Vorlesung 6b

Zufallsvariable mit Dichten

Teil 2

Normalverteilung

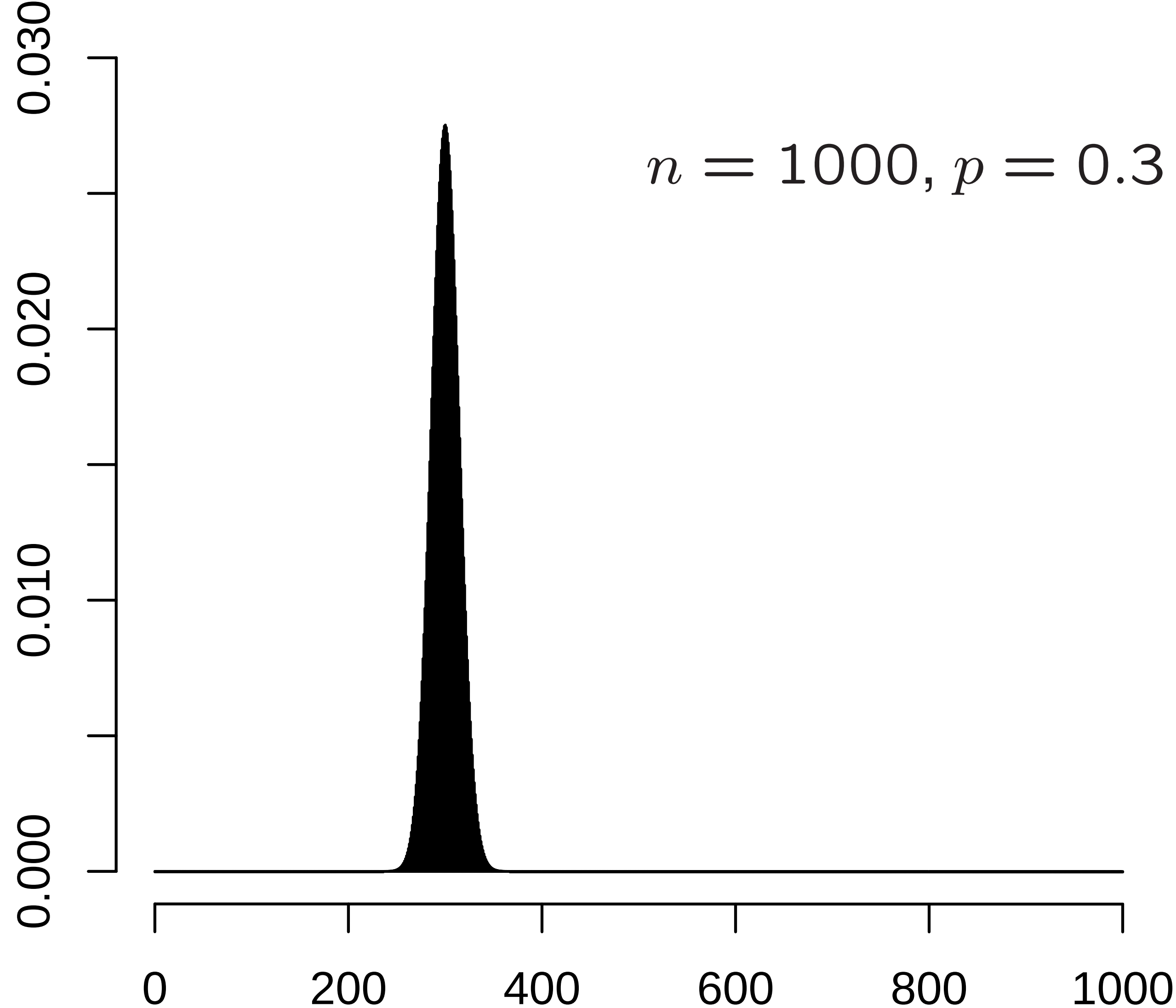
Wie sieht die Bin$(n, 1/2)$-Verteilung

für großes n aus,

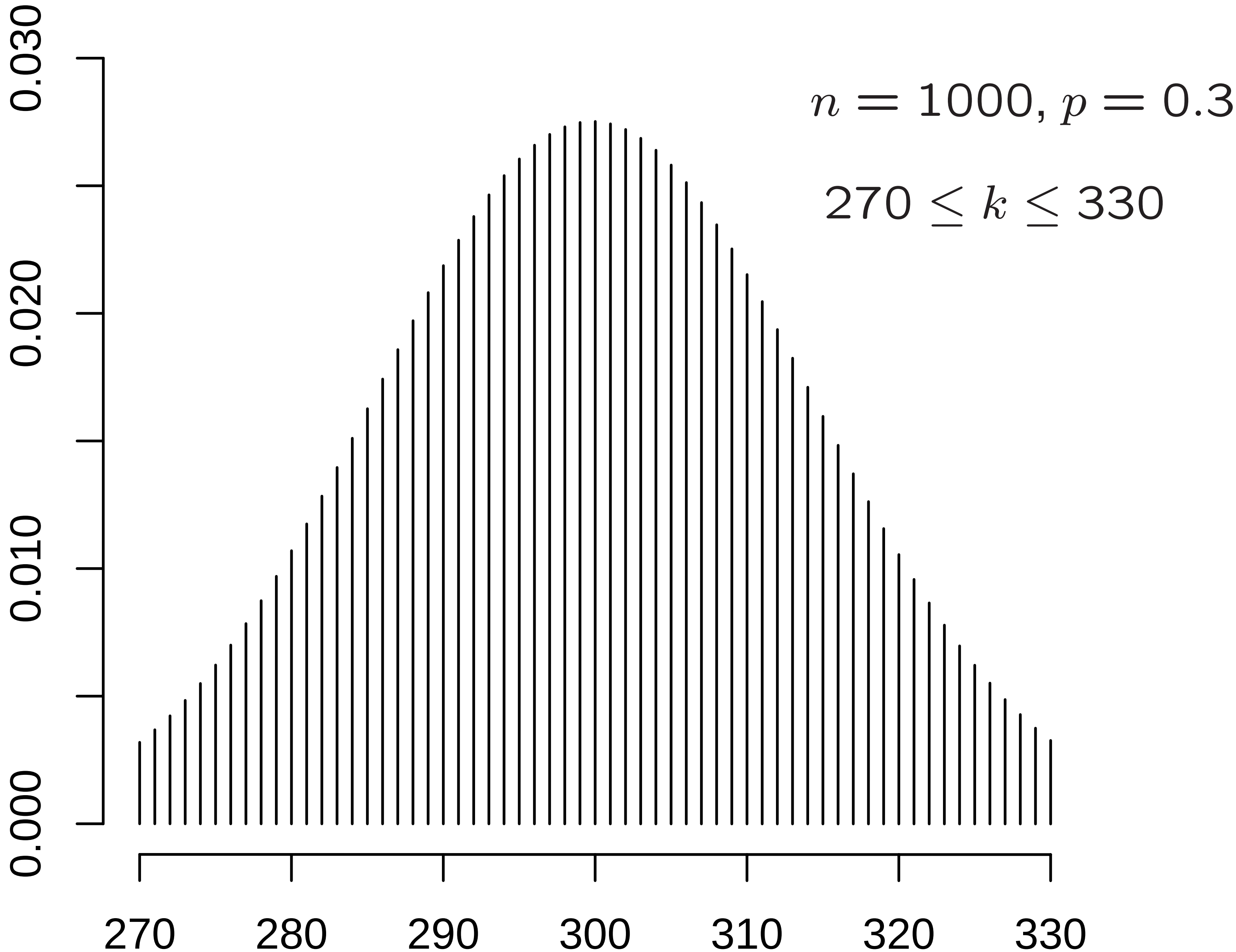
oder allgemeiner

die Bin(n, p)-Verteilung

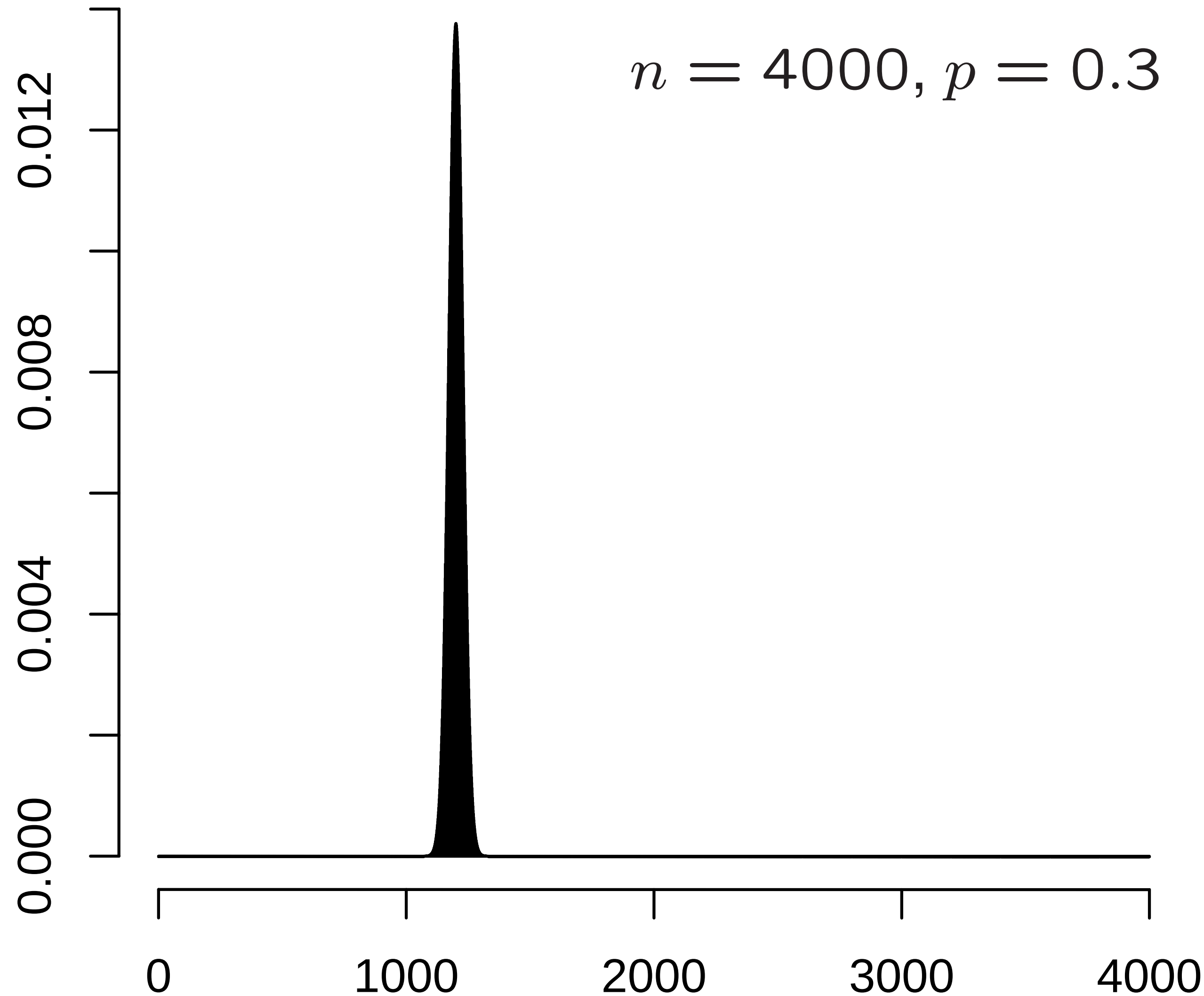
mit großem n und großem npq ?

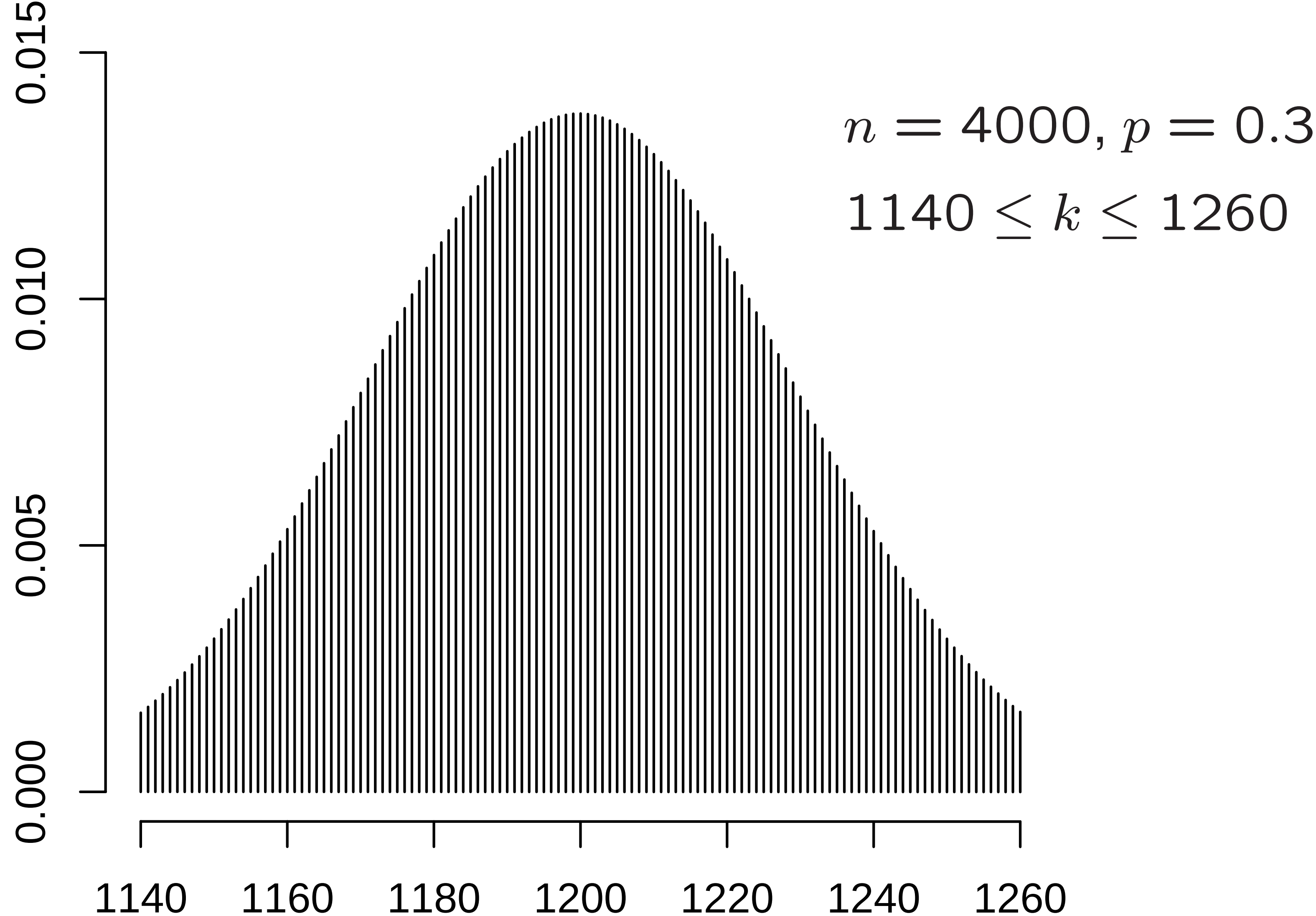


$n = 1000, p = 0.3$
0.000
0.010
0.020
0.030
0
200
400
600
800
1000



0.000
0.010
0.020
0.030
$n = 1000, p = 0.3$
$270 \leq k \leq 330$
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



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

Binomialverteilungen mit großem n und großer Varianz npq

sehen "glockenförmig" aus,

wenn man sie geeignet "ins Bild holt".

Dahinter steht eine Approximation der Binomialgewichte

mit “Stirling & Taylor”:

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

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

Dahinter steht eine Approximation der Binomialgewichte

mit “Stirling & Taylor”:

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

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

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

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

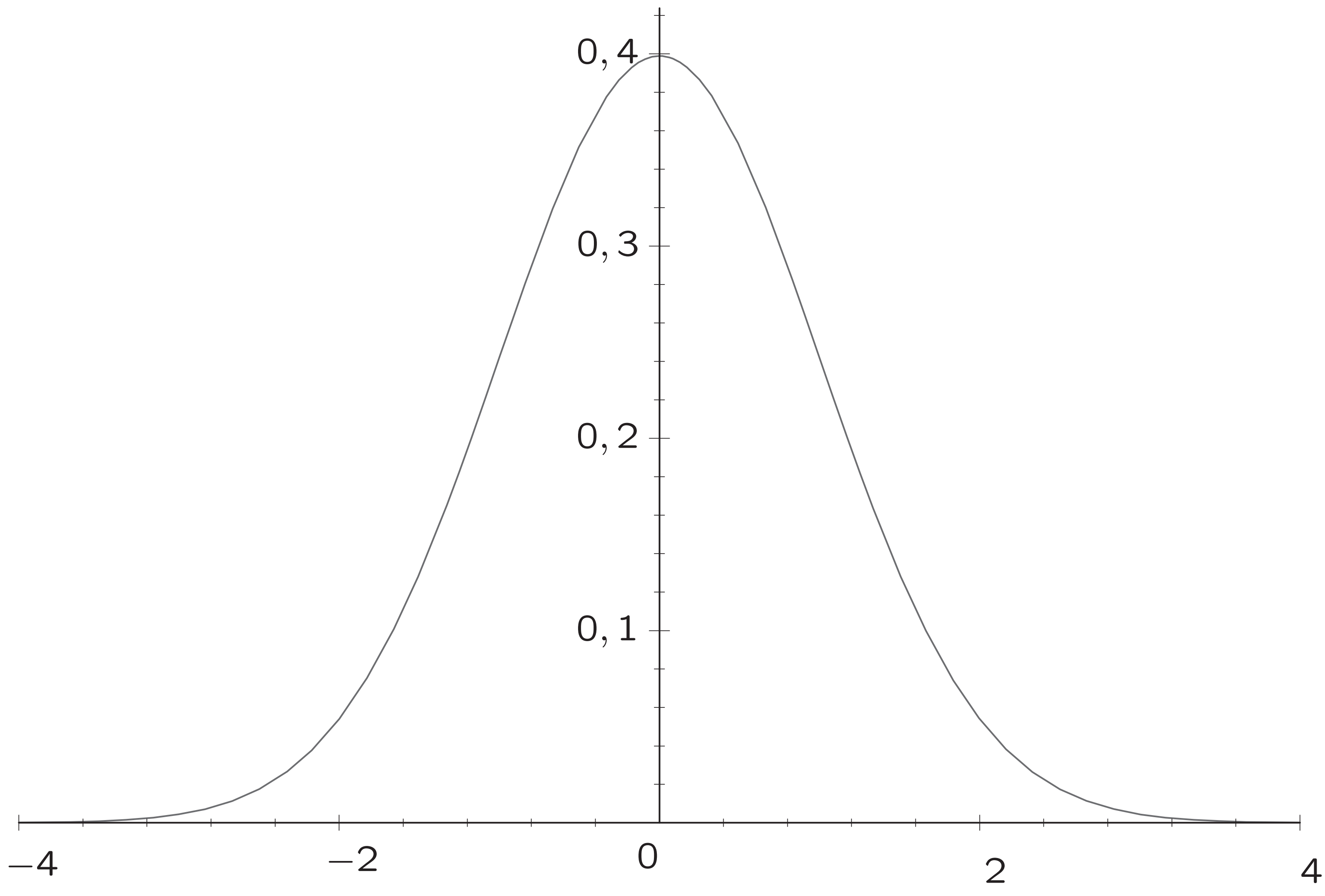
Die nächste Definition beinhaltet

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.


0, 4
0, 3
0, 2
0, 1
−4
−2
0
2
4

Nebenbei überzeugen wir uns, dass gilt

$$I := \int_{\mathbb{R}} \varphi(a)\,da = 1.$$

In der Tat ergibt der Übergang zu Polarkoordinaten:

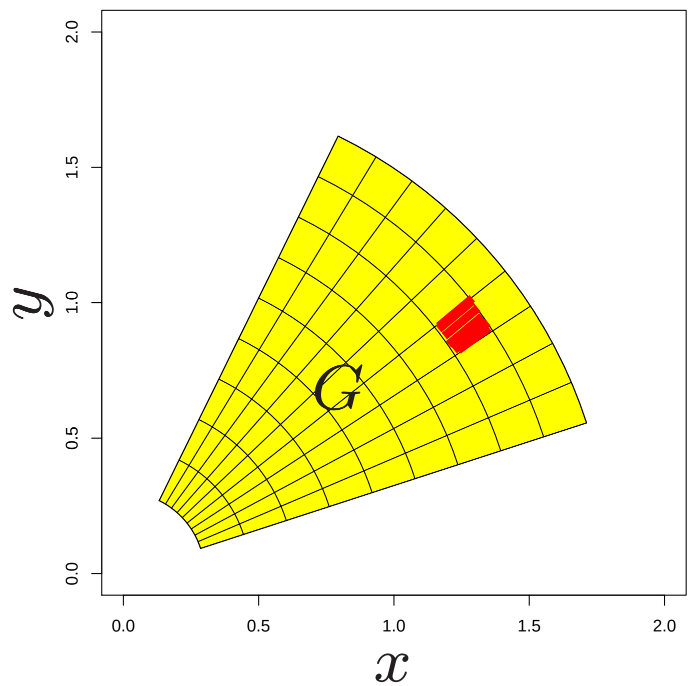
$$J^2 = \int_{\mathbb{R}} \varphi(x)\,dx \int_{\mathbb{R}} \varphi(y)\,dy = \iint\limits_{\mathbb{R}\times\mathbb{R}} \frac{1}{2\pi} e^{-(x^2+y^2)/2} dx\,dy$$

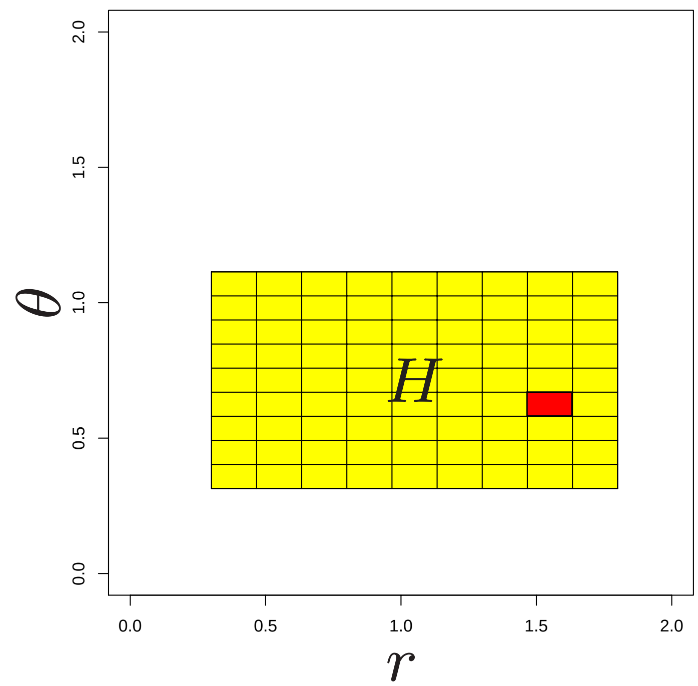
$$= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2}\; r\,dr\,d\theta$$

$$= -e^{-r^2/2}\Big|_0^{\infty} = 1. \quad \square$$

Zur Illustration der Polarkoordinatentransformation:

$$\iint\limits_G e^{-(x^2+y^2)/2}\, dx\, dy = \iint\limits_H e^{-r^2/2}\, r\, dr\, d\theta$$





Für ein standard-normalverteiltes Z ist

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

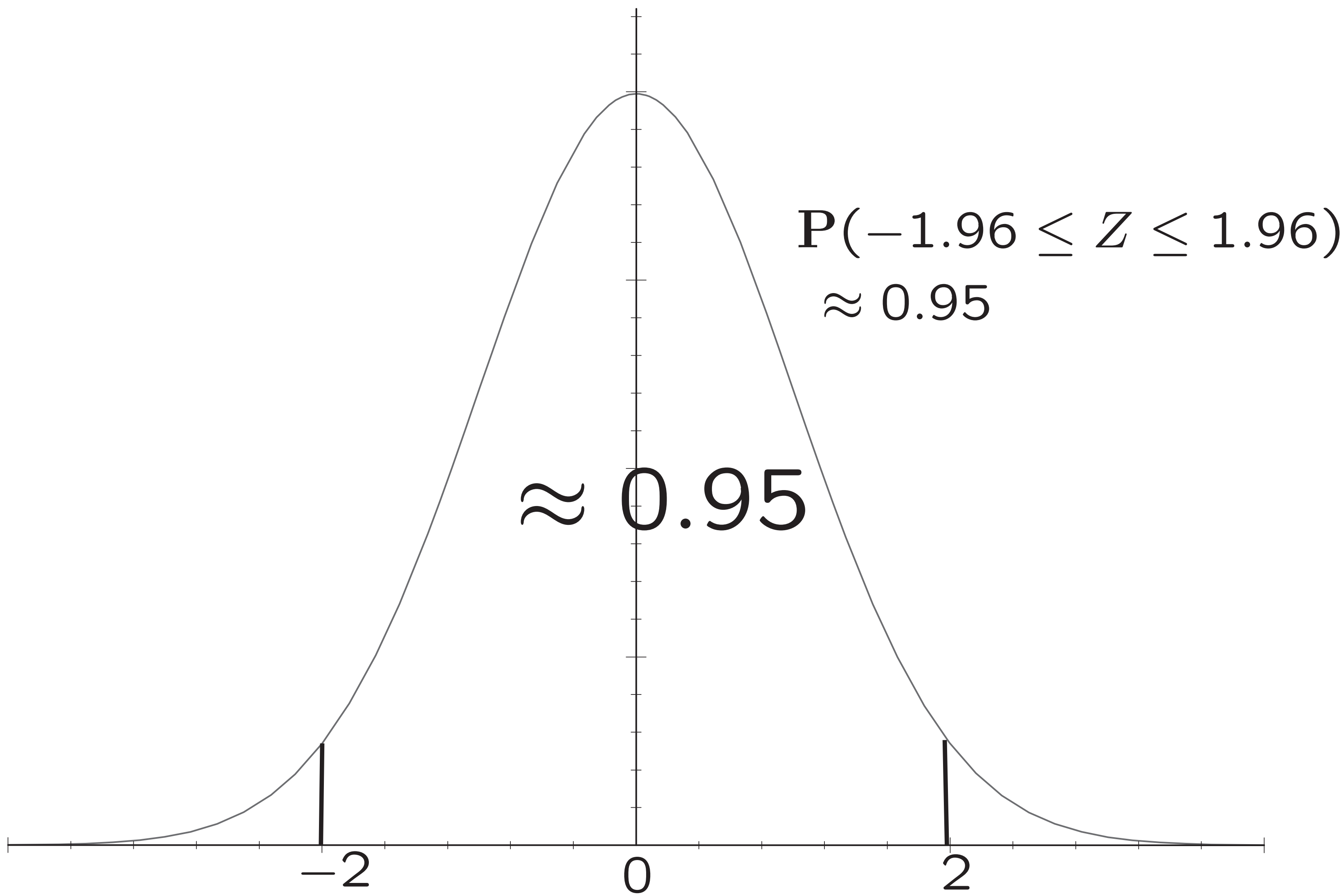
Denn aus Symmetriegründen ist $\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} a\, e^{-a^2/2}\, da = 0,$

und mit partieller Integration bekommt man

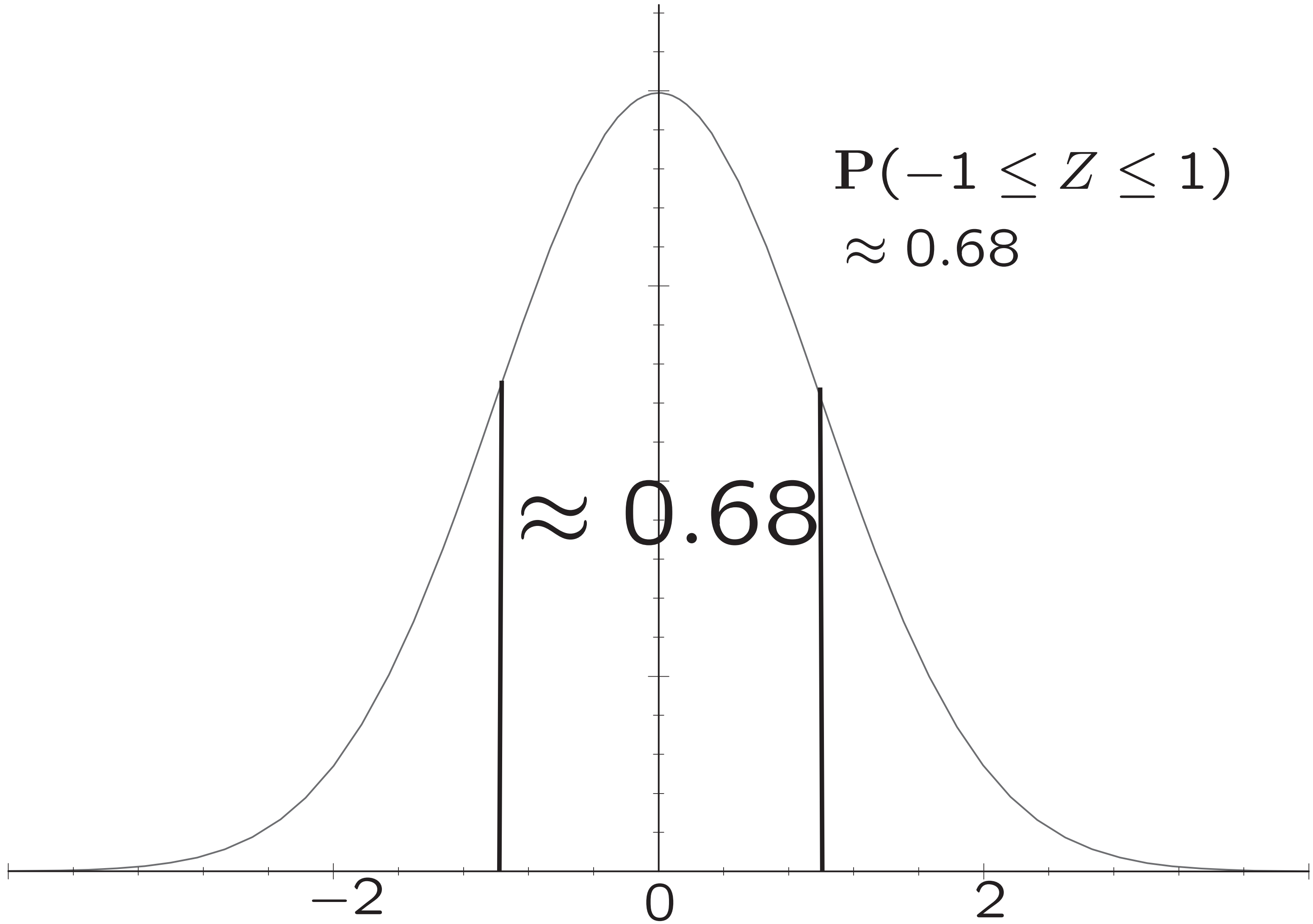
$$\int_{\mathbb{R}} a^2\, e^{-a^2/2}\, da = \int_{\mathbb{R}} a\, a e^{-a^2/2}\, da$$

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

Zwei für die Praxis wichtige Zahlen:



P(−1.96 ≤ Z ≤ 1.96)
≈ 0.95
≈ 0.95
−2
0
2


$\mathbf{P}(-1 \leq Z \leq 1)$
≈ 0.68
≈ 0.68
-2
0
2

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 Mittelwert 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.

Zurück zu unserer locker gestellten Frage:

Wie holt man bei großem n und großem npq eine

Bin(n, p)-verteilte Zufallsvariable X “zurück ins Bild?”

Durch **Standardisieren**, d.h. in diesem Fall

Verschieben um den Erwartungswert μ

und Teilen durch die Standardabweichung σ:

$\frac{X-\mu}{\sigma}$ ist dann annähernd N($0, 1$)-verteilt.

Genaues sagt der Satz von de Moivre-Laplace (Buch S. 44).

Satz (de Moivre (1733) für $p = 1/2$, Laplace (1812))

Sei $Y_1, Y_2, \ldots$ ein p-Münzwurf

und Z eine N(0, 1)-verteilte Zufallsvariable.

Dann gilt mit $n \to \infty$:

$$\mathbf{P}\left(c \le \frac{1}{\sqrt{npq}}\left(\sum_{i=1}^{n} Y_i - np\right) \le d\right) \to \mathbf{P}(c \le Z \le d)$$

für alle $-\infty \le c < d \le \infty$.

Der Wunder nicht genug:

Was dem p-Münzwurf recht ist,

- die Verteilungskonvergenz der reskalierten Summe gegen N(0,1) -

ist jeder Folge von “unabhängigen, identisch verteilten” Zufallsvariablen mit endlicher Varianz billig.

Diese Aussage wird präzisiert im klassischen Zentralen Grenzwertsatz (Buch S. 77); er verallgemeinert den Satz von de Moivre-Laplace. Mehr dazu demnächst !