Kapitel 4

Der zentrale Grenzwertsatz

Die Normalverteilung ist die wichtigste Wahrscheinlichkeitsverteilung: Viele in der Natur oder im Alltag auftretende Größen sind (näherungsweise) normalverteilt, zum Beispiel das Körpergewicht von Personen, das tatsächliche Gewicht von 1 kg-Zuckerpaketen, oder die Zahl der Borsten auf dem Abdomen von Drosophila. (Die letzte Größe ist diskret. Überlegen Sie sich, was die behauptete Aussage der näherungsweisen Normalität für dieses Beispiel bedeuten soll!) Grob gesprochen besteht die Ursache für das Aufreten der Normalverteilung darin, dass Mittelwerte (fast) beliebig verteilter Zufallsvariablen näherungsweise normalverteilt sind, wenn man nur über hinreichend viele Summanden mittelt! Aus diesem Grund sind auch viele Messfehler in guter Näherung normalverteilt. Diese sehr erstaunliche Aussage wollen wir jetzt präzisieren und begründen.

4.1 Aussage und Beweisskizze

Wir betrachten Folgen $\{X_i\}_{i\geq 1}$ von **i.i.d.**[1] Zufallsvariablen X_i, alle Folgenglieder sind also unabhängig und von derselben Verteilung bestimmt. Ebenso betrachten wir die Folge der n-ten Mittelwerte $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$. Man kann den Vektor $(X_1, \dots, X_n)$ als "mathematische Stichprobe" betrachten, also als diejenige ZV, die den Elementen der Stichprobe entspricht, bevor das Experiment durchgeführt wird. Nach dem Experiment haben wir die Realisierungen $x_1, \dots, x_n$ erhalten, beziehungsweise $\overline{x}$.

Satz 4.1 (Zentraler Grenzwertsatz)**.** Sei $\{X_i\}_{i\geq 1}$ eine Folge von i.i.d. Zufallsvariablen mit Erwartungswert $\mathbb{E}(X_i) = \mu$, wobei $|\mu| < \infty$ gelten soll, und Varianz $0 < \mathbb{V}(X_i) = \sigma^2 < \infty$. Definiere die Zufallsvariable

$$Z_n := \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) = \frac{\overline{X} - \mu}{(\sigma/\sqrt{n})}. \tag{4.1}$$

[1] independent and identically distributed (unabhängig und identisch verteilt)

Dann ist Z_n asymptotisch standard-normalverteilt, d.h.

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq b) \;=\; F(b), \tag{4.2}$$

wobei $F(x)$ die Verteilungsfunktion der Standard-Normalverteilung $\mathcal{N}(0,1)$ bezeichnet.

Bemerkung. Für die Verteilungsfunktion der Standard-Normalverteilung wird in der Regel das Symbol Φ verwendet, für die zugehörige Dichte das Symbol φ. Wir werden erst später zu diesen Bezeichnungen übergehen, um Verwechslungen mit der charakteristischen Funktion ϕ zu vermeiden.

Der Nenner $\sigma/\sqrt{n}$ in Gl. (4.1) hat eine konkrete Bedeutung: Er ist die Standardabweichung der Zufallsvariablen $\overline{X}$. Im Zähler ist μ der Erwartungswert von allen X_i, und zugleich von $\overline{X}$. Wir werden diesen Zusammenhängen später im Abschnitt "Eigenschaften des Mittelwertes" wieder begegnen.

Beweis-Skizze. Ohne Einschränkung der Allgemeinheit betrachten wir den Spezialfall $\mu = 0$ und $\sigma = 1$. Der allgemeine Fall kann durch die Transformation $X_i' = (X_i - \mu)/\sigma$ auf den Spezialfall zurückgeführt werden.

Wir betrachten jetzt die Taylor-Entwicklung (oder Taylor-Reihe) der charakteristischen Funktion $\phi_{X_i}(t) \;=\; \phi_X(t) \;=\; \mathbb{E}(\mathrm{e}^{\mathrm{i}tx})$ um 0. Da alle X_i identisch verteilt sind, schreiben wir einfach X anstelle von X_i. Weil die ersten drei Momente von X bekannt sind, nämlich $M_0 = 1$, $M_1 = \mu = 0$ und $M_2 = \sigma^2 = 1$, kennen wir auch die Funktionswerte und Ableitungen der charakteristischen Funktion,

$$\phi_X^{(k)} \;=\; \mathrm{i}^k M_k,$$

für $k \in \{0,1,2\}$. Also können wir die Taylor-Reihe darstellen als

$$\phi_X(t) \;=\; 1 - \mathrm{i}\cdot 0 + \frac{t^2}{2}\cdot(-1) + R(t) \;=\; 1 - \frac{t^2}{2} + R(t). \tag{4.3}$$

Das Restglied $R(\cdot)$ besitzt die Eigenschaft $R(t) \;=\; o(t^2)$, erfüllt also die asymptotische Bedingung $\lim_{t\to 0} R(t)/t^2 \;=\; 0$, wie wir auch weiter unten sehen und verwenden werden.

Es reicht nun zu zeigen, dass die charakteristischen Funktionen asymptotisch übereinstimmen, d.h. dass $\lim_{n\to\infty} \phi_{Z_n}(t) = \exp\bigl(-\frac{t^2}{2}\bigr)$ gilt. Dann folgt mit dem Stetigkeitssatz von Lévy — der hier nicht bewiesen wird — die punktweise Konvergenz der Verteilungsfunktion von Z_n gegen die der Standard-Normalverteilung.

Wegen $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i = \sum_{i=1}^n \frac{X_i}{\sqrt{n}}$ folgt sofort:

$$\phi_{Z_n}(t) \;=\; \left(\phi_{\frac{X}{\sqrt{n}}}(t)\right)^n \;=\; \left(\phi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n \;=\; \left(1 - \frac{t^2}{2n} + R\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

Die erste Gleichheit besteht wegen des Faltungssatzes, die zweite folgt aus der Identität $\phi_{\alpha X}(t) = \phi_X(\alpha t)$, und die dritte gilt wegen der Taylor-Entwicklung von ϕ_X aus (4.3). Nun müssen wir zeigen, dass das Restglied $R(\cdot)$ schnell genug verschwindet. Dazu benutzen wir eine der vielen Darstellungen des Restglieds im Satz von Taylor, nämlich die *Integraldarstellung* (sehen Sie in Ihrem Skript zur Analysis nach!). Wir erhalten in unserem Fall (nach partieller Integration)

$$R(t) \;=\; \int_0^t (t-\tau)\,\big(\phi_X''(\tau) - \phi_X''(0)\big)\,\mathrm{d}\tau .$$

Wir verwenden auf dem Intervall $[0,t]$ mit $t>0$ die Abschätzung

$$\big(\phi_X''(\tau) - \phi_X''(0)\big) \;\le\; \sup_{0\le\nu\le 1} \big|\phi_X''(\nu t) - \phi_X''(0)\big| \;=:\; c(t).$$

Es gilt $0 \le c(t) < \infty$ für jedes reelle t, und $c(t) \longrightarrow 0$ für $t \to 0$. Damit erhalten wir für das Integral die Abschätzung

$$|R(t)| \;\le\; c(t)\int_0^t (t-\tau)\,\mathrm{d}\tau \;=\; c(t)\cdot\frac{t^2}{2}.$$

Wir verwenden jetzt noch die Ungleichung $|a^n - b^n| \le n\cdot|a-b|$, welche für $|a|\le 1$ und $|b| \le 1$ gültig ist. Der Fall $n=2$ ergibt sich dabei sofort aus der dritten binomischen Formel, der allgemeine Fall ist für $a=b$ trivial und folgt sonst nach Teilen durch $a-b$. Damit schätzen wir folgendermaßen ab:

$$\left|\left(1-\frac{t^2}{2n}+R\left(\frac{t}{\sqrt{n}}\right)\right)^n - \left(1-\frac{t^2}{2n}\right)^n\right| \;\le\; n\cdot\left|R\left(\frac{t}{\sqrt{n}}\right)\right|$$
$$\le n\cdot c\left(\frac{t}{\sqrt{n}}\right)\frac{t^2}{2n} \;=\; c\left(\frac{t}{\sqrt{n}}\right)\frac{t^2}{2},$$

was für beliebige t und hinreichend große n gerechtfertigt ist.

Im Grenzfall $n\to\infty$ konvergiert der rechte Term gegen 0, wegen $c(s) \longrightarrow 0$ für $s \to 0$. Die Differenz der beiden Folgen strebt also gegen 0, und damit stimmen ihre Grenzwerte überein. Insgesamt erhalten wir

$$\lim_{n\to\infty}\phi_{Z_n}(t) \;=\; \lim_{n\to\infty}\left(1-\frac{t^2}{2n}+R\left(\frac{t}{\sqrt{n}}\right)\right)^n = \lim_{n\to\infty}\left(1+\frac{-t^2/2}{n}\right)^n \;=\; \mathrm{e}^{-\frac{t^2}{2}},$$

wobei wir $\lim_{n\to\infty}\big(1+\frac{x}{n}\big)^n = \mathrm{e}^x$ aus der Analysis verwendet haben. □

Bemerkung. Einige Hinweise und Ergänzungen:

(i) Dies ist nur die schwächste Version des zentralen Grenzwertsatzes (ZGWS). Stärkere Versionen (d.h. analoge Aussagen unter schwächeren Voraussetzungen) erhält man insbesondere für paarweise unabhängige oder für i.ni.d.[2] Zufallsvariable, und sogar für bestimmte Formen der Abhängigkeit (vgl. auch das Lehrbuch von Georgii).

[2] also unabhängige, aber nicht identisch verteilte Zufallsvariablen unter der sog. Lindeberg–Feller-Bedingung an deren Varianzen.

(ii) Achtung: Der ZGWS ist eine Aussage über *Verteilungsfunktionen*, nicht über Dichten! Das kann auch gar nicht anders sein, denn er gilt ja auch für *diskrete* ZVn, deren Mittelwert für jedes endliche n immer diskret bleibt. Für kontinuierliche ZVn gibt es auch Resultate über die Konvergenz der Dichten, sogenannte *lokale* zentrale Grenzwertsätze.

4.2 Anwendungen

Wir führen zur Beschreibung von Zufallsvariablen zwei neue Symbole ein. Es bedeuten $\sim$ "verteilt nach" und $\approx$ "asymptotisch verteilt nach" für Verteilungsfunktionen, im Sinne der Aussage des ZGWS. (Beachten Sie, dass das Symbol $\sim$ in anderem Zusammenhang auch zur Beschreibung von asymptotischer Äquivalenz benutzt wird.) Die folgenden Beziehungen ergeben sich unmittelbar aus dem zentralen Grenzwertsatz:

$$\begin{aligned} Z_n &= \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \approx \mathcal{N}(0,1), \\ \overline{X} &= Z_n \cdot \frac{\sigma}{\sqrt{n}} + \mu \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \\ \sum_{i=1}^{n} X_i &= n\overline{X} \approx \mathcal{N}(n\mu, n\sigma^2), \end{aligned} \tag{4.4}$$

wobei wir die zweite und dritte über die Rechenregeln in der Tabelle auf S. 21 aus der ersten gewonnen haben; vergleiche auch die Standardisierung (1.50). Für hinreichend große n kann man also oft die Normalapproximation anstelle der (meist unbekannten) exakten Verteilung verwenden, weil der Approximationsfehler dann vernachlässigbar klein ist. Letzteres muss man aber prüfen, wie wir später bei den statistischen Anwendungen noch sehen werden.

Beispiel. Betrachten wir dies etwas konkreter:

(i) Binomialverteilung: Mit dem Faltungssatz (rückwärts eingesetzt) ergibt sich, dass $Y \sim \mathcal{B}(n,p)$ gleichbedeutend ist mit $Y = \sum_{i=1}^{n} X_i$, wobei die X_i i.i.d. Bernoulli-ZVn mit Parameter p sind, d.h. $\mathbb{E}(X_i) = p$ und $\mathbb{V}(X_i) = p(1-p)$ (vergleichen Sie dies mit den Übungen und dem Beispiel des Münzwurfs). Also gilt $Y \approx \mathcal{N}(np, np(1-p))$ für hinreichend große n.

(ii) Poisson-Verteilung: $Y \sim \mathrm{Poi}(\lambda)$ bekommen wir zum Beispiel dann, wenn wir $Y = X_1 + \ldots + X_n$ mit i.i.d. ZVn $X_i \sim \mathrm{Poi}(\lambda/n)$ haben. Dies folgt aus dem Faltungssatz, wie auch in den Übungen besprochen. Damit ergibt sich wie oben $\mathrm{Poi}(\lambda) \approx \mathcal{N}(\lambda, \lambda)$ für hinreichend große Werte von λ.

(iii) Sei X hypergeometrisch verteilt, mit Parametern W, S, n. Falls dabei die Abschätzung $n/(W+S) \lesssim 0.05$ gilt, kann man näherungsweise eine Binomialverteilung annehmen. Dann folgt $X \approx \mathcal{N}(np, np(1-p))$ mit $p = W/(W+S)$.

Wann ist n groß genug, so dass die Normalapproximation anwendbar (hinreichend genau) ist? Das hängt natürlich von der gewünschten Genauigkeit ab; im Einzelfall muss man (bekannte) Fehlerschranken anwenden. Eine genaue Untersuchung würde hier zu weit führen. Deshalb geben wir nur einige *Faustregeln* an, die für die Zwecke dieser Vorlesung ausreichen.

$$\begin{aligned} \mathcal{B}(n,p) &\approx \mathcal{N}(np, np(1-p)) \; : \quad np(1-p) \geq 9, \\ \mathrm{Poi}(\lambda) &\approx \mathcal{N}(\lambda, \lambda) \; : \quad \lambda \geq 10, \end{aligned}$$

sowie

$$\mathrm{hyper}(W,S,n) \approx \mathcal{N}(np, np(1-p)) \; : \quad \begin{array}{c} np(1-p) \geq 9, \\ \frac{n}{W+S} \leq 0.05, \\ p = \frac{W}{W+S}. \end{array}$$