Zusammenfassend haben wir bewiesen:

Satz 5.1 (Satz von de Moivre-Laplace[2], lokale Normalapproximation der Binomialverteilung). *Sei* $C > 0$, $p \in (0,1)$,

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}z^2\right)$$

die Dichte der Standard-Normalverteilung. Mit

$$z_n(k) := \frac{k - np}{\sqrt{np(1-p)}}$$

gilt

$$\lim_{n\to\infty} \max_{k\,:\,|k-np|\le C\sqrt{n}} \left| \frac{\mathrm{Bin}_{n,p}(\{k\})}{\frac{1}{\sqrt{np(1-p)}}\varphi\big(z_n(k)\big)} - 1 \right| = 0.$$

also $|z_n(k)| \le \frac{C}{\sqrt{p(1-p)}}$

$$\underbrace{P(Z_n = k)}_{=\binom{n}{k} p^k (1-p)^{n-k}} = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{np(1-p)}} e^{-\frac{1}{2}\frac{1}{p(1-p)}\left(\frac{k-np}{\sqrt{n}}\right)^2} \cdot (1 + o(1))$$

(für k mit: $|k - n \cdot p| \le C \cdot \sqrt{n}$)

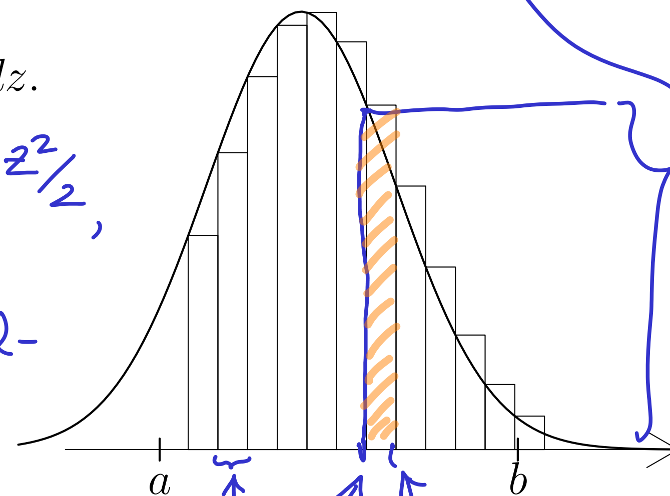
[2] Abraham de Moivre, 1667–1754; Pierre-Simon Laplace, 1749–1827

Korollar 5.2 (Normalapproximation der Binomialverteilung). *Sei* $p \in (0,1)$, $Z_n \sim \mathrm{Bin}_{n,p}$, *setze*

$$Z_n^* := \frac{Z_n - np}{\sqrt{np(1-p)}}.$$

Dann gilt für $-\infty \le a < b \le \infty$

$$\lim_{n\to\infty} P(a \le Z_n^* \le b) = \int_a^b \varphi(z)\,dz.$$



Sei $-\infty < a < b < \infty$.

(mit $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, Dichte der Standard-Normal-Vert.)

$\varphi(z_n(k))$

$$P(a \le Z_n^* \le b)$$

$$= \sum_{k:\ a \le \underbrace{\frac{k-n\cdot p}{\sqrt{np(1-p)}}}_{=z_n(k)} \le b} P(Z_n = k) = \left(\sum_{k:\ a \le z_n(k) \le b} \frac{\varphi(z_n(k))}{\sqrt{np(1-p)}} \right) \cdot (1 + o(1))$$

$z_n(k)$

$z_n(k) + \frac{1}{\sqrt{np(1-p)}} = z_n(k+1)$

Maschenweite $\frac{1}{\sqrt{np(1-p)}}$

$$\Big\downarrow\ n \to \infty$$

$$\int_a^b \varphi(z)\,dz$$

(Riemann-Approx. des Integrals)

Sei nun $a = -\infty < b < \infty$:

Zu $\varepsilon > 0$ wähle $a' < 0$ so, dass für $n \in \mathbb{N}$

$$P(Z_n^* < a') \leq P(|Z_n^*| > |a'|) \leq \frac{1}{(a')^2} < \frac{\varepsilon}{4}$$

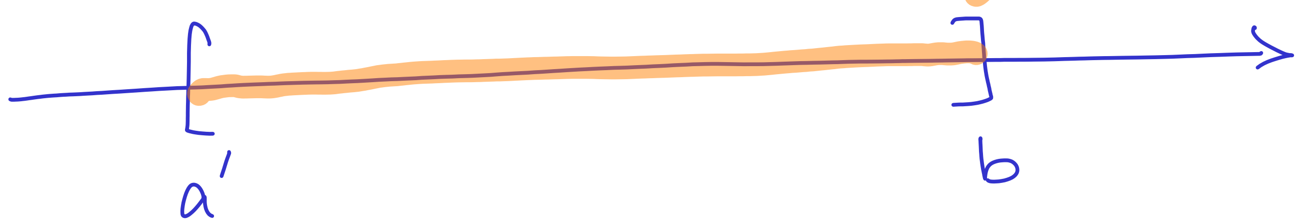
(mit Chebychev-Ungleichung, Satz 4.1) und

$$\int_{-\infty}^{a'} \varphi(z)\,dz < \frac{\varepsilon}{4}$$

gilt. Dann ist für n genügend groß

$$\left|P(Z_n^* \leq b) - \int_{-\infty}^{b} \varphi(z)\,dz\right| \leq \left|P(a' \leq Z_n^* \leq b) - \int_{a'}^{b} \varphi(z)\,dz\right| + \left|P(Z_n^* < a')\right| + \int_{-\infty}^{a'} \varphi(z)\,dz$$
$$\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon.$$

Der Fall $b = \infty$ kann analog behandelt werden (Übung). □

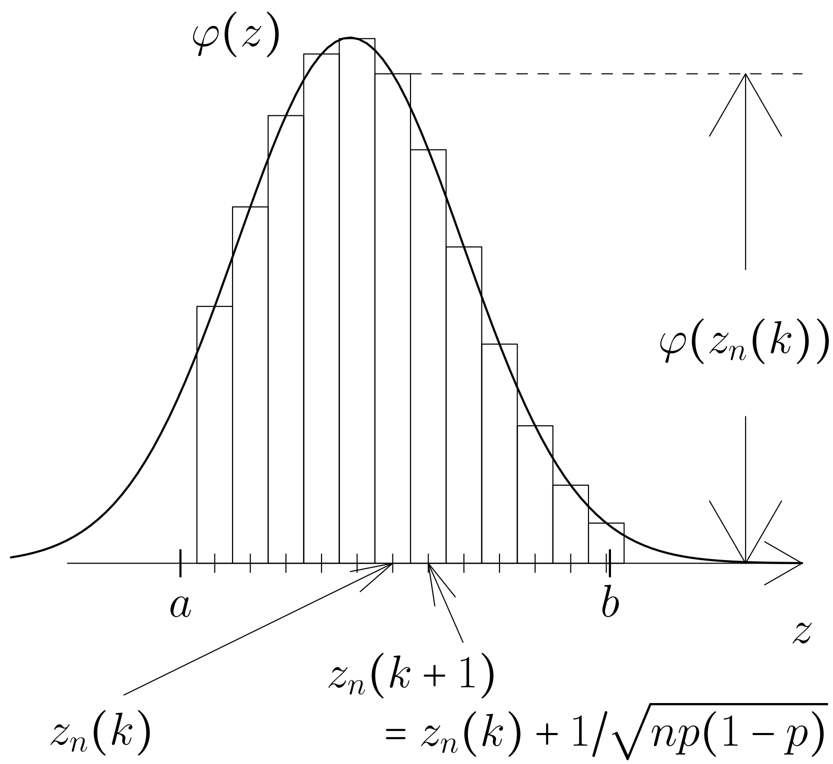


Bemerkung 5.3 (Stetigkeitskorrektur)**.** Für numerische Approximationen betrachtet man oft

$$\mathrm{Bin}_{n,p}(\{k, k+1, \ldots, \ell\}) \approx \Phi\left(\frac{\ell + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

für $0 \leq k \leq \ell \leq n$ (mit $\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-z^2/2}\, dz$ der Verteilungsfunktion der Standardnormalverteilung).

Man betrachtet also Histogramm-Balken wie im Beweis von Kor. 5.2, die aber jeweils an $z_n(k)$ „zentriert" sind.



$$\Phi\left(\frac{\ell - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)$$

$$\left(\mathrm{Bin}_{n,p}(\{k, k+1, \ldots, \ell\}) \approx \int_{(k-np)/\sqrt{np(1-p)}}^{(\ell-np)/\sqrt{np(1-p)}} \varphi(z)\, dz \right)$$

$$\mathrm{Bin}_{n,p}\big(\{k,k+1,\dots,\ell\}\big) \approx \Phi\Big(\frac{\ell+\frac{1}{2}-np}{\sqrt{np(1-p)}}\Big) - \Phi\Big(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\Big)$$

Die numerische Approximation ist (speziell für eher kleine Werte von n) meist besser mit „Stetigkeitskorrektur“, Beispiele:

n	p	k	ℓ	$Bin_{n,p}(\{k,k+1,\dots,\ell\})$	Approx. ohne	Approx. mit
15	0,3	5	9	0,4809	0,3835	0,4976
50	0,3	9	15	0,5509	0,4680	0,5389
1000	0,3	250	290	0,2567	0,2448	0,2558

Hierbei gibt „Approx. ohne“ jeweils den Wert $\Phi\Big(\frac{\ell-np}{\sqrt{np(1-p)}}\Big) - \Phi\Big(\frac{k-np}{\sqrt{np(1-p)}}\Big)$ und „Approx. mit“ den Wert $\Phi\Big(\frac{\ell+0{,}5-np}{\sqrt{np(1-p)}}\Big) - \Phi\Big(\frac{k-0{,}5-np}{\sqrt{np(1-p)}}\Big)$ an (jeweils auf 4 Nachkommastellen gerundet).

Bericht und Definition 5.4. Der Sachverhalt aus Korollar 5.2 wird auch ausgesprochen als „Konvergenz in Verteilung“:

$$X_n \xrightarrow[n\to\infty]{} X \text{ in Verteilung} \quad \left(\text{auch } X_n \xrightarrow[n\to\infty]{d} X \text{ oder } X_n \xrightarrow[n\to\infty]{\mathscr{L}} X \text{ geschrieben}\right),$$

wenn gilt

$$\lim_{n\to\infty} P(X_n \le x) = P(X \le x) \ \left(= F_X(x)\right)$$

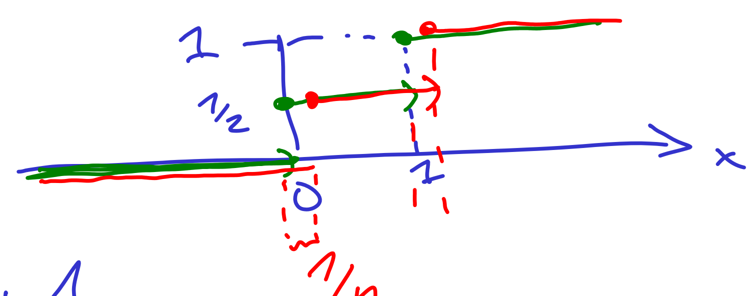
für jedes $x \in \mathbb{R}$, an dem F_X stetig ist.

$$\lim_{n\to\infty} P(Z_n^* \le b) = \int_{-\infty}^{b} \varphi(z)\,dz = P(Z \le b)$$

$\sim \mathcal{N}_{0,1}$

Nebenbei: Warum nur Konvergenz an Stetigkeitsstellen von F_X fordern?

Bsp. $X \sim \mathrm{Ber}_{1/2}$



$X_n = X + \frac{1}{n}$

$F_{X_n}(0) = 0 \neq \frac{1}{2} = F_X(0)$

Beispiel („Macht entschlossener Minderheiten"). An einer Wahl zwischen Vorschlag A und Vorschlag B nehmen 100.000 Wähler teil. Darunter sind 300, die fest entschlossen sind, für Vorschlag A zu stimmen; die übrigen sind unentschlossen und entscheiden sich (in unserem Modell) unabhängig per fairem Münzwurf zwischen den beiden Vorschlägen. Wie wahrscheinlich ist es, dass Vorschlag A die Mehrheit erhält?

Sei Y = Anz. d. unentschlossenen Wähler, die für A stimmen $(\sim \mathrm{Bin}_{\underbrace{99.700}_{=n}, 1/2})$,

am Ende: $Y + 300$ Stimmen für A

$$P(Y+300 \geq 50.001) = P\left(\frac{Y - \frac{1}{2} \cdot n}{\sqrt{n \cdot \frac{1}{2} \cdot \frac{1}{2}}} \geq \underbrace{\frac{49.701 - \frac{1}{2} n}{\sqrt{n \cdot \frac{1}{2} \cdot \frac{1}{2}}}}_{\approx -0{,}944..} \right)$$

$$\approx \int_{-0{,}944..}^{\infty} \varphi(z)\, dz \approx 0{,}827$$

Satz 5.5 („Zentraler Grenzwertsatz"). *Seien* $X_1, X_2, \ldots$ *u.i.v. reelle ZVn* $\in \mathscr{L}^2$ *mit* $\operatorname{Var}[X_1] \in (0, \infty)$, *dann gilt für* $-\infty \leq a < b \leq \infty$

$$\lim_{n\to\infty} P\left(a \leq \frac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n\operatorname{Var}[X_1]}} \leq b\right) = P(a \leq Z \leq b) \quad \textit{mit } Z \sim \mathcal{N}_{0,1}$$

= hat EW = 0 und Var = 1

Bemerkung. Korollar 5.2 ist ein Spezialfall dieses Satzes, indem man $Z_n \sim \operatorname{Bin}_{n,p}$ darstellt als $Z_n = X_1 + X_2 + \cdots + X_n$ mit X_i u.i.v., $X_1 \sim \operatorname{Ber}_p$.

Beobachtung. Die Aussage von Satz 5.5 gilt trivialerweise (sogar für festes $n \in \mathbb{N}$), wenn $X_i \sim \mathcal{N}_{0,1}$, da dann

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \sim \mathcal{N}_{0,1}$$

vgl. Beispiel 2.28.

Die Beweisidee für Satz 5.5 ist den allgemeinen Fall auf diesen Spezialfall zurückzuführen. Wir betrachten als Ergänzung in Abschnitt 5.1 den Beweis.

5.1 Beweis von Satz 5.5*

Seien $X_1, X_2, \dots$ u.i.v. reelle ZVn $\in \mathscr{L}^2$ mit $\mathrm{Var}[X_1] \in (0,\infty)$. Wir möchten zeigen, dass für $-\infty \leq a < b \leq \infty$ gilt

$$\lim_{n\to\infty} P\Big(a \leq \frac{X_1 + \dots + X_n - n\mathbb{E}[X_1]}{\sqrt{n\mathrm{Var}[X_1]}} \leq b\Big) = P(a \leq Z \leq b) \quad \text{mit } Z \sim \mathcal{N}_{0,1}$$

Wir können o.E. $\mathbb{E}[X_1] = 0$, $\mathrm{Var}[X_1] = 1$ annehmen, ansonsten betrachten wir

$$\widetilde{X}_i := \frac{X_i - \mathbb{E}[X_1]}{\sqrt{\mathrm{Var}[X_1]}}.$$

Lemma 5.6. *$X_1, X_2, \dots$ u.i.v. reelle ZVn, $X_i \in \mathcal{L}^2$ mit $\mathbb{E}[X_i] = 0$, $\mathrm{Var}[X_i] = 1$, $f : \mathbb{R} \to \mathbb{R}$ dreimal stetig differenzierbar, die ersten drei Ableitungen seien gleichmäßig beschränkt. Dann gilt*

$$\lim_{n\to\infty} \mathbb{E}\Big[f\Big(\frac{X_1 + \dots + X_n}{\sqrt{n}}\Big)\Big] = \mathbb{E}\big[f(Z)\big]$$

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

Beweis. Seien $Z_1, Z_2, \dots$ u.i.v., $\sim \mathcal{N}_{0,1}$, unabhängig von den X_i, schreibe

$$\begin{aligned} &f\Big(\frac{X_1 + \dots + X_n}{\sqrt{n}}\Big) - f\Big(\frac{Z_1 + \dots + Z_n}{\sqrt{n}}\Big) \\ &= \sum_{i=1}^n \Big(f\Big(W_{i,n} + \frac{X_i}{\sqrt{n}}\Big) - f\Big(W_{i,n} + \frac{Z_i}{\sqrt{n}}\Big)\Big) \end{aligned} \tag{5.1}$$

mit

$$W_{i,n} := \frac{1}{\sqrt{n}}\Big(X_1 + X_2 + \dots + X_{i-1} + Z_{i+1} + Z_{i+1} + \dots + Z_n\Big).$$

Taylor-Enwicklung (im Punkt $W_{i,n}$) liefert

$$
\begin{aligned}
&f\Big(W_{i,n}+\frac{X_i}{\sqrt{n}}\Big)-f\Big(W_{i,n}+\frac{Z_i}{\sqrt{n}}\Big)\\
&=f'\big(W_{i,n}\big)\frac{X_i-Z_i}{\sqrt{n}}+\frac{1}{2}f''\big(W_{i,n}\big)\frac{X_i^2-Z_i^2}{n}+R_{i,n}
\end{aligned}
$$

mit

$$
\big|R_{i,n}\big|\le\big|f''\big(W_{i,n}+\vartheta_{i,n}\big)-f''\big(W_{i,n}\big)\big|\frac{X_i^2}{2n}+\big|f''\big(W_{i,n}+\widetilde{\vartheta}_{i,n}\big)-f''\big(W_{i,n}\big)\big|\frac{Z_i^2}{2n}
$$

wobei $|\vartheta_{i,n}|\le\dfrac{|X_i|}{\sqrt{n}},\ |\widetilde{\vartheta}_{i,n}|\le\dfrac{|Z_i|}{\sqrt{n}}$.

Mit $C_2:=\sup\limits_{x\in\mathbb{R}}\big|f''(x)\big|,\ C_3:=\sup\limits_{x\in\mathbb{R}}\big|f'''(x)\big|$ gilt für jedes $K>0$:

$$
\big|R_{i,n}\big|\le C_3\frac{K^3}{2n^{3/2}}\mathbf{1}_{\{|X_i|\le K\}}+C_2\frac{X_i^2}{n}\mathbf{1}_{\{|X_i|>K\}}+C_3\frac{|Z_i|^3}{n^{3/2}}.
$$

Nehme Erwartungswert in (5.1):

$$
\begin{aligned}
&\Big|\mathbb{E}\Big[f\Big(\frac{X_1+\dots+X_n}{\sqrt{n}}\Big)\Big]-\mathbb{E}\big[f(Z)\big]\Big|\\
&=\Big|\sum_{i=1}^n\Big(\underbrace{\mathbb{E}\Big[f'\big(W_{i,n}\big)\frac{X_i-Z_i}{\sqrt{n}}\Big]}_{=\mathbb{E}[f'(W_{i,n})]\mathbb{E}[X_i-Z_i]/\sqrt{n}=0}+\underbrace{\mathbb{E}\Big[\frac{1}{2}f''\big(W_{i,n}\big)\frac{X_i^2-Z_i^2}{n}\Big]}_{=\mathbb{E}[f''(W_{i,n})]\mathbb{E}[X_i^2-Z_i^2]/(2n)=0}+\mathbb{E}\big[R_{i,n}\big]\Big)\Big|\\
&\le\sum_{i=1}^n\mathbb{E}\big[|R_{i,n}|\big]\le n\Big(C_3\frac{K^3}{2n^{3/2}}+\frac{C_2}{n}\mathbb{E}\big[X_1^2\mathbf{1}_{\{|X_1|>K\}}\big]+\frac{C_3}{n^{3/2}}\mathbb{E}\big[|Z_1|^3\big]\Big),
\end{aligned}
$$

also

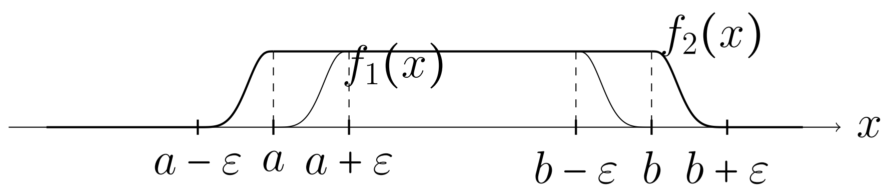
$$\limsup_{n\to\infty}\left|\mathbb{E}\left[f\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right]-\mathbb{E}\left[f(Z)\right]\right| \leq C_2\mathbb{E}\left[X_1^2\mathbf{1}_{\{|X_1|>K\}}\right]\xrightarrow[K\to\infty]{}0$$

□

Beweis von Satz 5.5. Seien $-\infty < a < b < \infty$. Zu $0 < \varepsilon < \frac{b-a}{2}$ wähle f_1, f_2, die den Voraussetzungen von Lemma 5.6 genügen und

$$\mathbf{1}_{[a+\varepsilon,b-\varepsilon]} \leq f_1 \leq \mathbf{1}_{[a,b]} \leq f_2 \leq \mathbf{1}_{[a-\varepsilon,b+\varepsilon]}$$

erfüllen (siehe Skizze).



Es ist

$$\begin{aligned} P\big(Z\in[a+\varepsilon,b-\varepsilon]\big) &\leq \mathbb{E}\left[f_1(Z)\right] = \lim_{n\to\infty}\mathbb{E}\left[f_1\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right] \\ &\leq \liminf_{n\to\infty} P\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\in[a,b]\right) \\ &\leq \limsup_{n\to\infty} P\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\in[a,b]\right) \\ &\leq \lim_{n\to\infty}\mathbb{E}\left[f_2\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right] = \mathbb{E}\left[f_2(Z)\right] \leq P\big(Z\in[a-\varepsilon,b+\varepsilon]\big), \end{aligned}$$

mit $\varepsilon \downarrow 0$ folgt die Behauptung.

Die Fälle $a = -\infty$ oder $b = +\infty$ kann man analog behandeln. □