Vorlesung 6b

Die Normalverteilung

Teil 1: Von den Binomialgewichten zur Gauß'schen Glockenkurve

Erinnerung und Auftakt:

Von der geometrischen zur Exponentialverteilung

Betrachten wir eine Geom(p)-verteilte Zufallsvariable Y mit

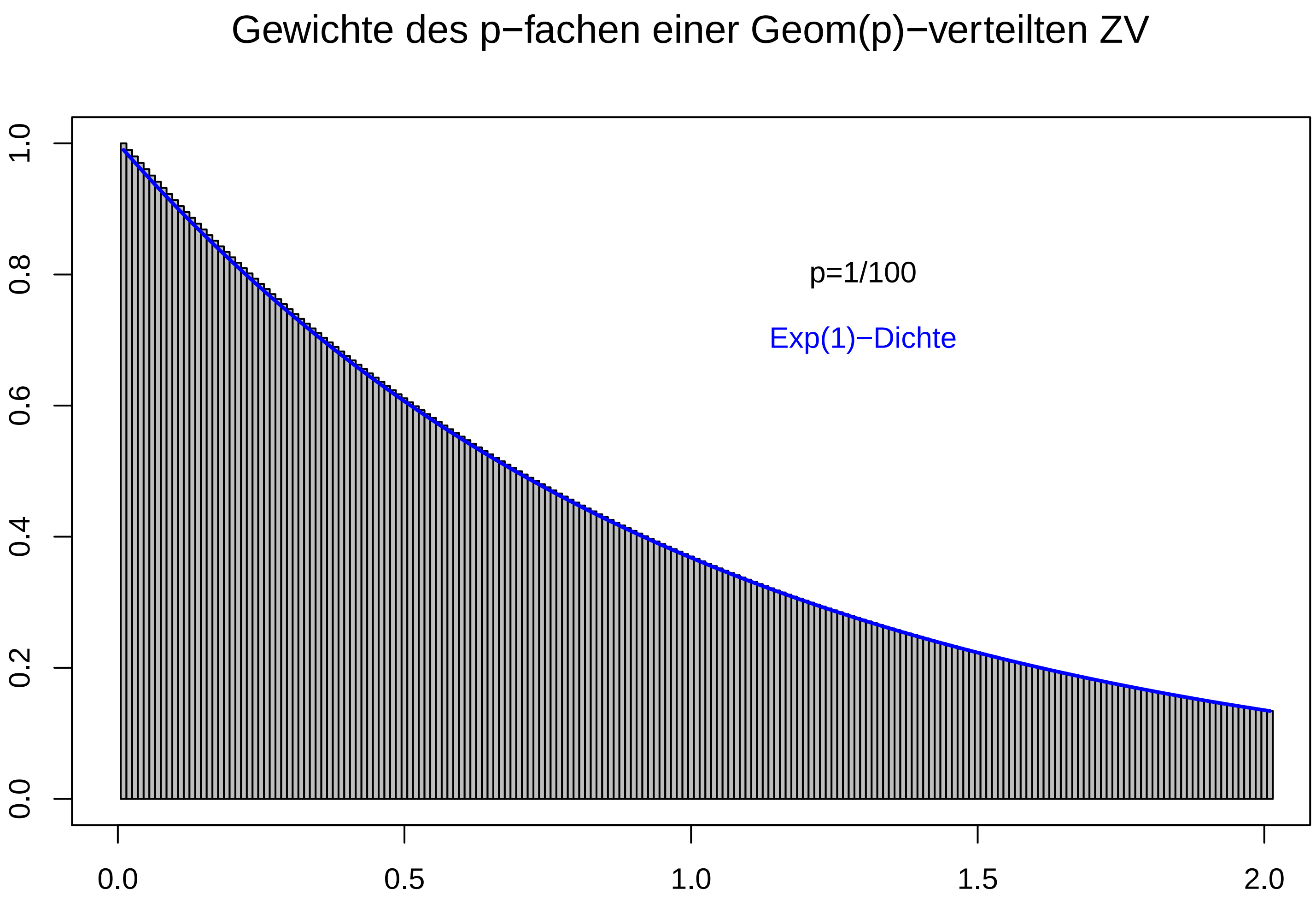
$$p = \frac{1}{100}.$$

Es gilt: $\mathbf{E}[Y] = 100$, also hat $\frac{Y}{100}$ Erwartungswert 1.

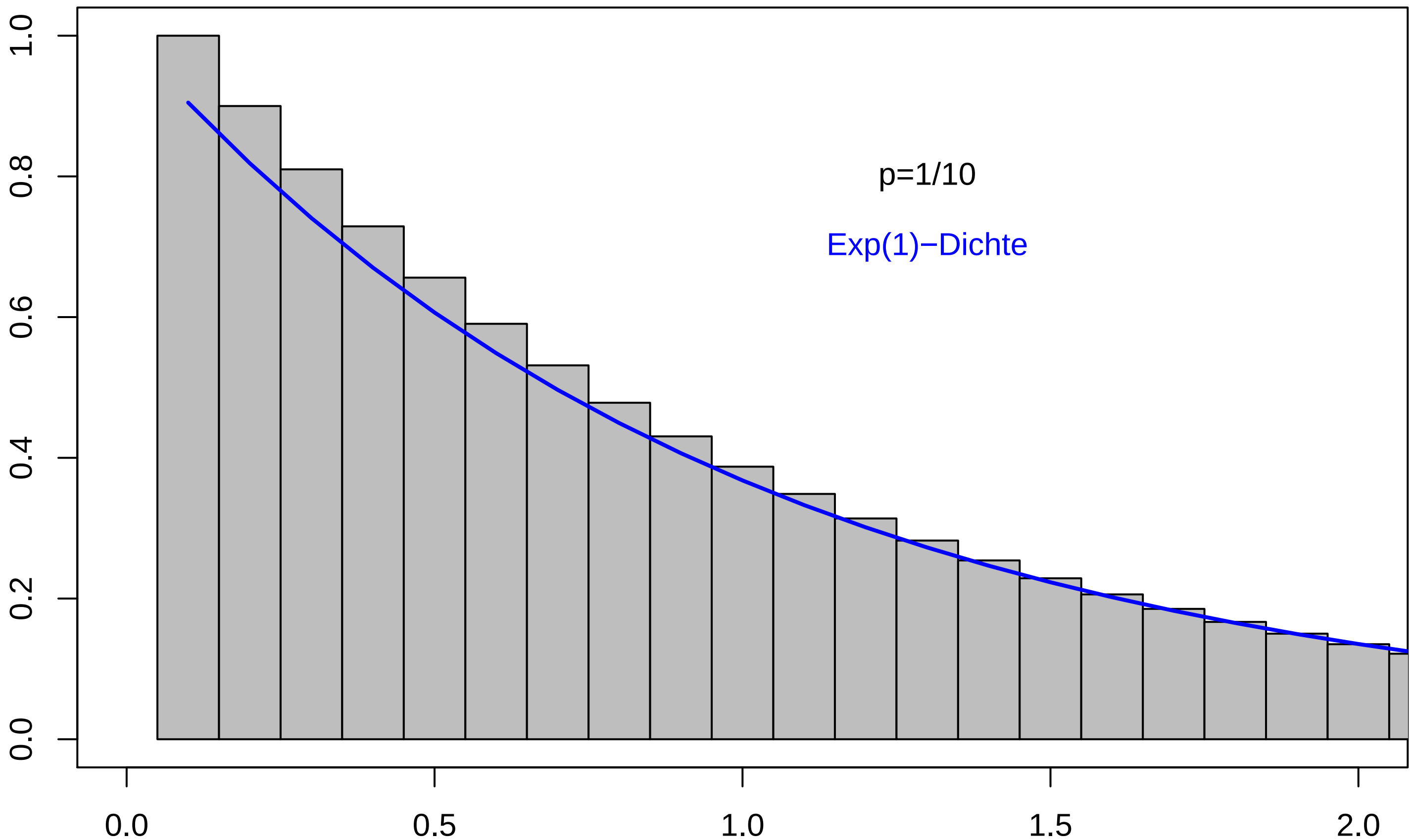
$$\mathbf{P}(\tfrac{Y}{100} = \tfrac{k}{100}) = \mathbf{P}(Y = k)$$
$$= \frac{1}{100}(1 - \frac{1}{100})^{k-1} = \frac{1}{100}\left(\left(1 - \frac{1}{100}\right)^{100}\right)^{(k-1)/100}$$
$$\approx \frac{1}{100}\, e^{-(k-1)/100} \approx \mathbf{P}(X \in [\tfrac{k-1}{100}, \tfrac{k}{100}])$$

für ein standard-exponentialverteiltes X.

Das wird durch das folgende Bild veranschaulicht:


Gewichte des p−fachen einer Geom(p)−verteilten ZV
p=1/100
Exp(1)−Dichte
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0

Gewichte des p−fachen einer Geom(p)−verteilten ZV



Salopp gesprochen:

Für kleines p hat eine Geom(p)-verteilte Zufallsvariable Y

einen großen Erwartungswert, nämlich $\frac{1}{p}$.

Man holt die Verteilung von Y zurück ins Schaubild,

indem man pY betrachtet.

Wir erinnern an die “Exponentialapproximation”:

(Vorlesung 4a und Buch Seite 42):

Ist für jedes $n \in \mathbb{N}$ die Zufallsvariable X_n geometrisch verteilt

mit $\mathbf{E}[X_n] \to \infty$ für $n \to \infty$, dann gilt:

$$\mathbf{P}\Big(\frac{X_n}{\mathbf{E}[X_n]} > t\Big) \to e^{-t} \quad \text{für alle } t \in \mathbb{R}_+.$$

$$\mathbf{P}\Big(\frac{X_n}{\mathbf{E}[X_n]} > t\Big) \to e^{-t} \quad \text{für alle } t \in \mathbb{R}_+.$$

Ist X eine standard-exponentialverteilte Zufallsvariable,

dann kann man dies schreiben als

$$\mathbf{P}\Big(\frac{X_n}{\mathbf{E}[X_n]} > t\Big) \to \mathbf{P}(X > t), \quad t \in \mathbb{R}_+.$$

Man sagt dafür auch:

Die Folge der Zufallsvariablen $X_n/\mathbf{E}[X_n]$

konvergiert in Verteilung

gegen die Zufallsvariable X.

Jetzt zum Thema:

Binomialverteilungen

mit großem Erwartungswert und großer Varianz

Wir wissen schon aus V4b4:

Für großes n und kleines p,

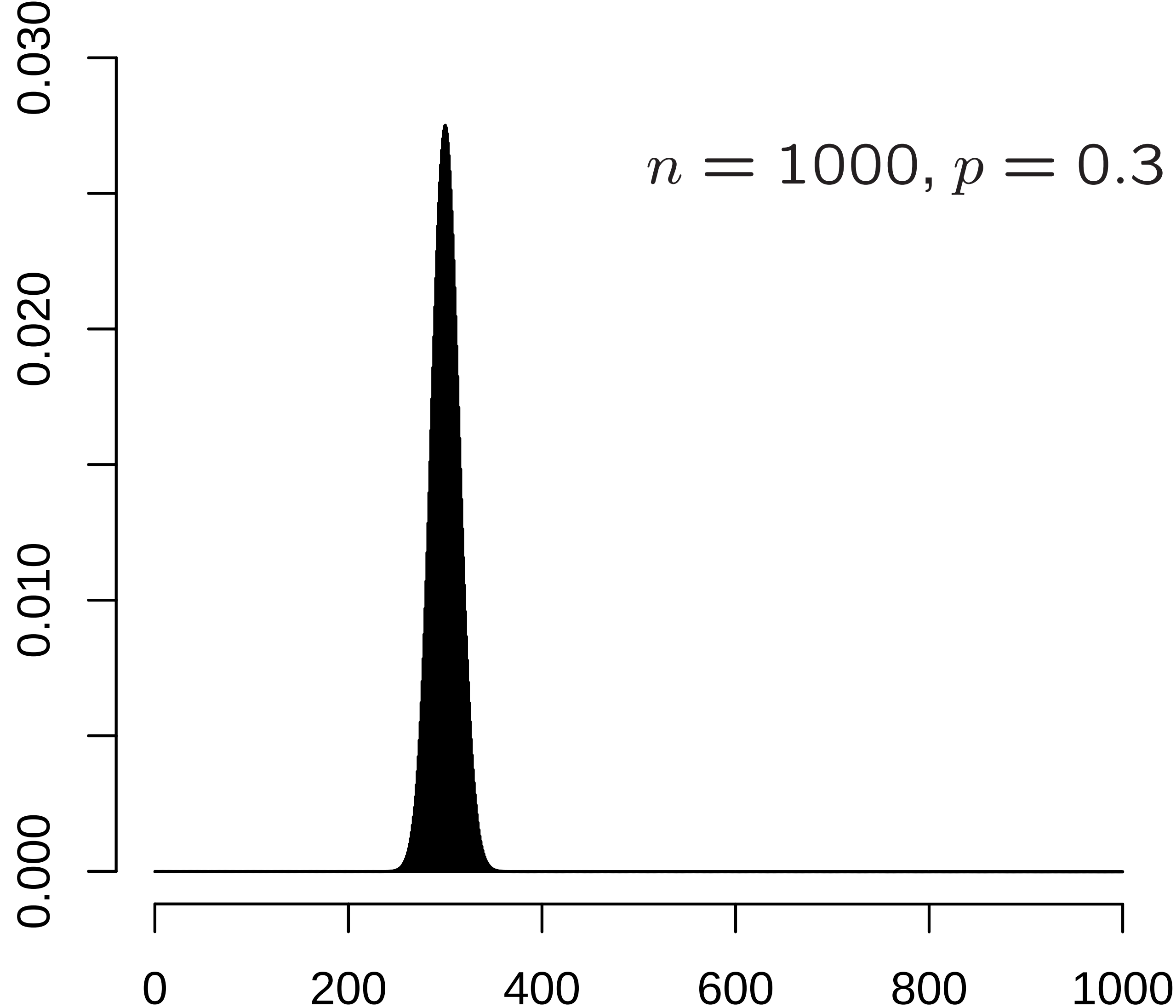
so dass $np \approx npq \approx \alpha$,

ist die Binomialverteilung Bin(n, p)

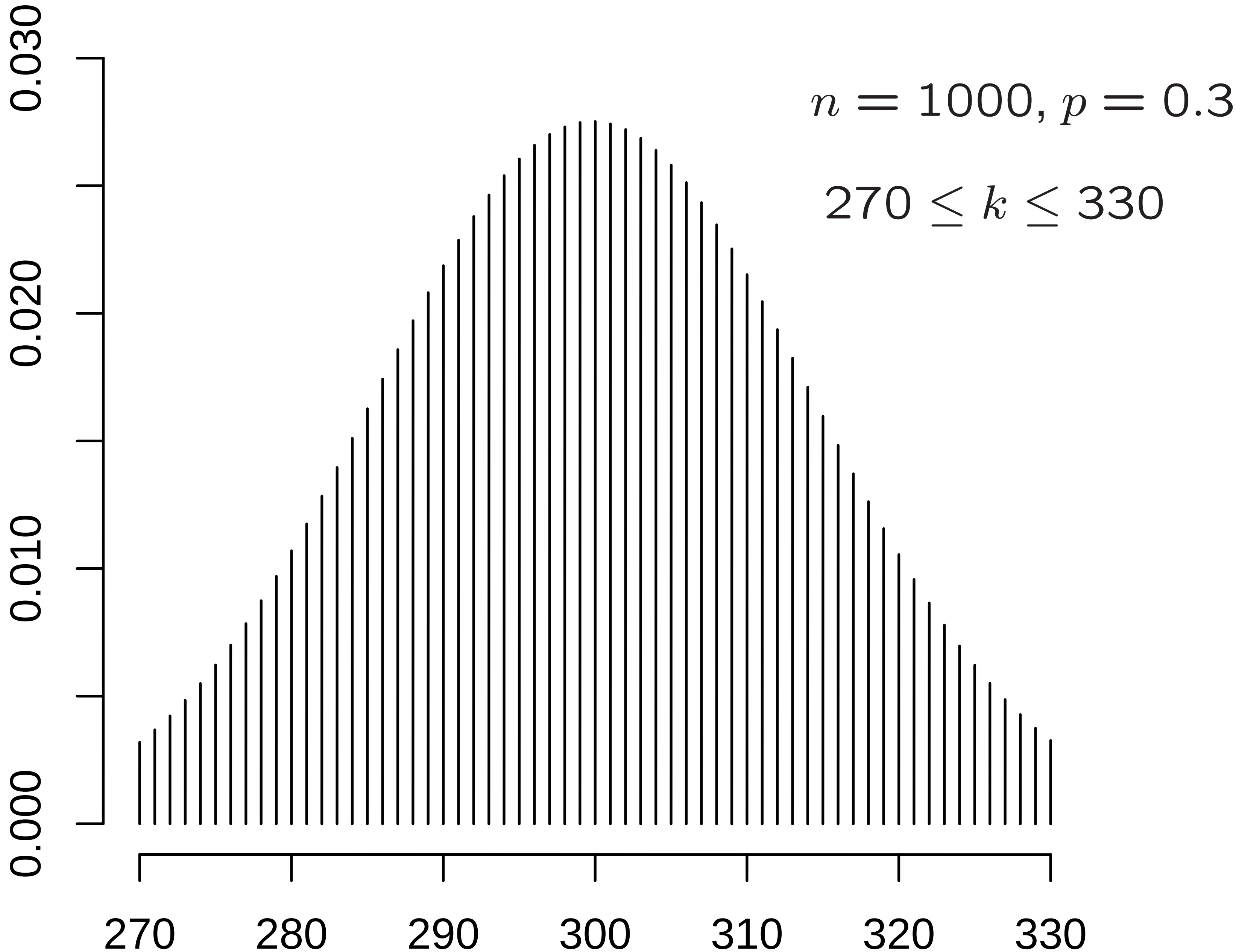
approximativ gleich Pois(α).

Wie sieht aber sieht die Bin(n, p)-Verteilung

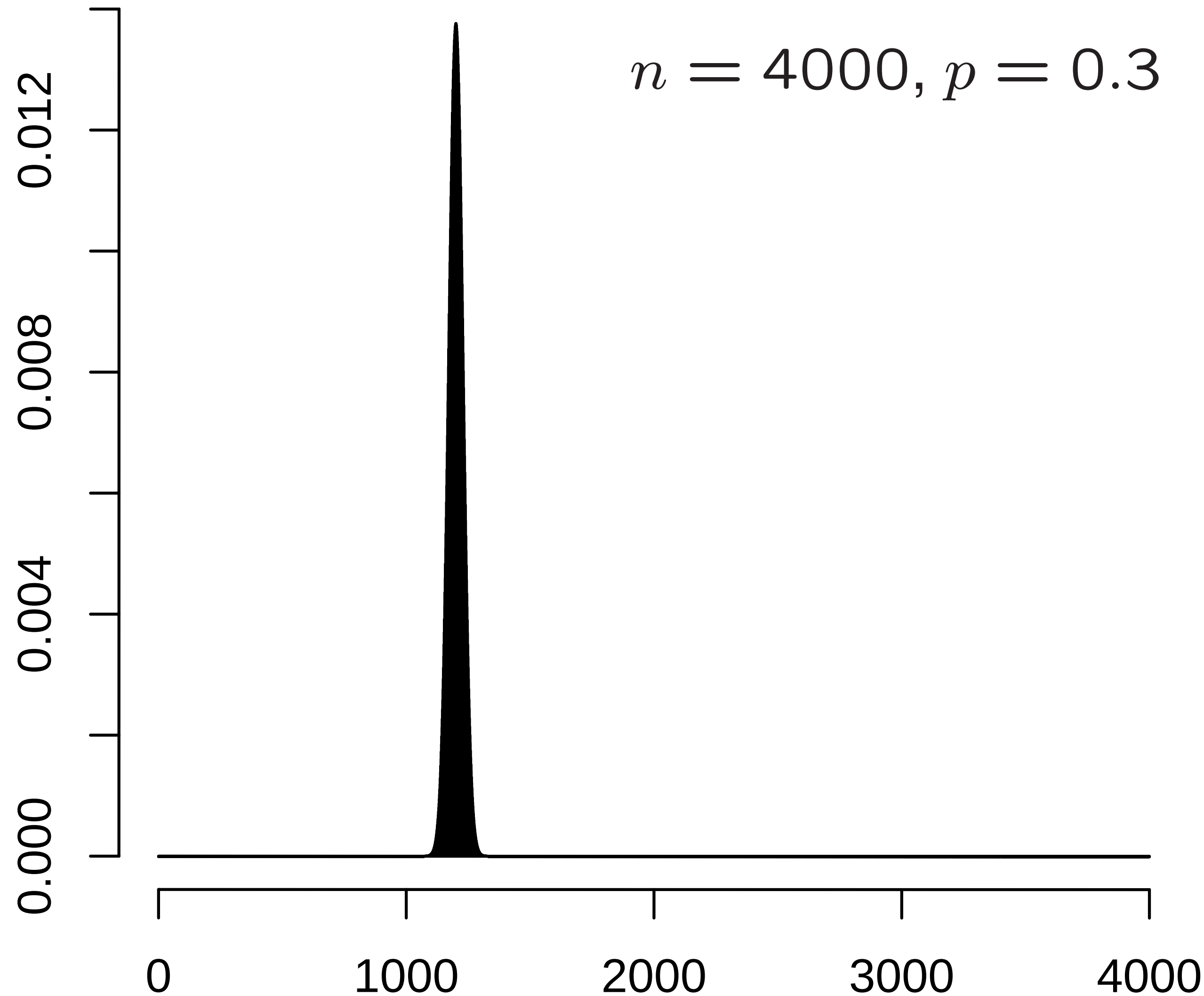
mit großem n und großem npq aus? ?



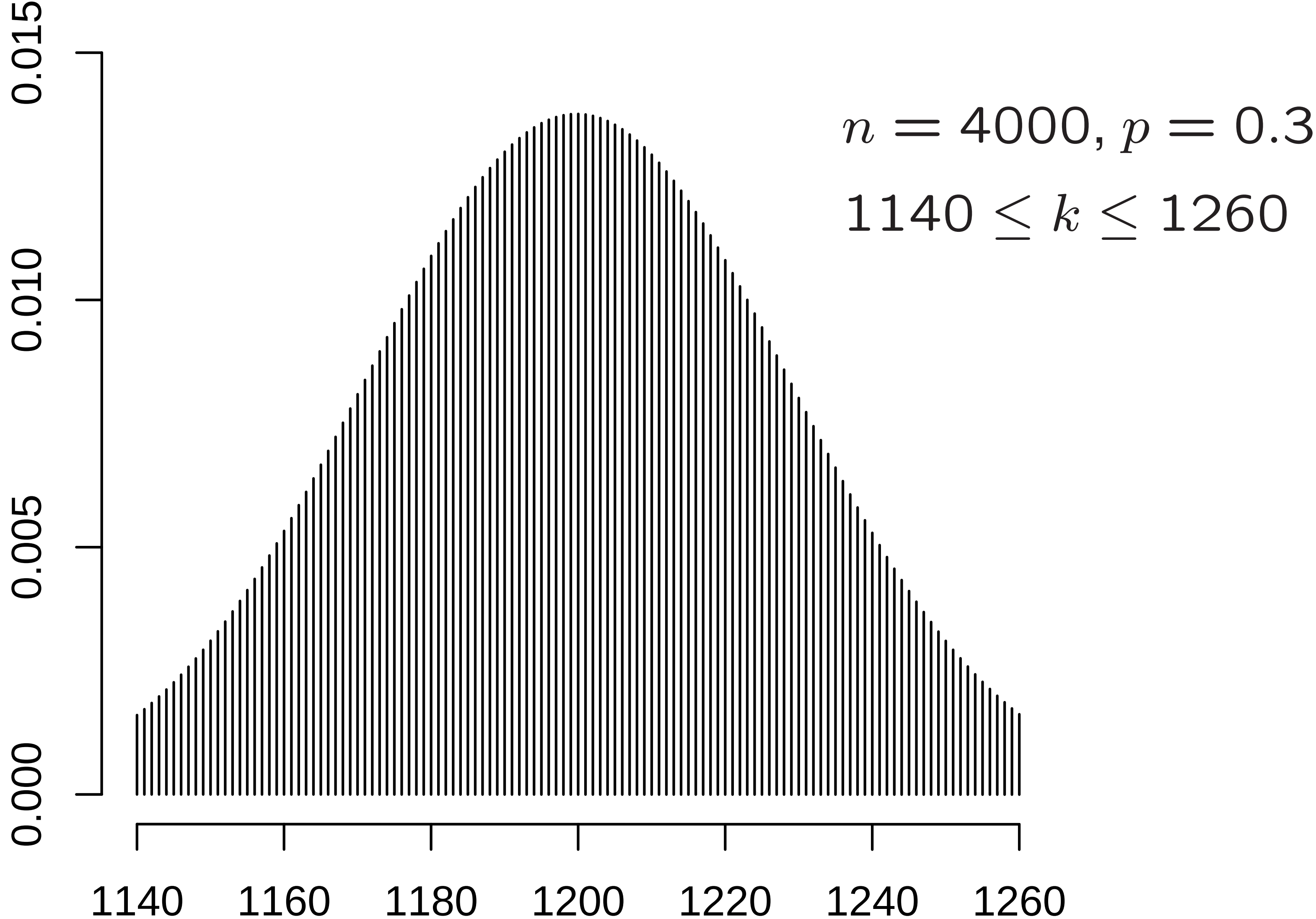
$n = 1000, p = 0.3$
0.000
0.010
0.020
0.030
0
200
400
600
800
1000



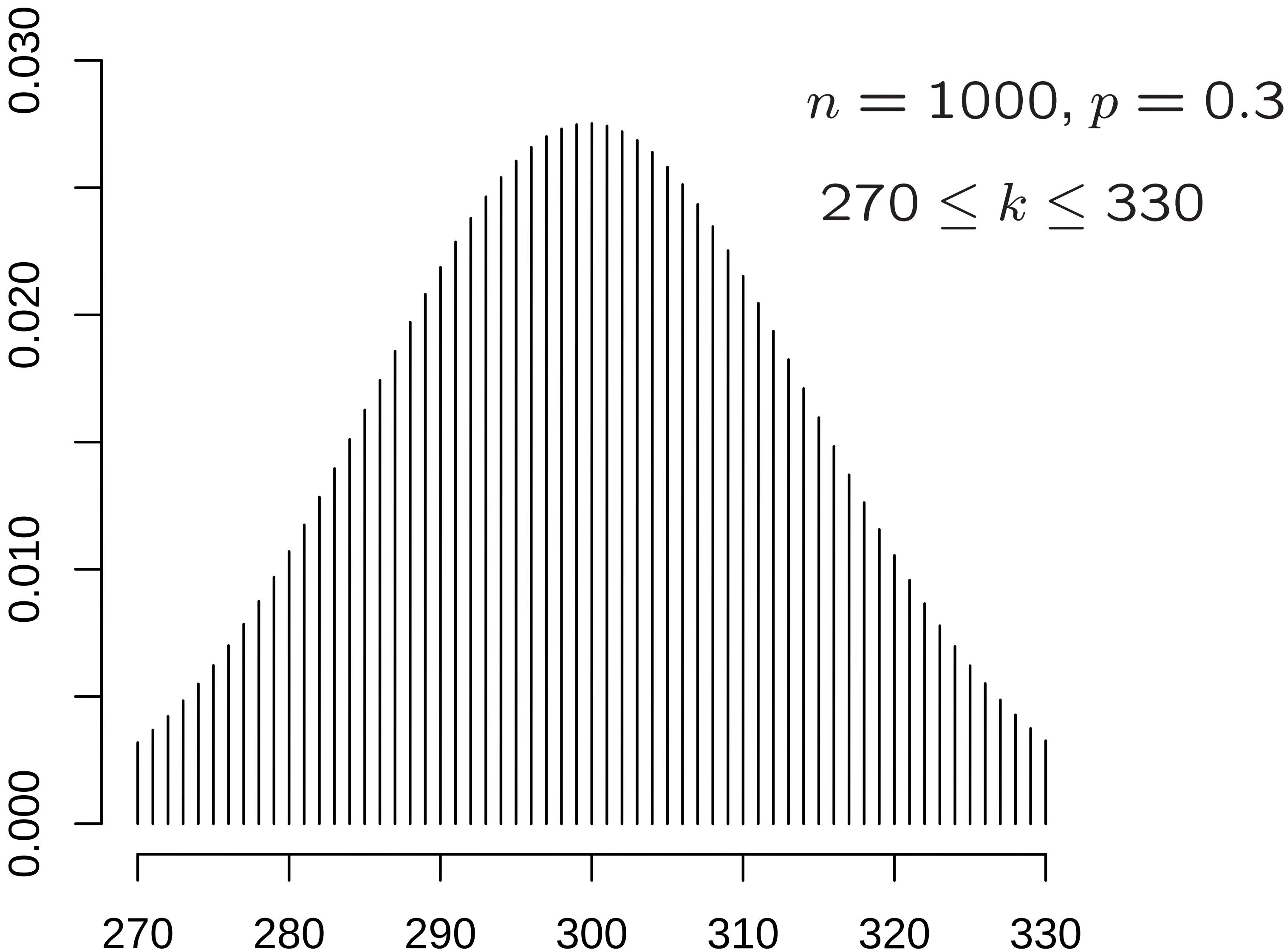
0.000
0.010
0.020
0.030
$n = 1000, p = 0.3$
$270 \le k \le 330$
270
280
290
300
310
320
330



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



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



$n = 1000, p = 0.3$
$270 \le k \le 330$
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):

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

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

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

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

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

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

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

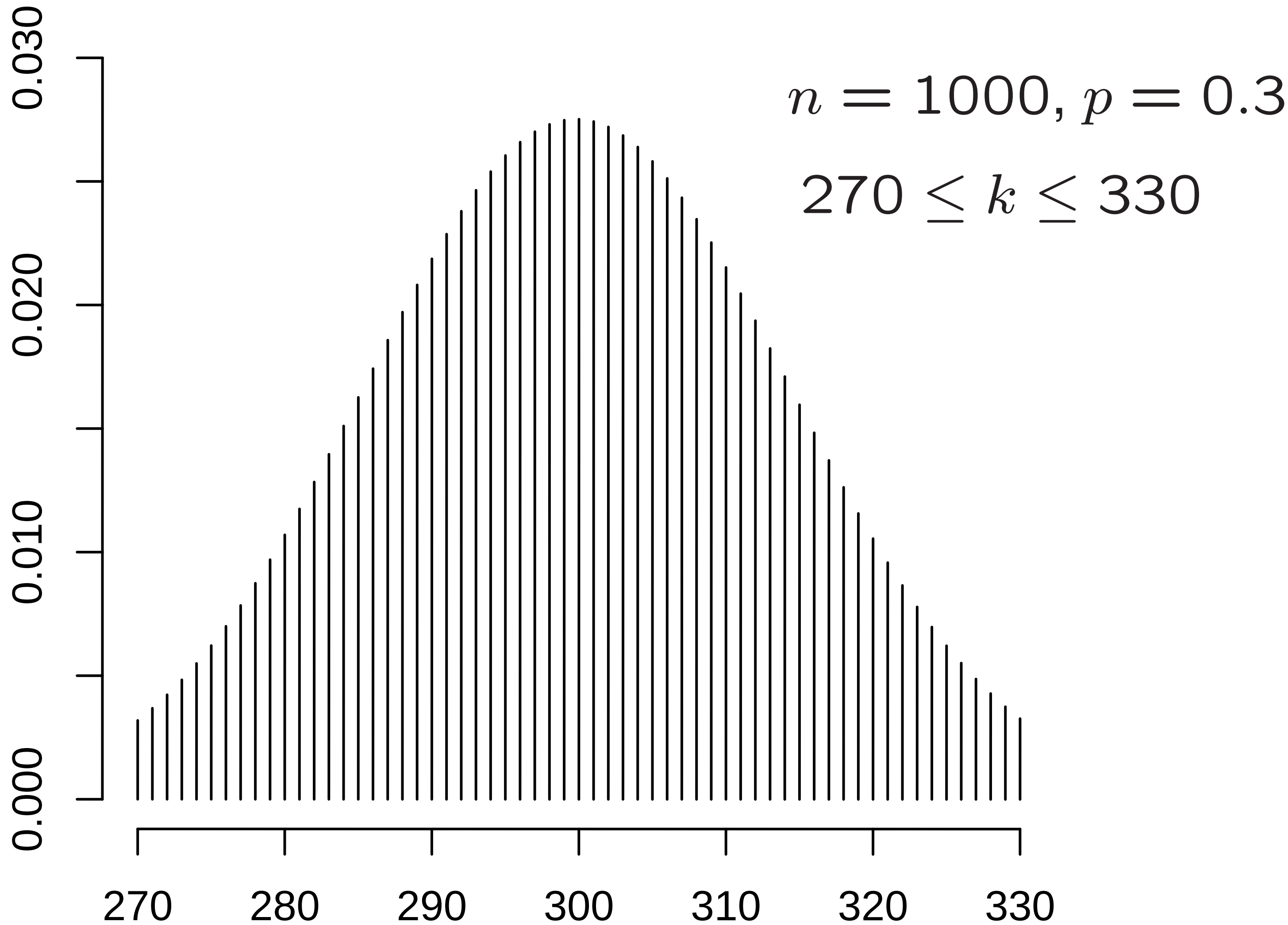
Wenn man n vervierfacht, verdoppelt sich σ.

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

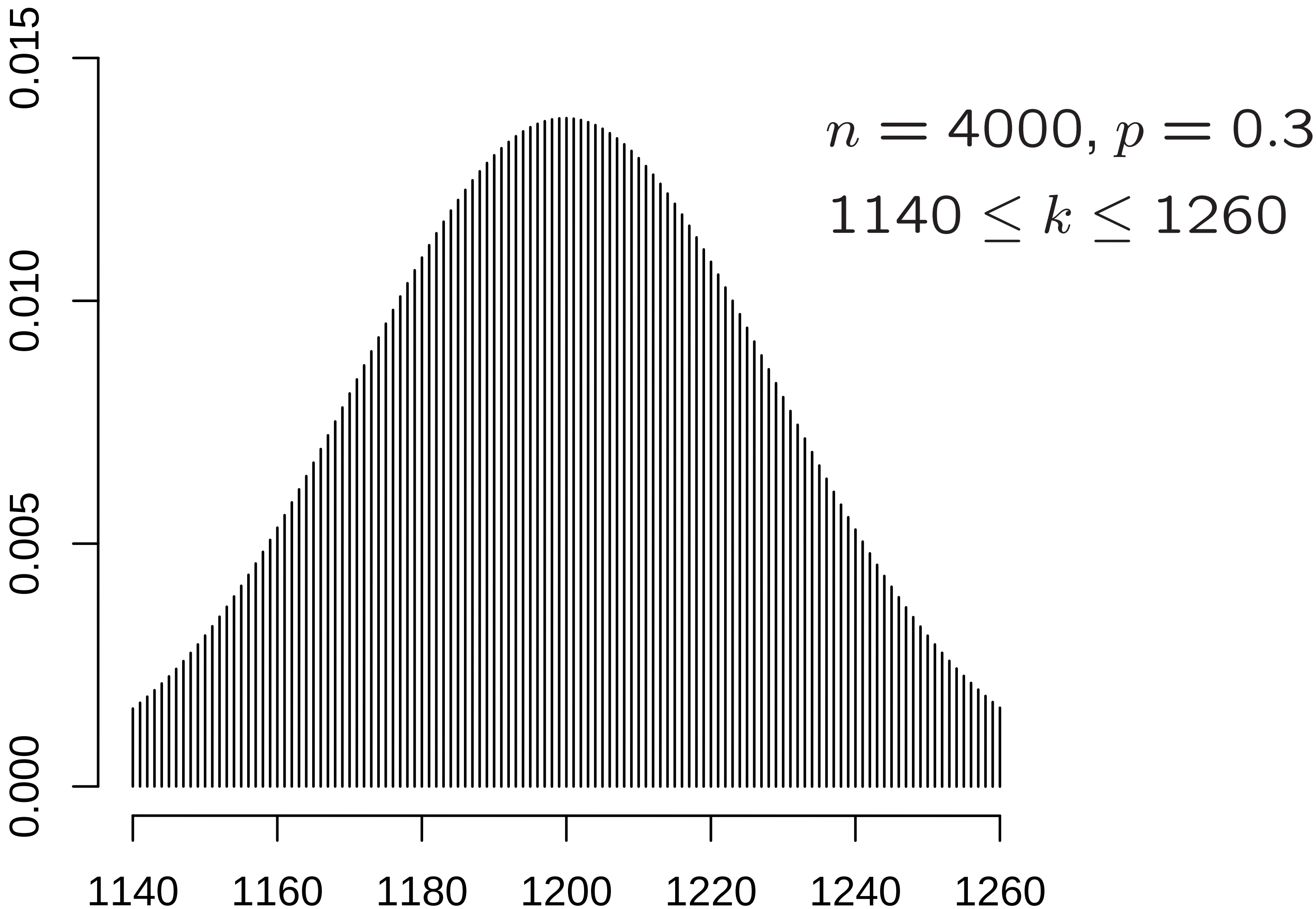
Im Bereich von einer Standardabweichung um das Zentrum μ

bringt man doppelt so viele k unter.

Das Gewicht jedes einzelnen k halbiert sich.



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


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

Die Ähnlichkeit ist unverkennbar:

Carl Friedrich Gauss (1777-1855)