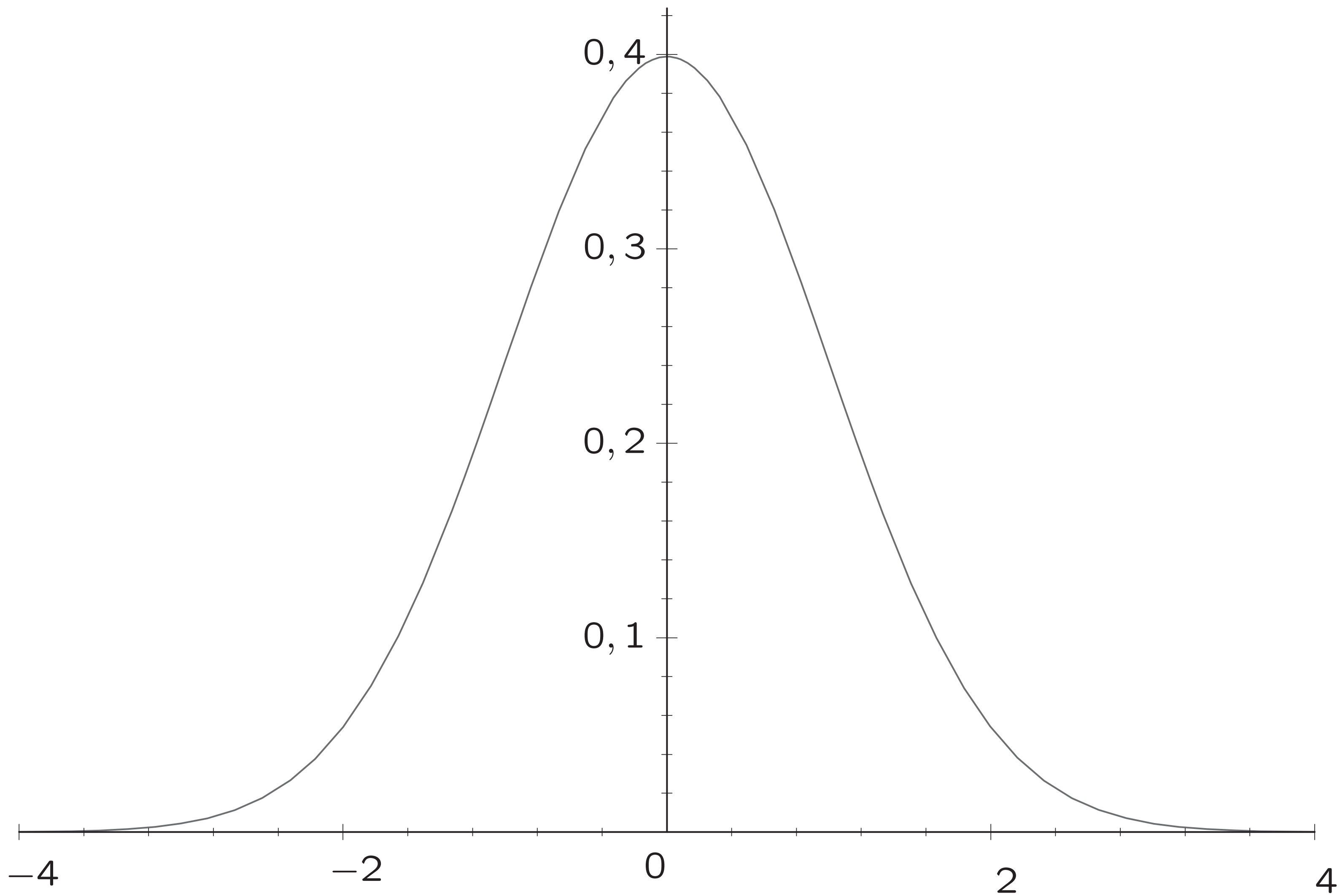
Zufallsvariable Z mit Dichte

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

heißt standard-normalverteilt.

Es lässt sich beweisen (vgl. die erste Aufgabe im Buch, S. 72):

$$\int_{-\infty}^{+\infty}\varphi(a)\,da = 1.$$


0,4
0,3
0,2
0,1
−4
−2
0
2
4

Für ein standard-normalverteiltes Z ist

$$\mathbf{E}[Z] = 0, \quad \mathbf{Var}\,[Z] = 1.$$

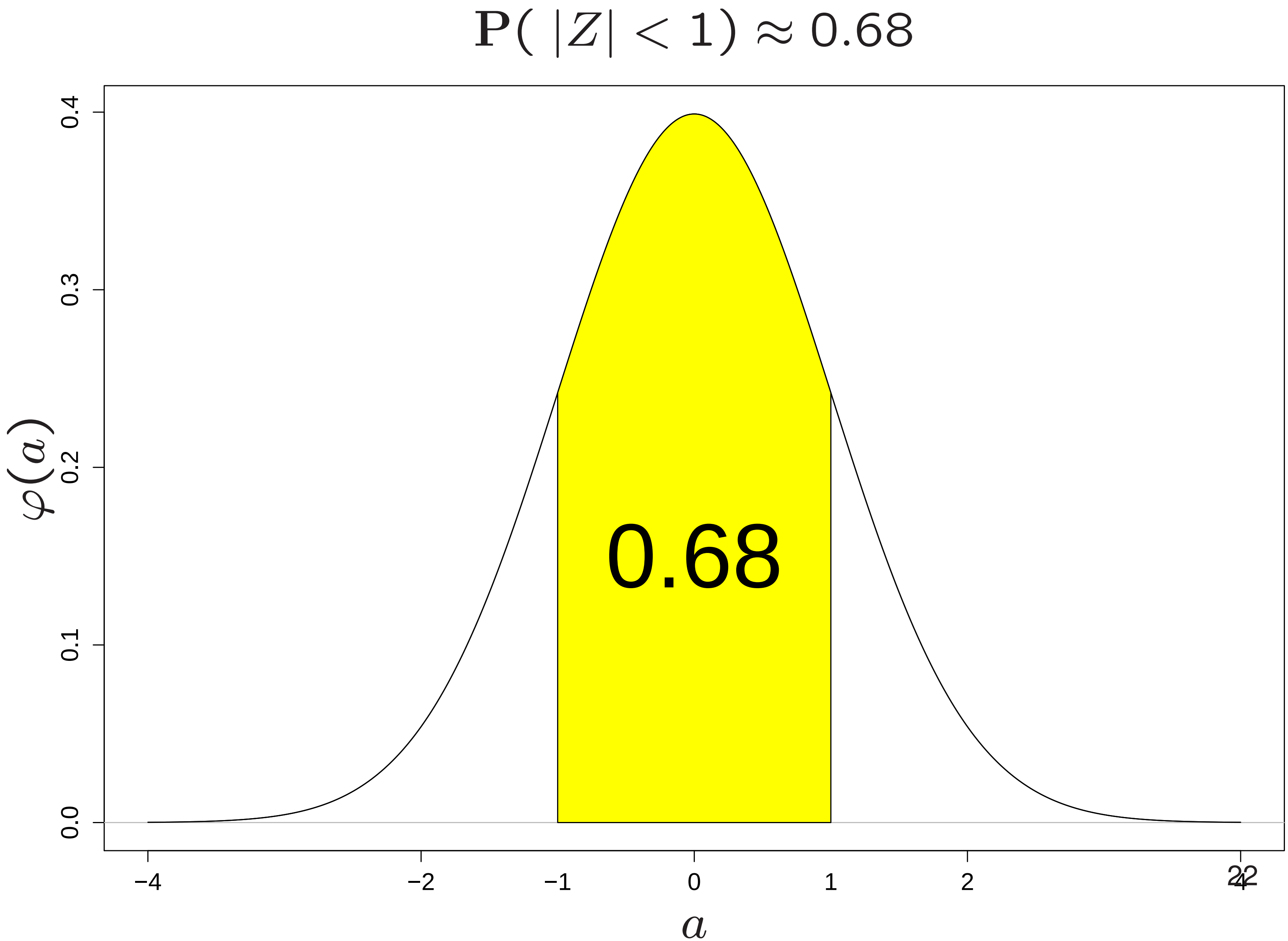
Denn aus Symmetriegründen ist $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} a\, e^{-a^2/2}\, da = 0$,

und mit partieller Integration bekommt man

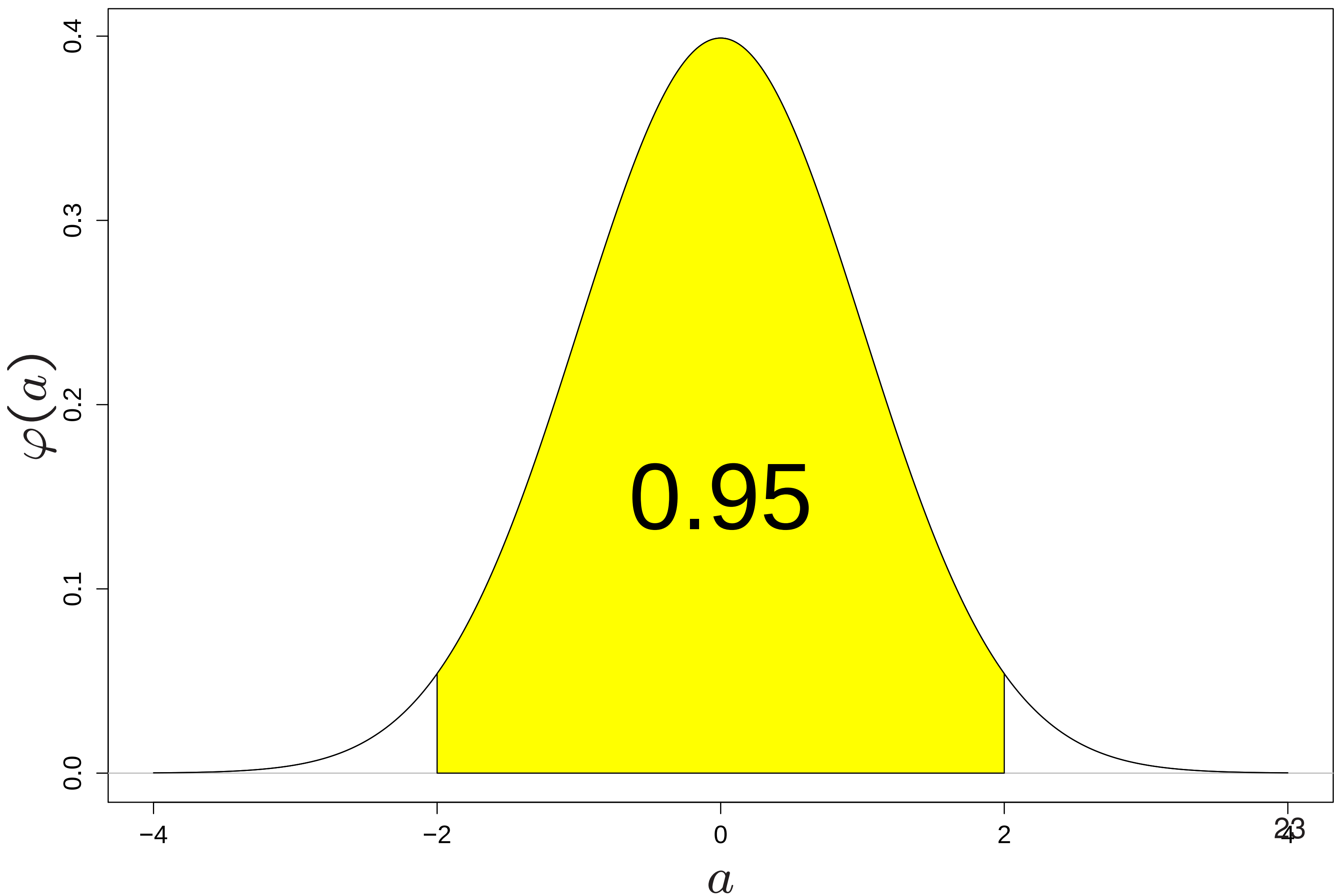
$$\int_{-\infty}^{\infty} a^2\, e^{-a^2/2}\, da = \int_{-\infty}^{\infty} a\, a e^{-a^2/2}\, da$$

$$= -a\, e^{-a^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-a^2/2}\, da = \sqrt{2\pi}.$$

Zwei für die Praxis wichtige Zahlen:


$\mathbf{P}(\ |Z| < 1) \approx 0.68$
0.68
$\varphi(a)$
0.0
0.1
0.2
0.3
0.4
−4
−2
−1
0
1
2
4
a

$\mathbf{P}(\ |Z| < 2\ ) \approx 0.95$



$$\Phi(b) := \int_{-\infty}^{b} \varphi(a)\, da$$

ist die Standard-Normalverteilungsfunktion.

Es gibt für sie keinen expliziten analytischen Ausdruck

(der ohne die Formulierung als Integral bzw. Stammfunktion auskommt).

Der R-Befehl dafür ist `pnorm(b)` .

3. Die $N(\mu, \sigma^2)$-Verteilung

Sei Z standard-normalverteilt, $\mu \in \mathbb{R}$, $\sigma > 0$. Dann gilt für

$$X := \sigma Z + \mu \ :$$

$$\mathbf{E}[X] = \mu, \quad \mathbf{Var}\,[X] = \sigma^2,$$

und die Dichte von X ist $\varphi_{\mu,\sigma^2}(a)\,da$ mit

$$\varphi_{\mu,\sigma^2}(a) := \frac{1}{\sigma}\varphi\left(\frac{a-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(a-\mu)^2}{2\sigma^2}}.$$

Eine Zufallsvariable mit Dichte $\varphi_{\mu,\sigma^2}(a)\,da$

heißt normalverteilt mit Erwartungswert μ und Varianz σ^2,

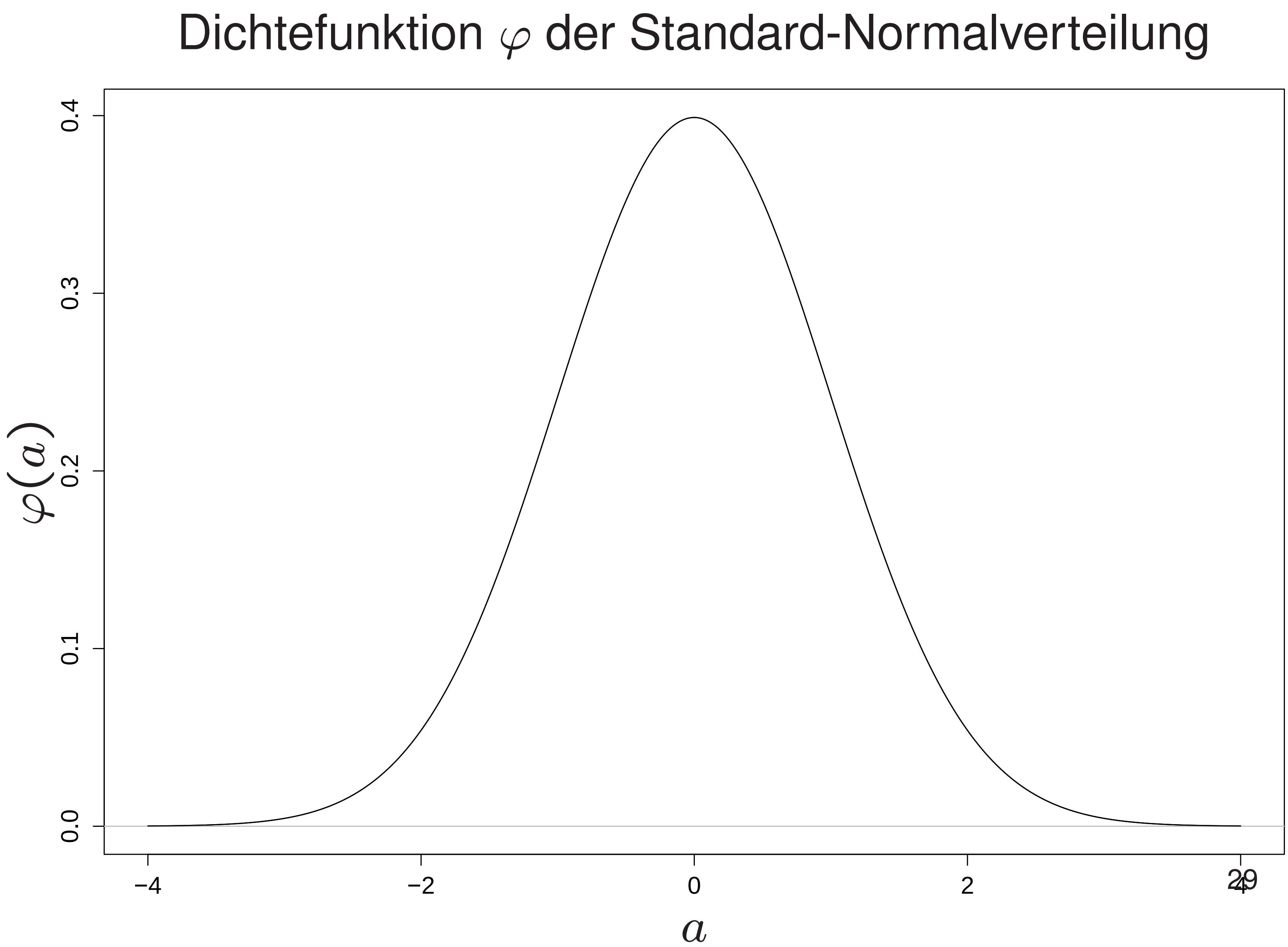
kurz

N(μ, σ^2)-verteilt.

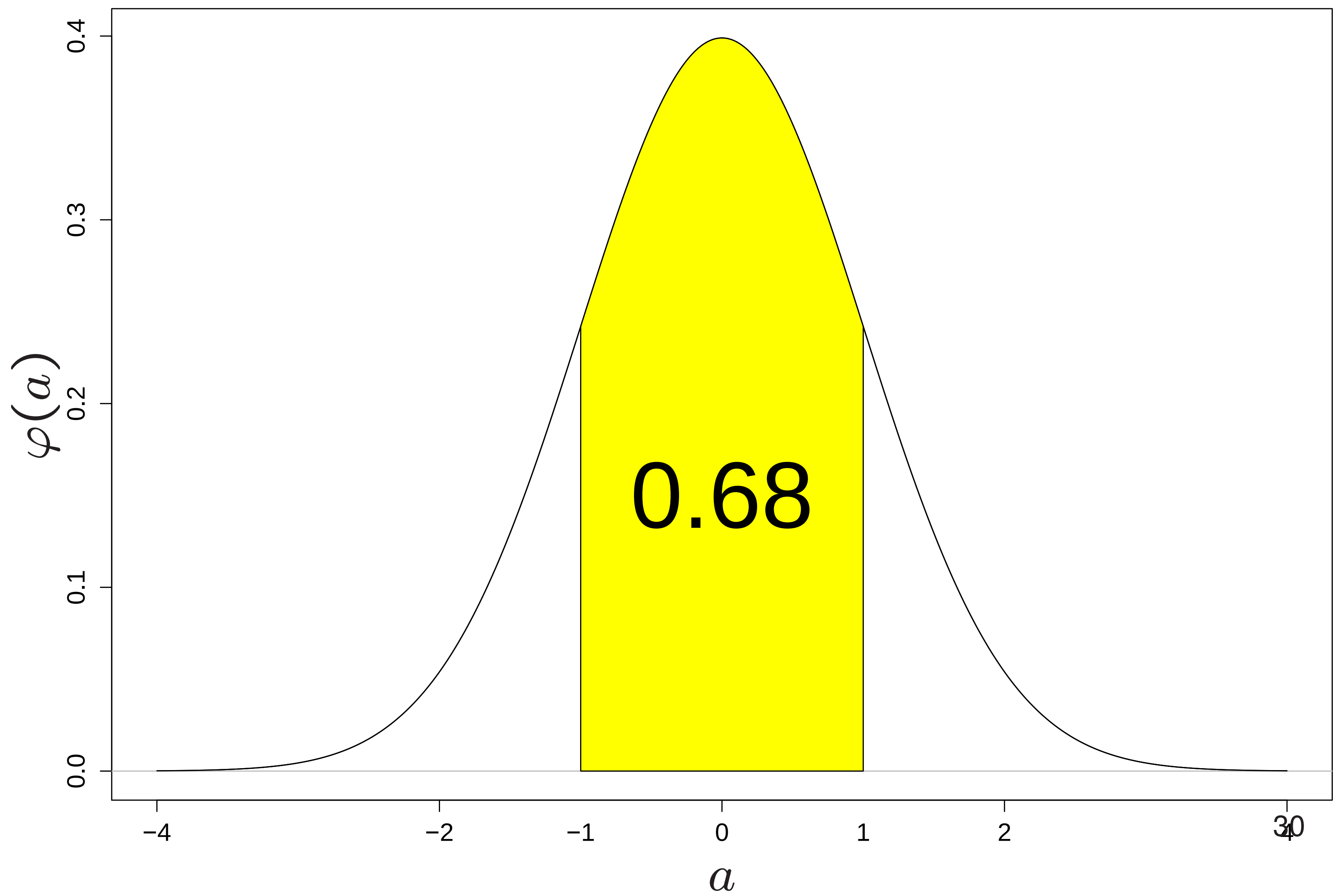
Ist Y N(μ, σ^2)-verteilt,

dann ist $\dfrac{Y-\mu}{\sigma}$ standard-normalverteilt.

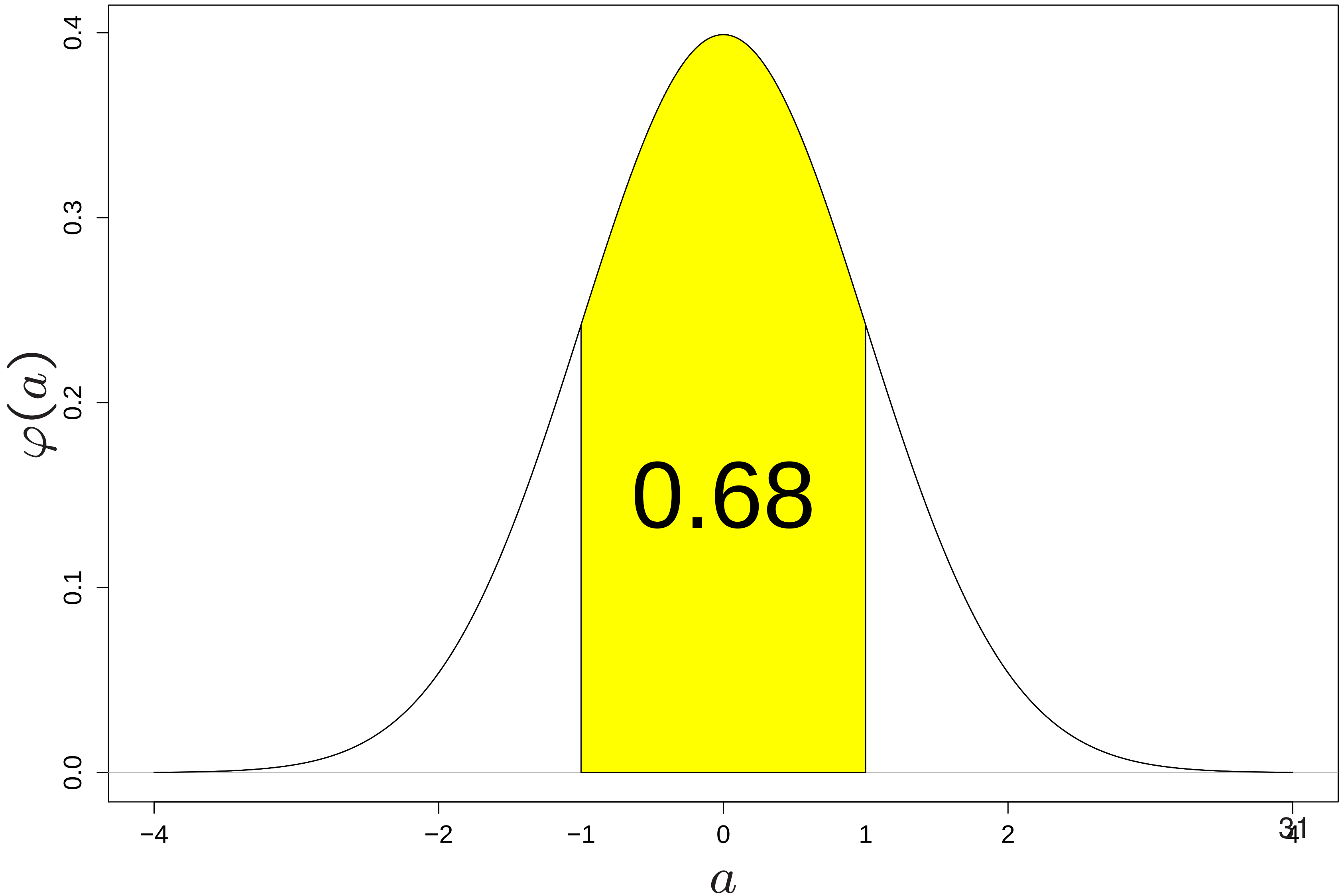
Carl Friedrich Gauss (1777-1855)


Dichtefunktion φ der Standard-Normalverteilung
$\varphi(a)$
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
a

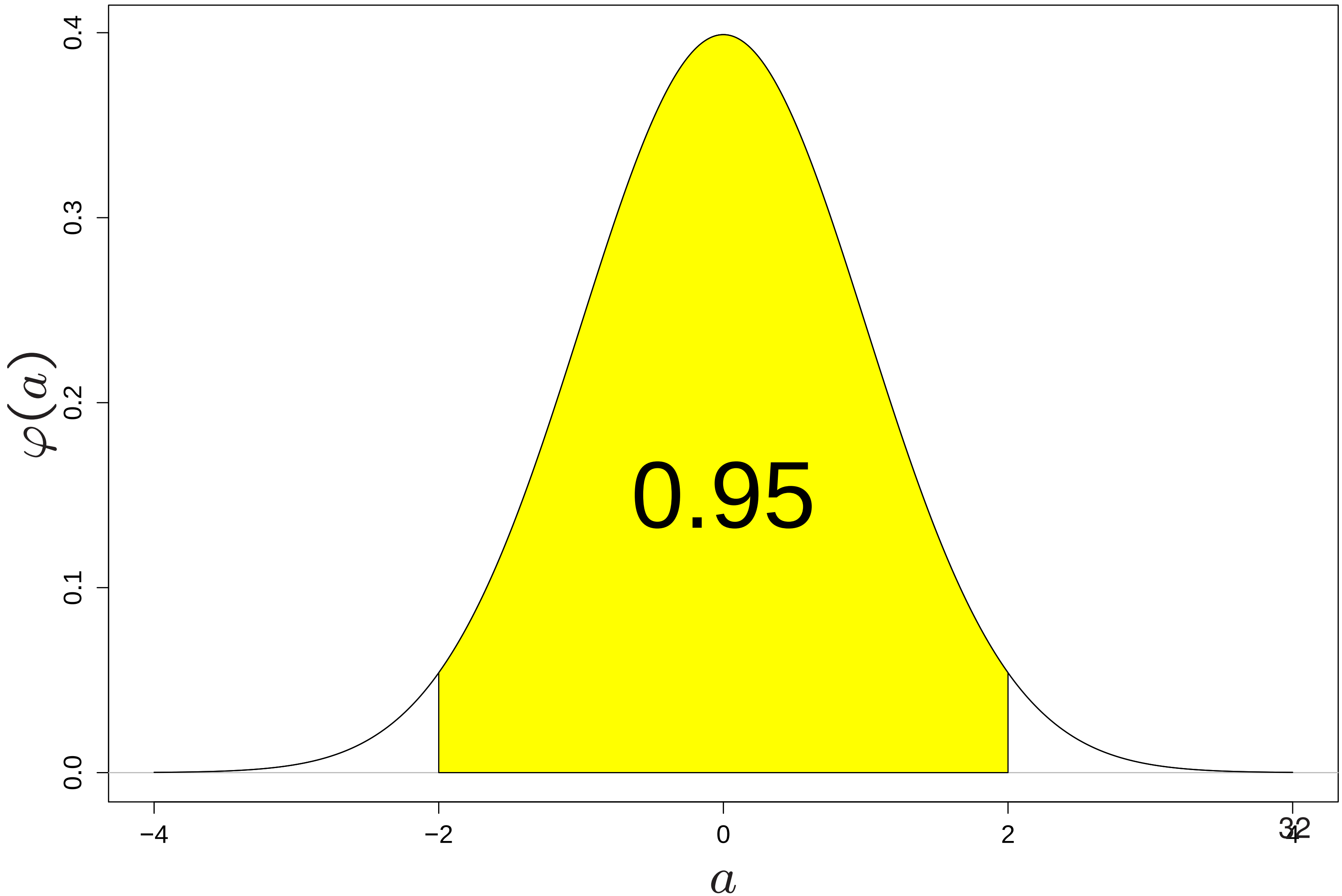
$\mathbf{P}(\,|Z| < 1) \approx 0.68$



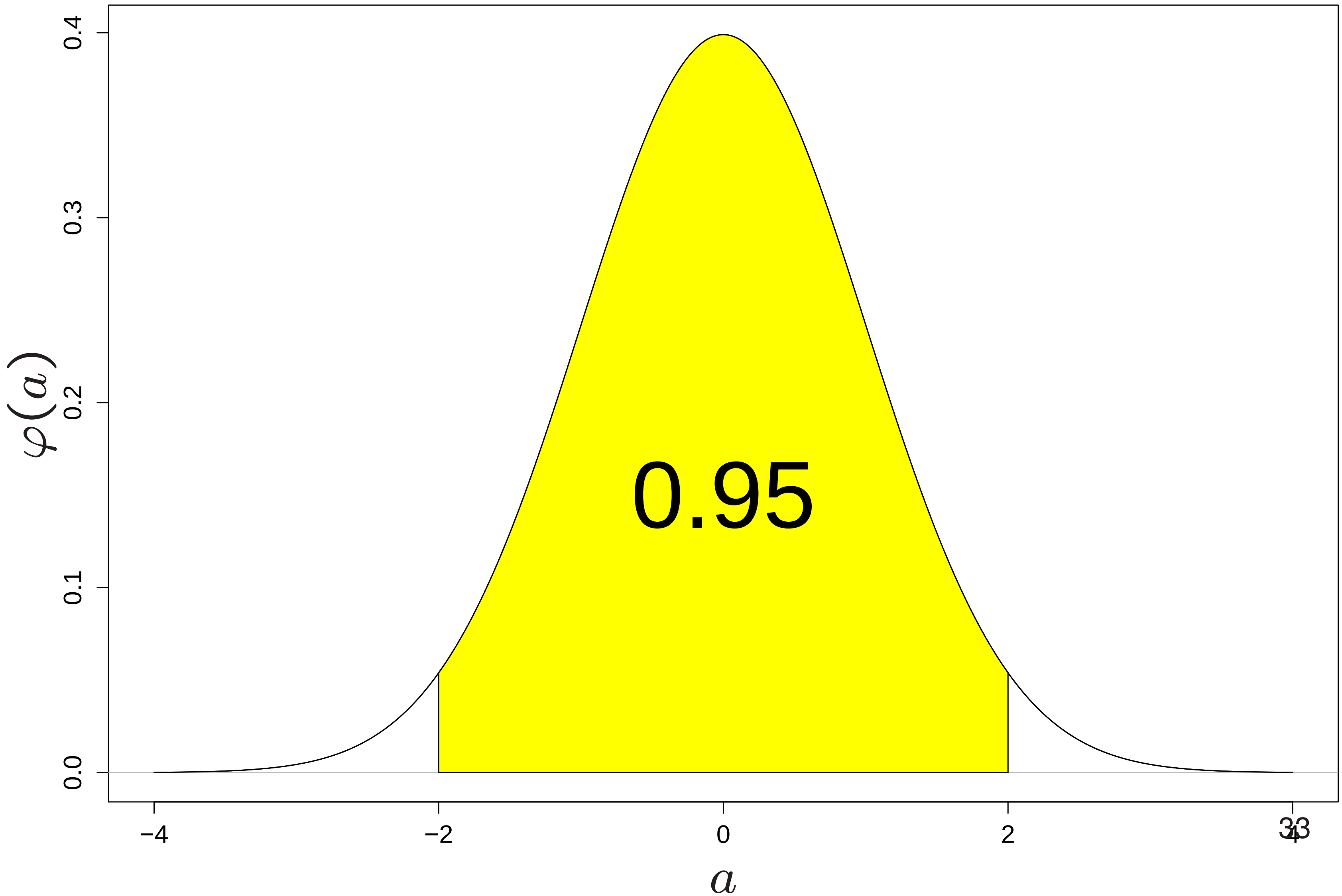
$\Phi(1) - \Phi(-1) \approx 0.68$



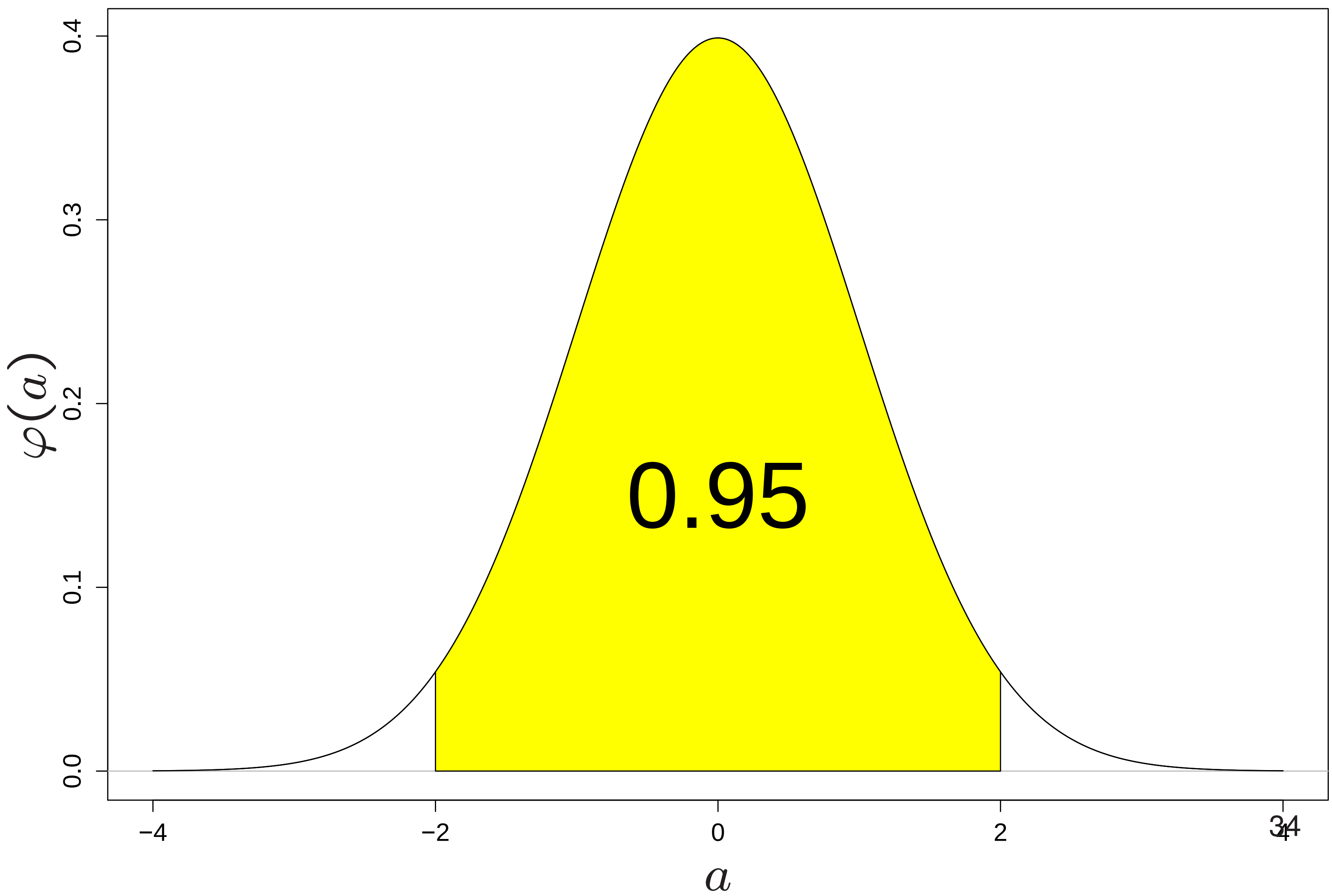
$\mathbf{P}(\ |Z| < 2\ ) \approx 0.95$



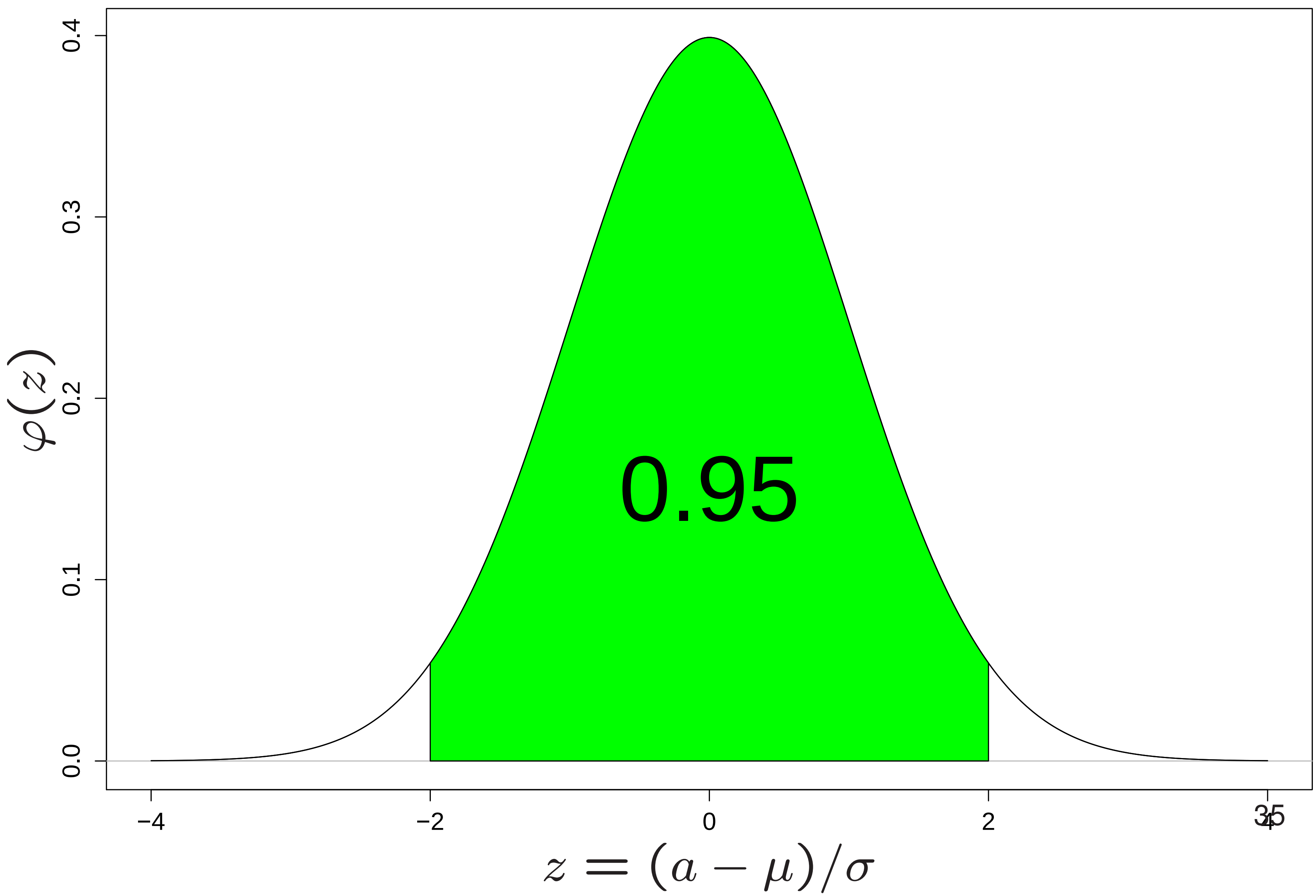
Und für N(μ, σ)-verteilte Zufallsgrößen X?

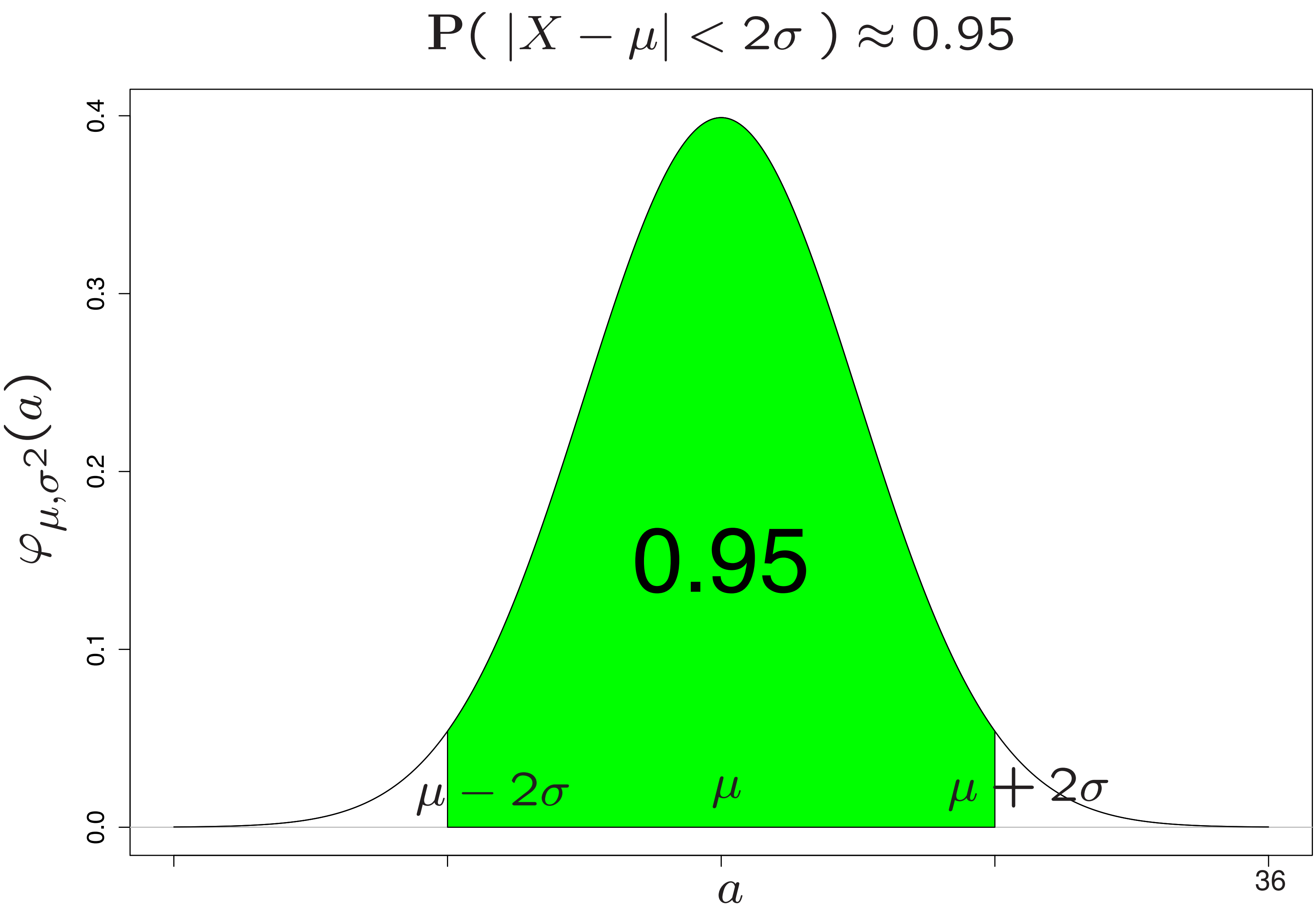


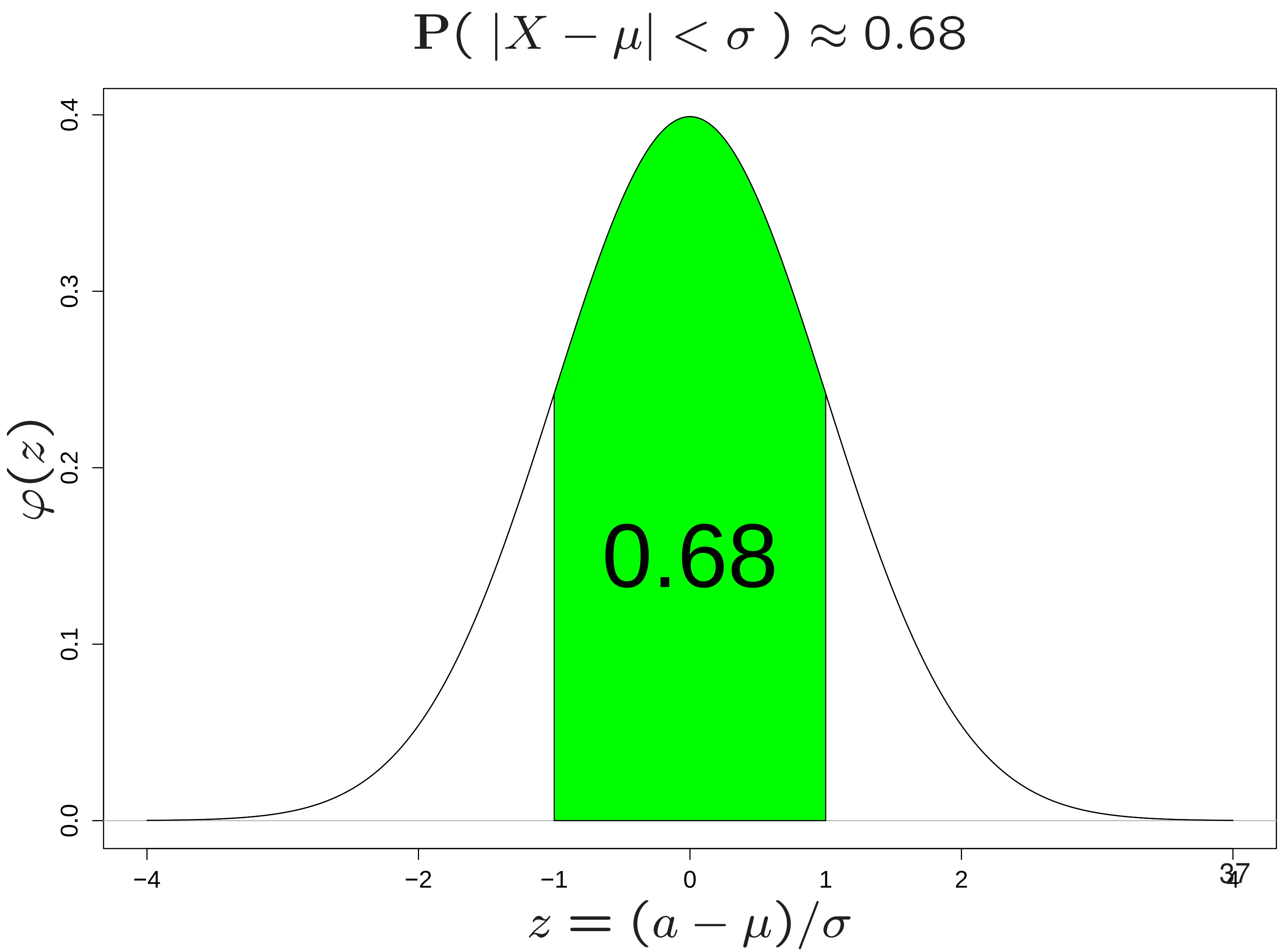
Dasselbe in grün.

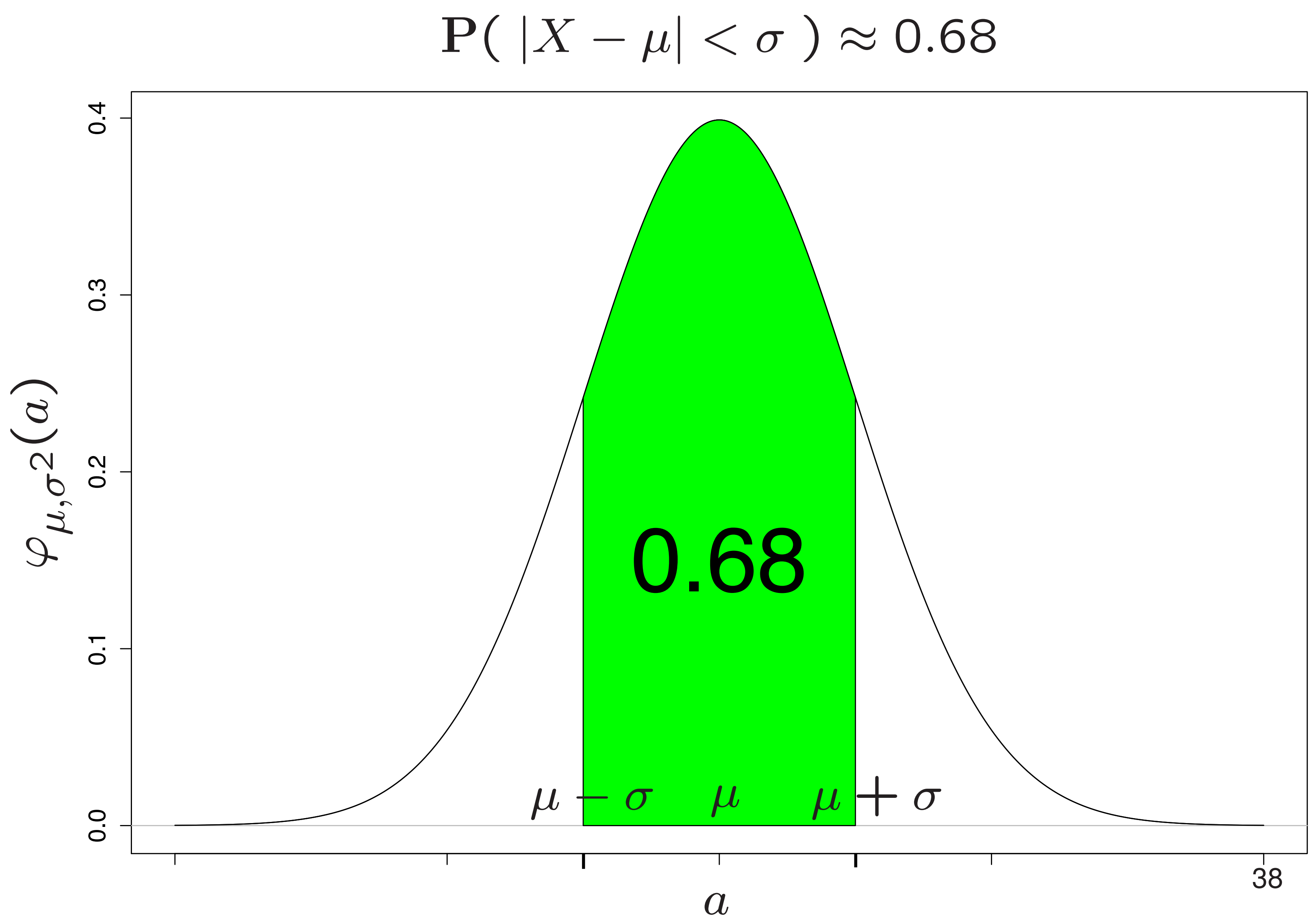


$\mathbf{P}(\ |X - \mu| < 2\sigma\ ) \approx 0.95$




$\mathbf{P}(\ |X - \mu| < 2\sigma\ ) \approx 0.95$
$\varphi_{\mu,\sigma^2}(a)$
0.4
0.3
0.2
0.1
0.0
0.95
$\mu - 2\sigma$
μ
$\mu + 2\sigma$
a


$\mathbf{P}(\ |X-\mu|<\sigma\ )\approx 0.68$
0.68
$\varphi(z)$
0.0
0.1
0.2
0.3
0.4
−4
−2
−1
0
1
2
4
$z=(a-\mu)/\sigma$


$\mathbf{P}(\ |X-\mu|<\sigma\ )\approx 0.68$
0.4
0.3
0.2
0.1
0.0
$\varphi_{\mu,\sigma^2}(a)$
0.68
$\mu-\sigma$
μ
$\mu+\sigma$
38
a

Approximation der Gewichte der Binomialverteilung

durch die Dichtefunktion der Normalverteilung

für große $\mu := np$ und $\sigma := \sqrt{npq}$:

Sei X_n Binomial(n, p)-verteilt. Dann gilt (siehe Folie 44):

$$\mathbf{P}(X_n = k) \approx \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right) = \varphi_{\mu,\sigma^2}(k) \approx \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \varphi_{\mu,\sigma^2}(a)da$$

$$\approx \mathbf{P}(k - \tfrac{1}{2} \le X \le k + \tfrac{1}{2})$$

für eine N(μ, σ^2)-verteilte ZV'e X.

4. Zufällige Paare und ihre Verteilung

Im Diskreten:

Verteilungsgewichte und ihre Summen über $A \subset S_1 \times S_2$:

Man darf die Summationsreihenfolge vertauschen.

$$\sum_{(a_1,a_2)} \mathbf{1}_A(a_1,a_2)\rho(a_1,a_2)$$

$$= \sum_{a_1} \left(\sum_{a_2} \mathbf{1}_A(a_1,a_2)\rho(a_1,a_2) \right)$$

$$= \sum_{a_2} \left(\sum_{a_1} \mathbf{1}_A(a_1,a_2)\rho(a_1,a_2) \right)$$

Im Kontinuierlichen:

Dichten und ihre Integrale über $A \subset \mathbb{R}^2$:

Man darf die Integrationsreihenfolge vertauschen.

$$\int\limits_{\mathbb{R}^2} \mathbf{1}_A(a_1, a_2) f(a_1, a_2)\, d(a_1, a_2)$$

$$= \int\limits_{\mathbb{R}} \left(\int\limits_{\mathbb{R}} \mathbf{1}_A(a_1, a_2) f(a_1, a_2)\, da_2 \right) da_1$$

$$= \int\limits_{\mathbb{R}} \left(\int\limits_{\mathbb{R}} \mathbf{1}_A(a_1, a_2) f(a_1, a_2)\, da_1 \right) da_2$$

Im Kontinuierlichen:

Dichten und ihre (iterierten) Integrale:

$$\int\limits_{\mathbb{R}^2} \mathbf{1}_A(a_1, a_2) f(a_1, a_2)\, d(a_1, a_2)$$

$$= \int\limits_{\mathbb{R}} \left(\int\limits_{\mathbb{R}} \mathbf{1}_A(a_1, a_2) f(a_1, a_2)\, da_2 \right) da_1$$

$$= \int\limits_{\mathbb{R}} \left(\int\limits_{\mathbb{R}} \mathbf{1}_A(a_1, a_2) f(a_1, a_2)\, da_1 \right) da_2$$

Im Kontinuierlichen:

Dichten und ihre (iterierten) Integrale:

$$\int_{\mathbb{R}^2} \mathbf{1}_A(a_1,a_2) f(a_1,a_2)\, d(a_1,a_2)$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(a_1,a_2) f(a_1,a_2)\, da_2 \right) da_1$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(a_1,a_2) f(a_1,a_2)\, da_1 \right) da_2$$

$$f(a_1,a_2)\, d(a_1,a_2)$$

$$= f(a_1,a_2)\, da_2\, da_1 = f(a_1,a_2)\, da_1\, da_2$$

Was Indikatorfunktionen $\mathbf{1}_A$ recht ist,

ist auch Abbildungen $h : \mathbb{R}^2 \to \mathbb{R}$ billig:

Für ein zufälliges Paar (X_1, X_2)

mit Dichte $f(a_1, a_2) d(a_1, a_2)$ auf $\mathbb{R}^2$

gilt die **Transformationsformel**

$$\begin{aligned}
\mathbf{E}[h(X_1, X_2)] &= \int_{\mathbb{R}^2} h(a_1, a_2) f(a_1, a_2)\, d(a_1, a_2) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(a_1, a_2) f(a_1, a_2)\, da_2 \right) da_1 \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(a_1, a_2) f(a_1, a_2)\, da_1 \right) da_2 \ .
\end{aligned}$$

5. Unabhängigkeit von Zufallsvariablen mit Dichten

Für diskrete Zufallsvariable
war die Unabhängigkeit von X_1 und X_2
äquivalent zur Produktform der gemeinsamen Gewichte:

$$\rho(a_1, a_2) = \rho_1(a_1)\,\rho_2(a_2).$$

Für Zufallsvariable mit Dichten
ist die Unabhängigkeit von X_1 und X_2
äquivalent zur Produktform der gemeinsamen Dichte:

$$f(a_1, a_2)\,da_1\,da_2 = f_1(a_1)\,da_1\;\; f_2(a_2)\,da_2$$

Allgemeiner gilt der

Satz über die Unabhängigkeit von Zv'en mit Dichten

$X_1, \ldots, X_n$ seien reellwertige Zufallsvariable,

$f_1, \ldots, f_n$ seien Dichtefunktionen.

Dann sind äquivalent:

(i) $X_1, \ldots, X_n$ sind unabhängig,

und X_i hat die Dichte $f_i(a_i)\, da_i, \;\; i = 1, \ldots, n$.

(ii) $(X_1, \ldots, X_n)$ hat die Dichte

$$f_1(a_1) \cdots f_n(a_n)\, da_1 \ldots da_n$$

6. Die uniforme Verteilung auf dem Einheitsquadrat

X_1, X_2 seien unabhängig und uniform verteilt auf $[0,1]$.

Dann hat (X_1, X_2) die Dichte

$$\mathbf{1}_{[0,1]}(a_1)\, da_1 \;\cdot \mathbf{1}_{[0,1]}(a_2)\, da_2$$

$$= \mathbf{1}_{[0,1]\times[0,1]}(a_1, a_2)\; da_1\, da_2,$$

und ist somit uniform verteilt auf $[0,1] \times [0,1]$.

7. Die Standardnormalverteilung auf $\mathbb{R}^2$

(vgl. Buch S. 71)

Zur Erinnerung:

Eine $\mathbb{R}$-wertige Zufallsvariable Z mit Dichte

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

heißt standard-normalverteilt (auf $\mathbb{R}^1$).

Wichtige Beobachtung:

Z_1, Z_2 seien standard-normalverteilt und unabhängig.

(Z_1, Z_2) hat dann die Dichte

$$\varphi(a_1)\, da_1\; \varphi(a_2)\, da_2$$

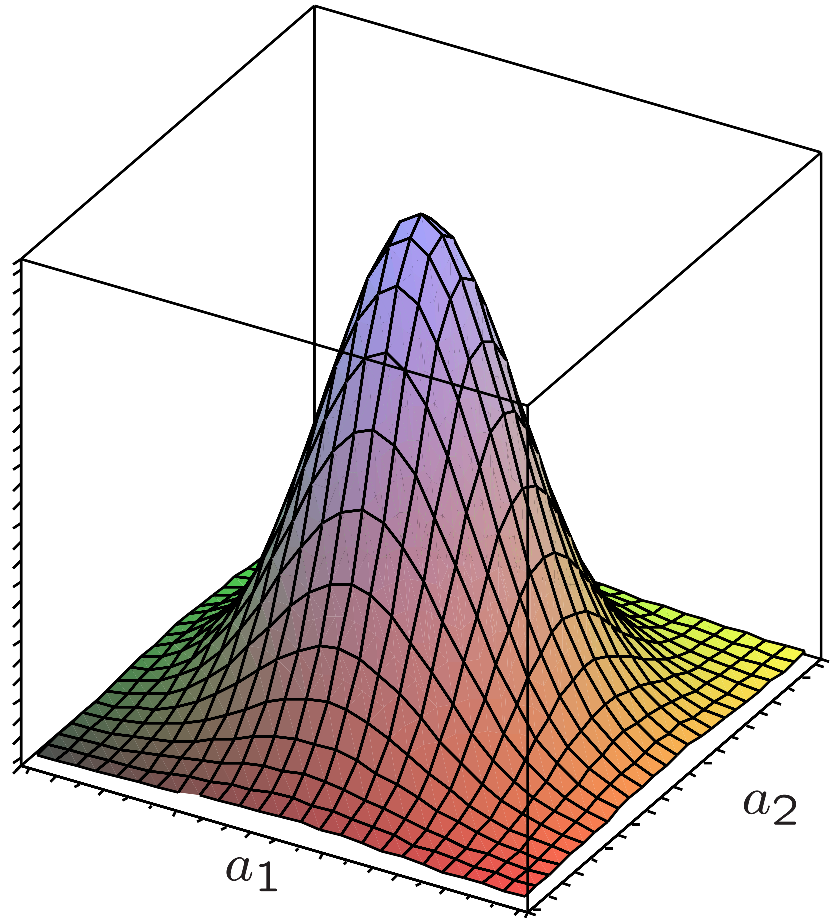
$$= \frac{1}{\sqrt{2\pi}} e^{-a_1^2/2} \frac{1}{\sqrt{2\pi}} e^{-a_2^2/2}\, da_1 da_2$$

$$= \frac{1}{2\pi} e^{-|a|^2/2}\, da, \qquad a = (a_1, a_2) \in \mathbb{R}^2,\ |a|^2 := a_1^2 + a_2^2.$$

Die Dichte ist rotationssymmetrtisch!

$$f(a_1, a_2) =$$

$$\frac{1}{2\pi} e^{-(a_1^2 + a_2^2)/2}$$



Definition:

Eine $\mathbb{R}^2$-wertige Zufallsvariable Z mit Dichte

$$f(a)\,da = \frac{1}{2\pi} e^{-|a|^2/2}\,da, \quad a \in \mathbb{R}^2,$$

heißt **standard-normalverteilt auf** $\mathbb{R}^2$.

8. Eine Folgerung aus der Rotationssymmetrie

(vgl Buch S. 71)

Z_1 und Z_2 seien unabhängig und standard-normalverteilt.

Wir wollen einsehen, dass dann auch

$$\frac{Z_1 + Z_2}{\sqrt{2}}$$

standard-normalverteilt ist.

Dafür gibt es ein wunderschönes geometrisches Argument.

Sei $\vec{e}_1$, $\vec{e}_2$ die Standardbasis in $\mathbb{R}^2$.

Wir fassen das zufällige Zahlenpaar $Z = (Z_1, Z_2)$ auf

als die Koordinaten des zufälligen Vektors

$$\vec{Z} = Z_1 \vec{e}_1 + Z_2 \, \vec{e}_2$$

und stellen fest:

- Z_1 ist die Koordinate von $\vec{Z}$ bgl. des Einheitsvektors $\vec{e}_1$.
- $\dfrac{Z_1 + Z_2}{\sqrt{2}}$ ist die Koordinate von $\vec{Z}$ bgl. des Einheitsvektors

$$\vec{u}_{\text{diag}} := \frac{1}{\sqrt{2}}\vec{e}_1 + \frac{1}{\sqrt{2}}\vec{e}_2.$$

- Die Verteilung von $\vec{Z}$ ist rotationssymmetrisch

(siehe voriger Abschnitt)

Also ist Z_1 so verteilt wie $\frac{Z_1+Z_2}{\sqrt{2}}$.

Und was dem Einheitsvektor $\vec{u}_{\text{diag}}$ recht ist,

ist jedem Einheitsvektor $\vec{u} = \tau_1\vec{e}_1 + \tau_2\vec{e}_2$ billig !

Anders gesagt:

Sind Z_1, Z_2 unabhängig und N(0, 1)-verteilt,

dann gilt für jedes Zahlenpaar (τ_1, τ_2) mit $\tau_1^2 + \tau_2^2 = 1$:

$Y := \tau_1 Z_1 + \tau_2 Z_2$ ist N(0, 1)-verteilt.

(Denn Y ist die Koordinate von $\vec{Z} = Z_1 \vec{e}_1 + Z_2 \vec{e}_2$

zum Einheitsvektor $\vec{u} := \tau_1 \vec{e}_1 + \tau_2 \vec{e}_2$.)

Eine wichtige Folgerung hieraus:

Die Summe von

unabhängigen, normalverteilten Zufallsvariablen

ist wieder normalverteilt.

Denn:

$$(\sigma_1 Z_1 + \mu_1) + (\sigma_2 Z_2 + \mu_2) = (\sigma_1 Z_1 + \sigma_2 Z_2) + (\mu_1 + \mu_2),$$

und

$$\sigma_1 Z_1 + \sigma_2 Z_2 = \sqrt{\sigma_1^2 + \sigma_2^2} \left(\frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} Z_1 + \frac{\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} Z_2 \right)$$

ist N$(0, \sigma_1^2 + \sigma_2^2)$-verteilt.

Eine wichtige Folgerung hieraus:

Die Summe von

unabhängigen, normalverteilten Zufallsvariablen

ist wieder normalverteilt.

Denn:

$$(\sigma_1 Z_1 + \mu_1) + (\sigma_2 Z_2 + \mu_2) = (\sigma_1 Z_1 + \sigma_2 Z_2) + (\mu_1 + \mu_2),$$

und

$$\sigma_1 Z_1 + \sigma_2 Z_2 = \sqrt{\sigma_1^2 + \sigma_2^2} \left(\frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} Z_1 + \frac{\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} Z_2 \right)$$

ist N$(0, \sigma_1^2 + \sigma_2^2)$-verteilt.

9. Die Standardnormalverteilung auf $\mathbb{R}^n$

(vgl. Buch S. 71)

In Abschnitt 5 der heutigen Vorlesung formulierten wir den

Satz über die Unabhängigkeit von Zv'en mit Dichten

$X_1, \ldots, X_n$ seien reellwertige Zufallsvariable,

$f_1, \ldots, f_n$ seien Dichtefunktionen.

Dann sind äquivalent:

(i) $X_1, \ldots, X_n$ sind unabhängig,

und X_i hat die Dichte $f_i(a_i)\, da_i$, $\; i = 1, \ldots, n$.

(ii) $(X_1, \ldots, X_n)$ hat die Dichte

$$f_1(a_1) \cdots f_n(a_n)\, da_1 \ldots da_n$$

Dazu passt die folgende Situation:

Sei $Z := (Z_1, \ldots, Z_n)$. Dann gilt:

$Z_1, \ldots, Z_n$ sind unabhängig und N(0, 1)-verteilt

$$\Longleftrightarrow$$

$$\mathbf{P}(Z \in da) = \frac{1}{(2\pi)^{n/2}} \exp\Big(-\frac{|a|^2}{2}\Big)\, da\ , \quad a \in \mathbb{R}^n\ ,$$

mit $|a|^2 := a_1^2 + \cdots + a_n^2$.

Z heißt dann *standard-normalverteilt auf* $\mathbb{R}^n$.

Die Dichte der Standard-Normalverteilung auf $\mathbb{R}^n$ ist rotationssymmetrisch. Analog zum Fall $n = 2$ gilt deshalb:

Ist $Z = (Z_1, \ldots, Z_n)$ standard-normalverteilt auf $\mathbb{R}^n$ und sind $\tau_1, \ldots, \tau_n$ reelle Zahlen mit $\tau_1^2 + \cdots + \tau_n^2 = 1$, dann ist $Y := \tau_1 Z_1 + \cdots + \tau_n Z_n$ N$(0, 1)$-verteilt.

(Denn Y ist die Koordinate von $\vec{Z} = Z_1 \vec{e}_1 + \cdots + Z_n \vec{e}_n$ zum Einheitsvektor $\vec{u} := \tau_1 \vec{e}_1 + \cdots + \tau_n \vec{e}_n$.)

Insbesondere ergibt sich:

$$\frac{Z_1 + \cdots + Z_n}{\sqrt{n}} \quad \text{ist N}(0,1)\text{-verteilt.}$$