7. Grenzwertsätze

Die **Grenzwertsätze** bilden den Abschluss der Wahrscheinlichkeitsrechnung und sind von zentraler Bedeutung vor allem für die induktive Statistik.

- **Gesetz der großen Zahlen**

  Aussage über die Genauigkeit der Abschätzung eines unbekannten Mittelwertes der Grundgesamtheit durch den Stichprobenmittelwert.

- **Zentrale Grenzwertsatz**

  Aussage über die Verteilung Summen und Durchschnitte beliebig verteilter Zufallsvariablen bei großem Stichprobenumfang

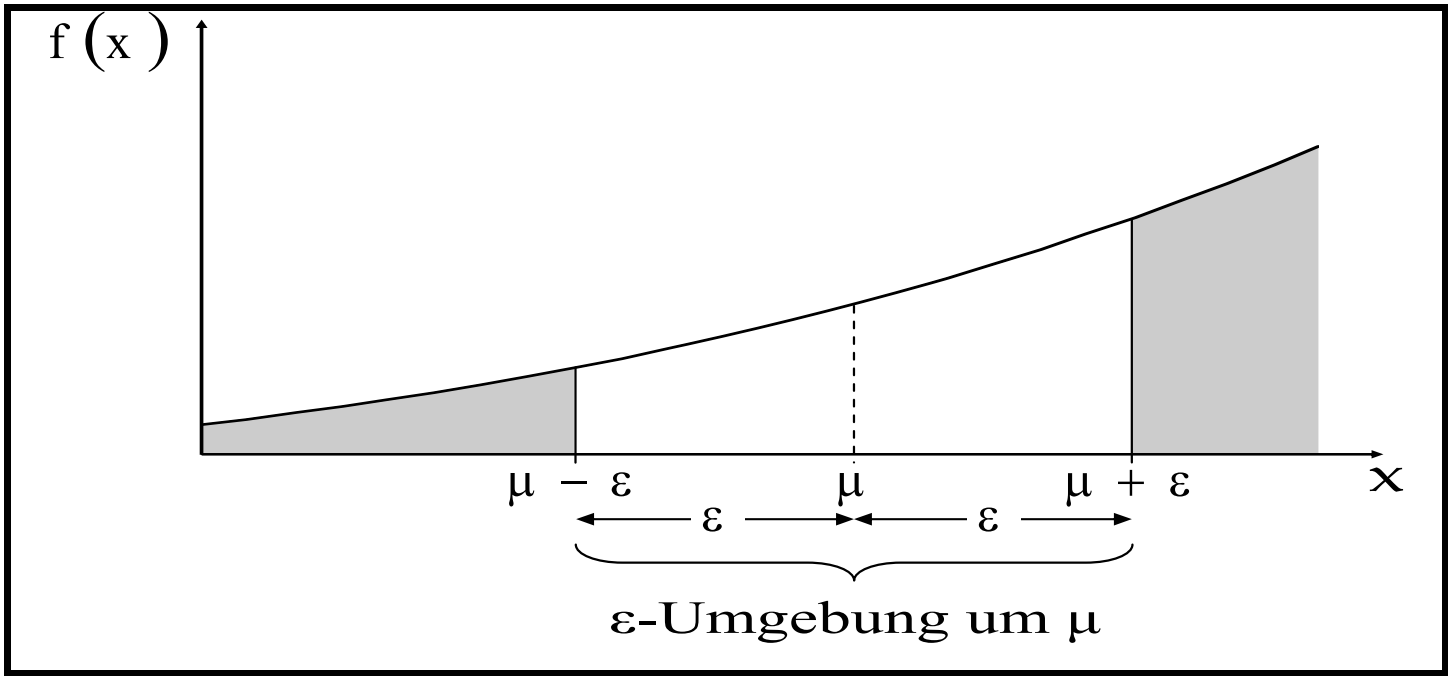
Die **Tschebyscheffsche Ungleichung** wird einerseits für den Beweis des Gesetzes der großen Zahlen benötigt. Andererseits lässt sie sich aber auch eigenständig zur Abschätzung von Wahrscheinlichkeiten bei einem unbekannten Verteilungstyp verwenden.

7.1 Tschebyscheffsche Ungleichung

Sofern für eine Zufallsvariable X die Verteilung, also die Wahrscheinlichkeits- oder Dichtefunktion, bekannt ist, lässt sich die Wahrscheinlichkeit dafür bestimmen, dass X in einem bestimmten Intervall liegt. Wie ist aber eine solche **Wahrschein-lichkeit** zu ermitteln, wenn der **Verteilungstyp nicht bekannt** ist?

Wir fragen genauer nach der **Wahrscheinlichkeit**, dass eine **beliebig verteilte Zufallsvariable innerhalb oder außerhalb** einer **ε-Umgebung um** den **Erwartungswert μ** liegt.

Abbildung: ε-Umgebung um den Erwartungswert μ



Mit der **Tschebyscheffsche Ungleichung** kann eine **Höchstwahrscheinlichkeit** dafür bestimmt werden, dass die Zufallsvariable X mindestens um ε vom Mittelwert μ abweicht.

Die **Wahrscheinlichkeit**, dass eine **beliebig verteilte Zufallsvariable** mit dem Erwartungswert μ und der Varianz σ^2 **um mindestens ε von ihrem Mittelwert abweicht**, ist **höchstens** gleich $\boldsymbol{\sigma^2/\varepsilon^2}$:

(7.1) $$P\left(|X-\mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Beweis von (7.1):

Wir beschränken uns auf den diskreten Fall und gehen von der in (5.14) gegebenen Definition der Varianz aus:

(5.14) $$V(X) = \sigma^2 = \sum_{j=1}^{m} \left(x_j - \mu\right)^2 \cdot p_j \ .$$

Werden nur Summanden betrachtet, deren Abweichung vom Erwartungswert μ mindestens gleich ε ist, $\left|x_j - \mu\right| \geq \varepsilon$, gilt

(7.2) $$\sigma^2 \geq \sum_{\left|x_j-\mu\right| \geq \varepsilon} \left(x_j - \mu\right)^2 p_j \ ,$$

da die rechte Seite von (7.2) eine Teilsumme von (5.14) ist. Da die Abweichungen $\left|x_j - \mu\right|$ mindestens gleich ε sind, gilt außerdem

(7.3) $$\sum_{\left|x_j-\mu\right|\geq\varepsilon} \left(x_j - \mu\right)^2 \cdot p_j \geq \sum_{\left|x_j-\mu\right|\geq\varepsilon} \varepsilon^2 \cdot p_j = \varepsilon^2 \sum_{\left|x_j-\mu\right|\geq\varepsilon} p_j .$$

Aus (7.2) und (7.3) folgt

(7.4) $$\sigma^2 \geq \varepsilon^2 \sum_{\left|x_j-\mu\right|\geq\varepsilon} p_j .$$

Da die Summe die Wahrscheinlichkeit angibt, dass die Zufallsvariable X mindestens um ε von ihrem Erwartungswert abweicht, $\sum_{\left|x_j-\mu\right|\geq\varepsilon} p_j = P\left(\left|X-\mu\right| \geq \varepsilon\right)$, lässt sich (7.4) in der Form

(7.5) $$\sigma^2 \geq \varepsilon^2 \cdot P\left(\left|X-\mu\right| \geq \varepsilon\right)$$

schreiben. Nach Division beider Seiten der Ungleichung (7.5) durch ε^2 und Vertauschung der Seiten erhält man hieraus unmittelbar die Tschebyscheffsche Ungleichung (7.1). □

Aufgrund der Wahrscheinlichkeit des Komplementärereignisses ist

(7.6) $$P(|X-\mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

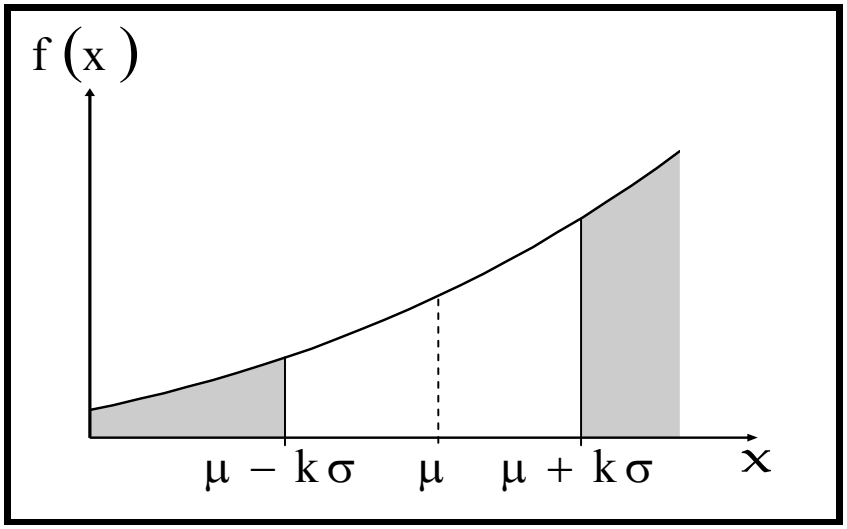
eine zu (7.1) gleichwertige Formulierung der **Tschebyscheffschen Ungleichung**. In der Form (7.6) gibt die Tschebyscheffsche Ungleichung eine **Mindestwahrscheinlichkeit** an.

Wird in (7.1) speziell $\varepsilon = k \cdot \sigma$ gesetzt, dann folgt

(7.7) $$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Die **Höchstwahrscheinlichkeit**, dass eine Zufallsvariable X um mindestens dem k-fachen ihrer Standardabweichung vom Erwartungswert abweicht, beträgt $1/k^2$.

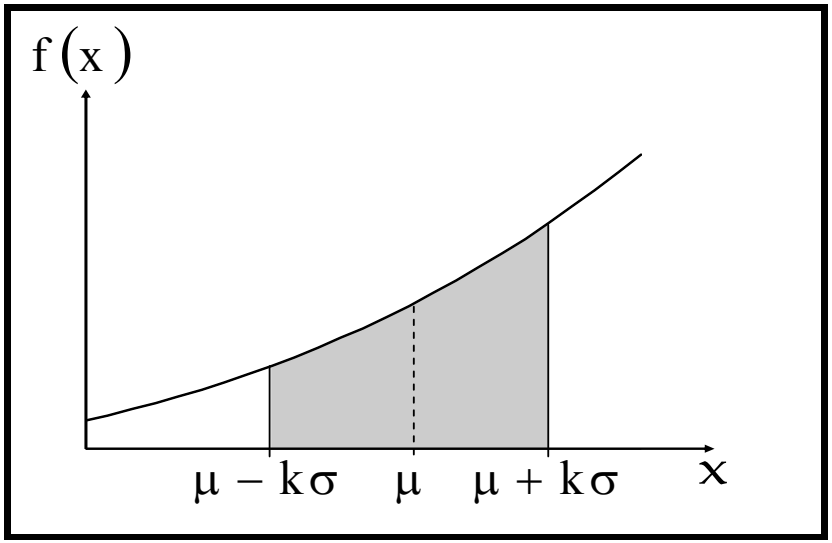
Abbildung: Höchstwahrscheinlichkeit bei ε-Umgebung als k-fache Standardabweichung (grauer Bereich)



Wird die Tschebyscheffsche Ungleichung für **Mindestwahrscheinlichkeiten** formuliert, ergibt sich

(7.8) $$P\left(\left|X-\mu\right|<k\sigma\right)\geq 1-\frac{1}{k^2}\ .$$

Abbildung: Mindestwahrscheinlichkeit bei ε-Umgebung als k-fache Standardabweichung (grauer Bereich)



Auswirkung von k auf die Höhe der Mindestwahrscheinlichkeiten

Wir berechnen die Mindestwahrscheinlichkeit, dass eine beliebig verteilte Zufallsvariable X um nicht mehr als das k = 1, 2, 3-fache ihrer Standardabweichung von ihrem Erwartungswert abweicht. Mit (7.8) erhält man:

k	Intervall	Mindestwahrscheinlichkeit
1	$(\mu - \sigma, \mu + \sigma)$	$P(\lvert X-\mu\rvert < \sigma) \geq 1 - \frac{1}{1^2} = 1 - 1 = 0$
2	$(\mu - 2\sigma, \mu + 2\sigma)$	$P(\lvert X-\mu\rvert < 2\sigma) \geq 1 - \frac{1}{2^2} = 1 - 0{,}25 = 0{,}75$
3	$(\mu - 3\sigma, \mu + 3\sigma)$	$P(\lvert X-\mu\rvert < 3\sigma) \geq 1 - \frac{1}{3^2} = 1 - 0{,}111 = 0{,}889$

- für k=1: Kein Informationsgewinn, da Wahrscheinlichkeiten stets nicht negativ sind
- für k>1: Mindestwahrscheinlichkeiten nehmen zu, wenn k steigt

Beispiel 7.1:

Der Benzinpreis ist bei gegebenem Nachfrageverhalten der Autofahrer von Einflussgrößen wie z. B. dem Rohölpreis und dem Bestand an Kraftfahrzeugen abhängig. Bei einer Mineralölfirma rechnet man für eine Planungsperiode mit einem mittleren Benzinpreis von 1,35 € pro Liter. Da Preisschwankungen nicht ausgeschlossen werden können, wird von einer Standardabweichung in Höhe von 0,12 € ausgegangen.

a) Wie groß ist mindestens die Wahrscheinlichkeit, dass der Benzinpreis innerhalb eines 2σ-Bereichs um den Erwartungswert liegen wird?

Es ist ε = 2σ. Mit der Tschebyscheffschen Ungleichung (7.8) erhält man für k=2 die Mindestwahrscheinlichkeit

$$P\left(|X-\mu| < 2\sigma\right) \geq 1-\frac{1}{k^2} = 1-\frac{1}{2^2} = 1-0{,}25 = 0{,}75.$$

b) Wie groß ist höchstens die Wahrscheinlichkeit, dass der Benzinpreis um mehr als 0,15 € vom Erwartungswert abweicht?

Hier gilt ε = 0,15. Mit der Tschebyscheffschen Ungleichung in der Formulierung (7.1) ergibt sich daraus die Höchstwahrscheinlichkeit

$$P\left(|X-\mu| \geq (\varepsilon = 0{,}15)\right) \leq \frac{\sigma^2}{\varepsilon^2} = \frac{0{,}12^2}{0{,}15^2} = \frac{0{,}0144}{0{,}0225} = 0{,}64.$$

♦

7.2 Gesetz der großen Zahlen

Das Gesetz der großen Zahlen liefert eine theoretische Begründung für die Verwendung des Stichprobenmittelwertes zur Abschätzung des unbekannten Mittelwertes einer Grundgesamtheit.

- **Schwaches Gesetz der großen Zahlen von Tschebyscheff**

gegeben: Zufallsvariable X mit
Erwartungswert $E(X) = \mu$ und Varianz $V(X) = \sigma^2$

Annahme: n-malige Wiederholung eines Zufallsvorgangs mit Zurücklegen

X_i: Zufallsvariable, die angibt, welchen Wert X bei der i-ten Durchführung des Zufallsvorgangs , i=1,2,…n, annehmen wird

x_i: Der nach der i-ten Durchführung beobachtete Wert x_i, i=1,2,…,n ist eine Realisation der Zufallsvariablen

Wenn die **Zufallsvariablen X_i** alle derselben Grundgesamtheit mit Zurücklegen entnommen werden, sind sie nicht nur **unabhängig**, sondern auch **identisch verteilt** (gleiche Wahrscheinlichkeits- bzw. Dichtefunktion). Insbesondere haben die Zufallsvariablen $X_1, X_2, \ldots, X_n$ denselben **Erwartungswert** und dieselbe **Varianz**:

(7.9) $$E(X_i) = \mu \quad \text{für alle } i=1,2,\ldots,n$$

und

(7.10) $$V(X_i) = \sigma^2 \quad \text{für alle } i=1,2,\ldots,n.$$

Stabilitätseigenschaft des arithmetischen Mittels $\overline{X}_n$

Arithmetisches Mittel der Zufallsvariablen X_i	Nach den n Wiederholungen beobachteter Durchschnittswert
(7.11) $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$	(7.12) $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$

Erwartungswert von $\overline{X}_n$:

(7.13) $$E(\overline{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu = \mu$$

Der Erwartungswert der Zufallsvariablen $\overline{X}_n$ ist gleich dem Erwartungswert der Zufallsvariablen X.

Varianz von $\overline{X}_n$:

$$V(\overline{X}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$$

(7.14) $$V(\overline{X}_n) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$ (wegen (7.10) und der Unabhängigkeit der X_i)

Die Varianz von $\overline{X}_n$ ist um den Faktor 1/n kleiner ist als die Varianz der Zufallsvariablen X.

Das **schwache Gesetzes der großen Zahlen von Tschebyscheff** besagt, dass die Wahrscheinlichkeit einer Abweichung des arithmetischen Mittels $\overline{X}_n$ vom Erwartungswert μ von weniger als ein beliebig kleiner Wert ε (ε > 0) mit wachsendem n gegen 1 strebt:

(7.15) $$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) = 1.$$

Beweis von (7.15):

Das Gesetz der großen Zahlen lässt sich leicht mit der Ungleichung von Tschebyscheff beweisen, die für eine beliebige Zufallsvariable X definiert ist. Wenn man die Tschebyscheffsche Ungleichung (7.15) für das arithmetische Mittel $\overline{X}_n$ formuliert, erhält man

(7.16) $$P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) \geq 1 - \frac{V\left(\overline{X}_n\right)}{\varepsilon^2}.$$

Wegen (7.14) geht diese Ungleichung in die Form

$$P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$$

über. Aufgrund von

$$\lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0$$

folgt für $n\to\infty$ hieraus unmittelbar das Gesetz der großen Zahlen (7.15). □

Die Folge von Zufallsvariablen $(\overline{X}_n)$ **konvergiert stochastisch** gegen μ. Diese stochastische Konvergenz bedeutet:

- nicht, dass die Abweichung $|\overline{X}_n - \mu|$ immer kleiner wird, sondern nur,
- dass es bei n gegen unendlich fast sicher ist, dass die Zufallsvariable $\overline{X}_n$ Werte in einer ε-Umgebung um μ annimmt.

Bei einer großen Anzahl von Durchführungen des Zufallsvorgangs liegt $\overline{X}_n$ daher mit hoher Wahrscheinlichkeit in dem beliebig kleinen Intervall

$$(\mu - \varepsilon,\ \mu - \varepsilon).$$

Man kann also mit einer hohen Wahrscheinlichkeit den **Erwartungswert μ** der Zufallsvariablen X **durch** eine konkrete **Realisierung der Zufallsvariablen** $\overline{X}_n$ **abschätzen**, sofern **n** nur **genügend groß** ist.

Beispiel 7.2:

Beim Lottospiel werden in jeder Ziehung 7 Zahlen (6 Zahlen und 1 Zusatzzahl) aus 49 Zahlen gezogen. Nach dem Gesetz der großen Zahlen konvergiert das arithmetische Mittel $\overline{X}_n$ mit zunehmender Zahl der Wiederholungen stochastisch gegen den Erwartungswert μ [=(m+1)/2=(49+1)/2=25 nach (6.2b)].

Wie groß muss die Zahl der Wiederholungen sein, damit mit einer Wahrscheinlichkeit von 98% gesichert ist, dass das arithmetische Mittel $\overline{X}_n$ der Lottozahlen in einer ε-Umgebung von $\left|\overline{X}_n - \mu\right| = 0{,}5$ um den Erwartungswert μ liegt?

Die gesuchte Wahrscheinlichkeit lässt sich in der für das Gesetz der großen Zahlen formulierten Tschebyscheffschen Ungleichung (7.16) bestimmen:

$$P\left(\left|\overline{X}_n - \mu\right| < \underset{\substack{\uparrow \\ \varepsilon}}{0{,}5}\right) \geq 1 - \frac{V(\overline{X})}{\varepsilon^2} = 1 - \frac{\sigma^2/n}{\varepsilon^2}.$$

Da die Lottozahlen x=1,2,…,49 gleichverteilt sind, ist ihre Varianz nach (6.3b) durch

$$V(X) = \frac{m^2 - 1}{12} = \frac{49^2 - 1}{12} = \frac{2401 - 1}{12} = 200$$

gegeben.

Damit ergibt sich

$$P(|\overline{X}_n - \mu| < 0{,}5) \geq 1 - \frac{200}{n \cdot 0{,}5^2} = 1 - \frac{200}{n \cdot 0{,}25} \overset{!}{=} 0{,}98.$$

Löst man die letzte hierin enthaltene Gleichung nach n auf, erhält man

$$1 - \frac{200}{\mathrm{n} \cdot 0{,}25} = 0{,}98 \quad | \qquad 1 - \frac{800}{\mathrm{n}} = 0{,}98$$

$$0{,}02 = \frac{800}{\mathrm{n}} \quad | \qquad 0{,}02 \cdot \mathrm{n} = 800$$

$$\Rightarrow \mathrm{n} = \frac{800}{0{,}02} = \underline{40.000}.$$

Bei 40.000 Wiederholungen des Zufallsexperiments ist also mit einer Wahrscheinlichkeit von 98% gesichert, dass das arithmetische Mittel $\overline{\mathrm{X}}_\mathrm{n}$ der Lottozahlen in einer ε-Umgebung von ±0,5 um den Erwartungswert 25 liegt. ♦

- **Schwaches Gesetz der großen Zahlen von Bernoulli**

<u>Spezialfall</u>: Die Größen $\mathbf{X_i}$ sind **unabhängige Bernoulli-verteilte Zufallsvariablen**:

$$X_i = \begin{cases} 1 \text{ falls A eintritt} \\ 0 \text{ sonst} \end{cases}$$

Summe der Bernouli-verteilten Zufallsvariable $X_1, X_2, \ldots, X_n$:

$$Y = \sum_{i=1}^{n} X_i .$$

Binomialverteilte Zufallsvariable Y: Absolute Häufigkeit für das Eintreten von A

Erwartungswert von X_i	Varianz von X_i
$E(X_i) = p$	$V(X_i) = p\cdot(1-p)$
Erwartungswert von Y	Varianz von Y
$E(Y) = n\cdot p$	$V(Y) = n\cdot p\cdot(1-p)$

(7.17) $$\overline{P}_n = \frac{1}{n} Y$$

Zufallsvariable $\overline{P}_n$: Relative Häufigkeit für das Eintreten von A (Anteilswert)

Erwartungswert von $\overline{P}_n$:

(7.18) $$E(\overline{P}_n) = E\left(\frac{1}{n}Y\right) = \frac{1}{n}E(Y) = \frac{1}{n}np = p$$

Der Erwartungswert der Zufallsvariablen $\overline{P}_n$ ist gleich der Wahrscheinlichkeit für das Eintreten des Ereignisses A.

Varianz von $\overline{P}_n$:

(7.19) $$V(\overline{P}_n) = V\left(\frac{1}{n}Y\right) = \frac{1}{n^2}V(Y) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}.$$

Die Varianz von $\overline{P}_n$ ist um den Faktor $1/n^2$ kleiner ist als die Varianz der Zufallsvariablen Y. Darin kommt die **Stabilität** der relativen Häufigkeiten mit wachsender Zahl der Wiederholungen des Zufallsvorgangs zum Ausdruck.

Das **schwache Gesetzes der großen Zahlen von Bernoulli** besagt, dass die Wahrscheinlichkeit einer Abweichung der relativen Häufigkeit $\overline{P}_n$ von der Wahrscheinlichkeit p des Ereignisses A von weniger als einem beliebig kleinen Wert ε (ε > 0) mit wachsendem n gegen 1 strebt:

(7.20) $$\lim_{n\to\infty} P\left(\left|\overline{P}_n - p\right| < \varepsilon\right) = 1.$$

Das Gesetz von Bernoulli liefert also die Begründung dafür, dass man eine unbekannte Wahrscheinlichkeit durch eine konkret ermittelte relative Häufigkeit beliebig genau abschätzen kann, sofern die Anzahl der Beobachtungen n nur groß genug ist.

Beispiel 7.3:

Eine Münze mit den Seiten Kopf und Zahl wird immer wieder geworfen. Nach dem Gesetz der großen Zahlen von Bernoulli konvergiert die relative Häufigkeit (=Anteilswert) $\overline{P}_n$ der Seite „Kopf“ (= Ereignis A) und zunehmendem n stochastisch gegen die Wahrscheinlichkeit P(A) = p = 0,5.

Wie oft muss die Münze geworfen werden, damit der Anteilswert mit einer Wahrscheinlichkeit von mindestens 96% in der ε-Umgebung (0,49; 0,51) liegen wird?

Dem Gesetz der großen Zahlen von Bernoulli (7.20) liegt die für die Zufallsvariable $\overline{P}_n$ formulierte Tschebyscheffsche Ungleichung (7.6) zugrunde:

(7.21) $$P\left(\left|\overline{P}_n - p\right| < \varepsilon\right) \geq 1 - \frac{V\left(\overline{P}_n\right)}{\varepsilon^2} = 1 - \frac{p(1-p)/n}{\varepsilon^2}.$$

Mit ε = 0,1 erhält man für die hier geforderte Mindestwahrscheinlichkeit von 96%

$$P\left(\left|\overline{P}_n - p\right| < \underset{=\varepsilon}{\underset{\uparrow}{0{,}01}}\right) \geq 1 - \frac{V\left(\overline{P}_n\right)}{\varepsilon^2} = 1 - \frac{p(1-p)/n}{\varepsilon^2}$$

$$= 1 - \frac{0{,}5 \cdot 0{,}5/n}{0{,}01^2} = 1 - \frac{0{,}25}{n \cdot 0{,}0001} \overset{!}{=} 0{,}96$$

Löst man die letzte hierin enthaltene Gleichung nach n auf, so ergibt sich

$$1-\frac{0{,}25}{n\cdot 0{,}0001}=0{,}96 \quad | \qquad 1-\frac{2500}{n}=0{,}96$$

$$0{,}04=\frac{2500}{n} \;\Rightarrow \qquad 0{,}04\cdot n=2500$$

$$\Rightarrow n=\frac{2500}{0{,}04}=\underline{62.500}.$$

Die Münze muss also 62.500-mal geworfen werden, damit die relative Häufigkeit für “Kopf” mit einer Warscheinlichkeit von mindestens 96% in dem vorgegebenen Intervall (0,49; 0,51) liegt. ♦

7.3 Zentraler Grenzwertsatz

Der **Zentrale Grenzwertsatz** macht eine Aussage über das Grenzverhalten einer Folge von Verteilungsfunktionen $\left(F_{X_n}\right)$, die zu einer gegebenen Folge von Zufallsvariablen (X_n) gehört. Für den Zentralen Grenzwertsatz gibt es unterschiedliche Verallgemeinerungsgrade und zahlreiche Varianten. Im Kern erhält man das Ergebnis, dass bei Zufallsvariablen, die als **Summen oder Durchschnitte** interpretiert werden können, bei großer Anzahl von Wiederholungen eines Zufallsvorgangs die **Normalverteilung** anwendbar ist.

- **Grenzwertsatz von Lindeberg und Lévy**

Beim Zentralen Grenzwertsatz von Lindeberg und Lévy setzt man keinen speziellen Verteilungstyp voraus. Wir betrachten eine Summe von n unabhängig beliebig verteilten Zufallsvariablen, die aus einer identischen Grundgesamtheit stammen, mit

$$E(X_i) = \mu \quad \text{und} \quad V(X_i) = \sigma^2.$$

Summenvariable: (7.18) $$Y_n = \sum_{i=1}^{n} X_i$$

Erwartungswert von Y_n	Varianz von Y_n
(7.19) $E(Y_n) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = n\mu$	(7.20) $V(Y_n) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = n\sigma^2$

Der **Grenzwertsatz von Lindeberg und Lévy** beinhaltet, dass die Folge der Verteilungsfunktionen $(F_{Y_n}) = F_{Y_1}, F_{Y_2}, F_{Y_3}, \cdots$, die zur Folge der Zufallsvariablen $(Y_n)=Y_1, Y_2, Y_3, \ldots$ gehört, sich mit wachsendem n immer besser der Verteilungsfunktion F(y) einer **Normalverteilung** mit den Parametern

$$E(Y_n) = n \cdot \mu \quad \text{und} \quad V(Y_n) = n \cdot \sigma^2$$

anpasst:

(7.21) Y_n ist bei großem n approximativ $N(n \cdot \mu, n \cdot \sigma^2)$-verteilt

Die exakte Grenzwertaussage gilt für die standardisierte Summenvariable Z_n. Nach dem **Grenzwertsatz von Lindeberg und Lévy** ist die Zufallsvariable

(7.22) $$Z_n = \frac{Y_n - n\mu}{\sigma \cdot \sqrt{n}}$$

asymptotisch standardnormalverteilt:

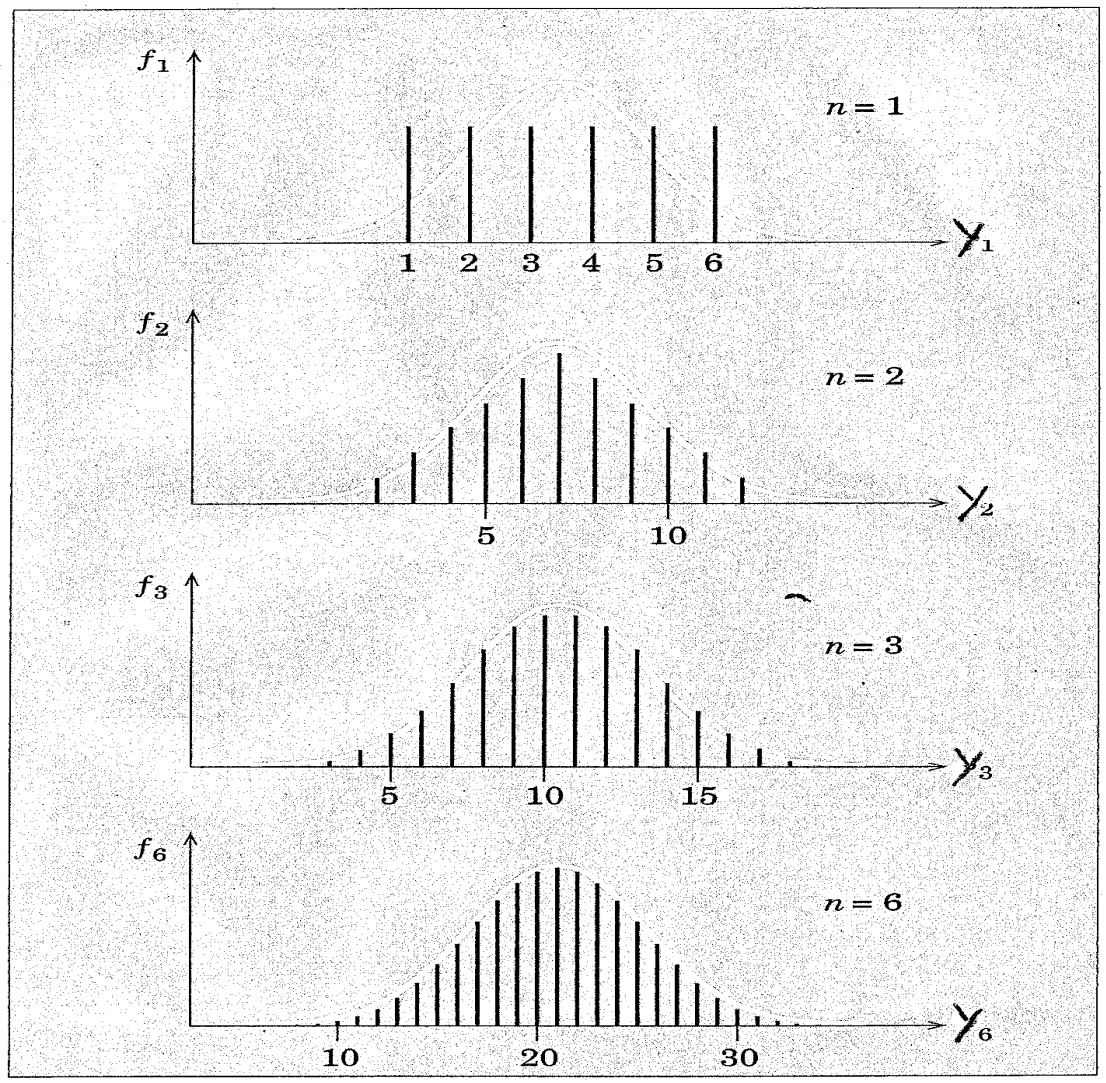
(7.23) $$Z_n \overset{a}{\sim} N(0,1)$$

In der Anwendung wird die Normalapproximation für n>30 (Faustregel) verwendet.

Beispiel zum Zentralen Grenzwertsatz

Zufallsexperiment: Würfelwurf mit n=1, 2, 3 und 6 Würfeln

Zufallsvariable Y: Summe der Augenzahlen



Beispiel 7.4:

Eine Einzelhandelskette plant, in einer Region mit 1 Mio. Haushalten ein Filialnetz zu errichten. Das Marktvolumen beträgt gegenwärtig 150 Mio. € pro Quartal, was einem durchschnittlichen Umsatz von 150 € je Haushalt entspricht. Die Standardabweichung der Haushaltsausgaben pro Quartal liegt ebenfalls bei 150 €. Eine Testfiliale, die in angemessener Zeit für 900 Haushalte erreichbar ist, hat in einem Quartal einen Gesamtumsatz von 144.000 € erzielt.

Die Marktforschungsabteilung wird damit beauftragt, einzuschätzen, welche Wahrscheinlichkeit einem in dieser Höhe erzielten oder größeren Umsatz unter unveränderten Marktbedingungen zukommt.

Die Zufallsvariable X_i bezeichnet die Ausgaben des i-ten Haushalts in €. Dann ist

$$E(X_i) = \mu = 150 \quad \text{und} \quad \sqrt{V(X_i)} = \sigma = 150$$

für alle i = 1, 2, ...,n (n = 900). Wir können annehmen, dass die X_i unabhängige und identisch verteilte Zufallsvariablen sind. Das bedeutet insbesondere, dass sich die Ausgaben unterschiedlicher Haushalte nicht gegenseitig beeinflussen. Dann gibt die Summenvariable

$$Y_n = \sum_{i=1}^{900} X_i$$

die gesamten in der Testfiliale erzielten Ausgaben der Haushalte an. Für den Erwartungswert und die Standardabweichung der Summenvariablen Y_n erhält man

$$E(Y_n) = n \cdot \mu = 900 \cdot 150 = 135.000$$

und

$$\sqrt{V(Y_n)} = \sigma \cdot \sqrt{n} = 150 \cdot \sqrt{900} = 150 \cdot 30 = 4.500.$$

Die Summe der n = 900 Haushaltsausgaben, 144.000 €, ist die konkrete Realisation der Zufallsvariablen Y_n. Gesucht wird nun die Wahrscheinlichkeit, dass Y_n Werte annimmt, die größer oder gleich 144.000 sind:

$$P(Y_n \geq 144.000).$$

Wegen der Wahrscheinlichkeit des Komplementärereignisses gilt

$$P(Y_n \geq 144.000) = 1 - P(Y_n < 144.000).$$

Da hier die Voraussetzungen des Zentralen Grenzwertsatzes von Lindeberg und Lévy erfüllt sind (n = 900 > 30), lässt sich die gesuchte Wahrscheinlichkeit approximativ über die Verteilungsfunktion der Standardnormalverteilung bestimmen:

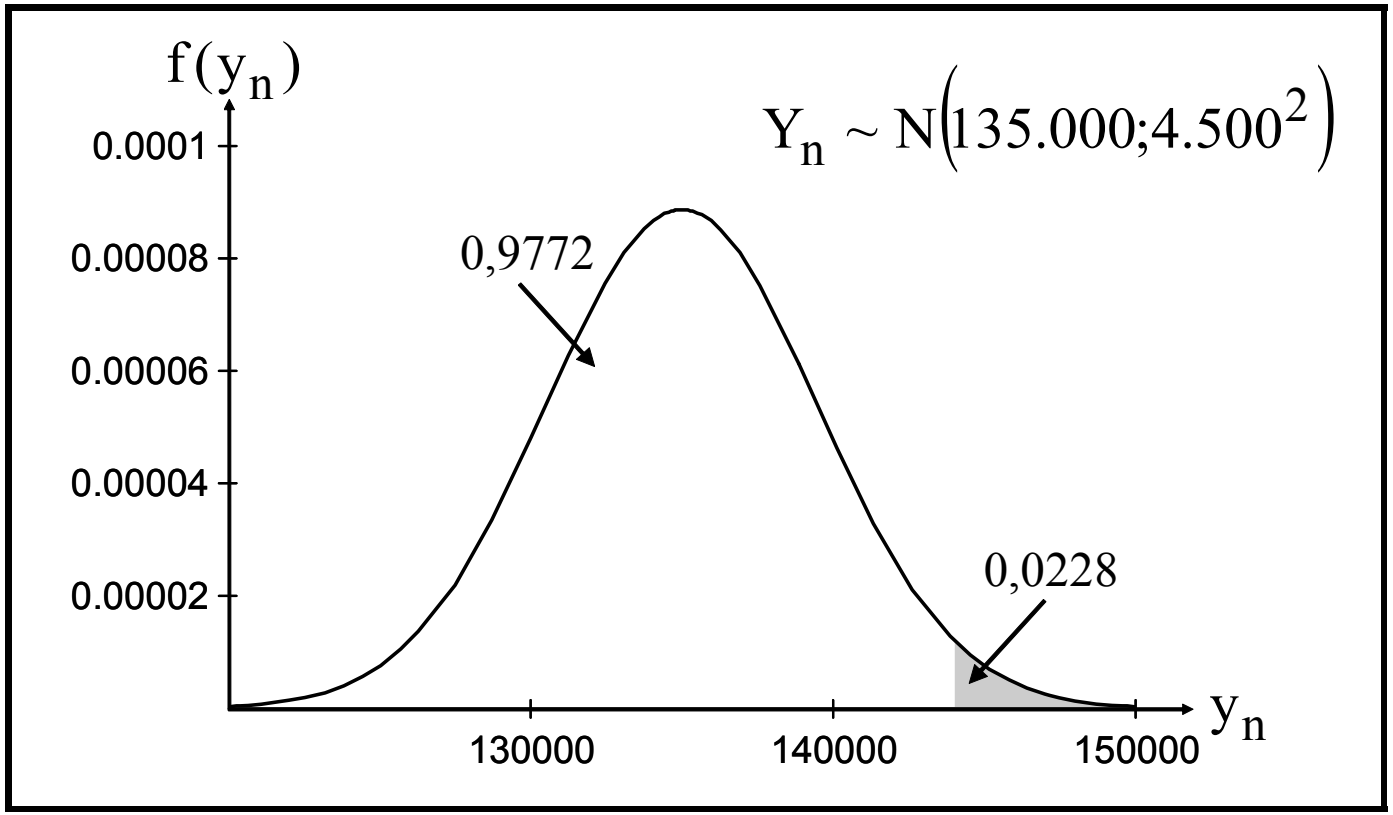
$$P(Y_n \leq 144.000) = \Phi\left(\frac{144.000 - 135.000}{4.500}\right) = \Phi(2) = 0{,}9772.$$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Die Komplementärwahrscheinlichkeit ist damit durch

$$P(Y_n \geq 144.000) = 1 - \Phi(2) = 1 - 0{,}9772 = 0{,}0228$$

gegeben. Die Wahrscheinlichkeit, bei unveränderten Marktbedingungen einen Umsatz von mindestens 144.000 € zu erzielen, liegt bei 2,3 %.



♦

Approximative Normalverteilung des arithmetischen Mittels $\overline{X}_n$

Das standardisierte arithmetische Mittel

(7.24) $$Z_n^* = \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}}$$

unterscheidet sich von der standardisierten Summenvariablen (7.22) nur um den Faktor n im Zähler und im Nenner. Der **Zentrale Grenzwertsatz von Lindeberg und Lévy** gilt daher auch für das standardisierte arithmetische Mittel (7.25):

(7.25) $$Z_n^* \overset{a}{\sim} N(0,1)$$

Das arithmetische Mittel $\overline{X}_n$ ist daher bei großen n (Faustregel: n>30) unter den Voraussetzungen des Zentralen Grenzwertsatzes von Lindeberg und Lévy approximativ normalverteilt mit den Parametern

$$E(\overline{X}_n) = \mu \quad \text{und} \quad V(\overline{X}_n) = \sigma^2 / n$$

- **Grenzwertsatz von de Moivre und Laplace**

<u>Spezialfall</u>: Die Größen $\mathbf{X_i}$ sind **unabhängige Bernoulli-verteilte Zufallsvariablen**:

$$X_i = \begin{cases} 1 \text{ falls A eintritt} \\ 0 \text{ sonst} \end{cases}$$

Die Summenvariable $Y_n = \sum_{i=1}^{n} X_i$ ist daher binomialverteilt.

Der **Grenzwertsatz von de Moivre und Laplace** beinhaltet, dass sich die Folge der Verteilungsfunktionen $(F_{Y_n}) = F_{Y_1}, F_{Y_2}, F_{Y_3}, \ldots$ mit wachsendem n immer besser der Verteilungsfunktion einer **Normalverteilung** F(y) mit den Parametern

$$E(Y_n) = n \cdot p \quad \text{und} \quad V(Y_n) = n \cdot p \cdot (1-p)$$

anpasst

(7.26) Y_n ist bei großem n approximativ $N(n \cdot p, n \cdot p \cdot (1-p))$-verteilt

Die Summe unabhängiger, Bernoulli-verteilter Zufallsvariablen X_i, d.h. die **binomialverteilte Zufallsvariable $\mathbf{Y_n}$** ist daher **approximativ normalverteilt**.

.

Die exakte Grenzwertaussage gilt auch hier wieder für die standardisierte Summenvariable Z_n. Nach dem **Grenzwertsatz von de Moivre und Laplace** ist die Zufallsvariable

(7.28) $$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}}$$

asymptotisch standardnormalverteilt:

(7.29) $$Z_n \overset{a}{\sim} N(0,1)$$

Für die Praxis benötigen wir die Information, wie groß n mindestens sein muss, damit diese **Approximation anwendbar** ist. Hierzu ist die **Faustregel**

(7.30) $$n > \frac{9}{p(1-p)}$$

nützlich.

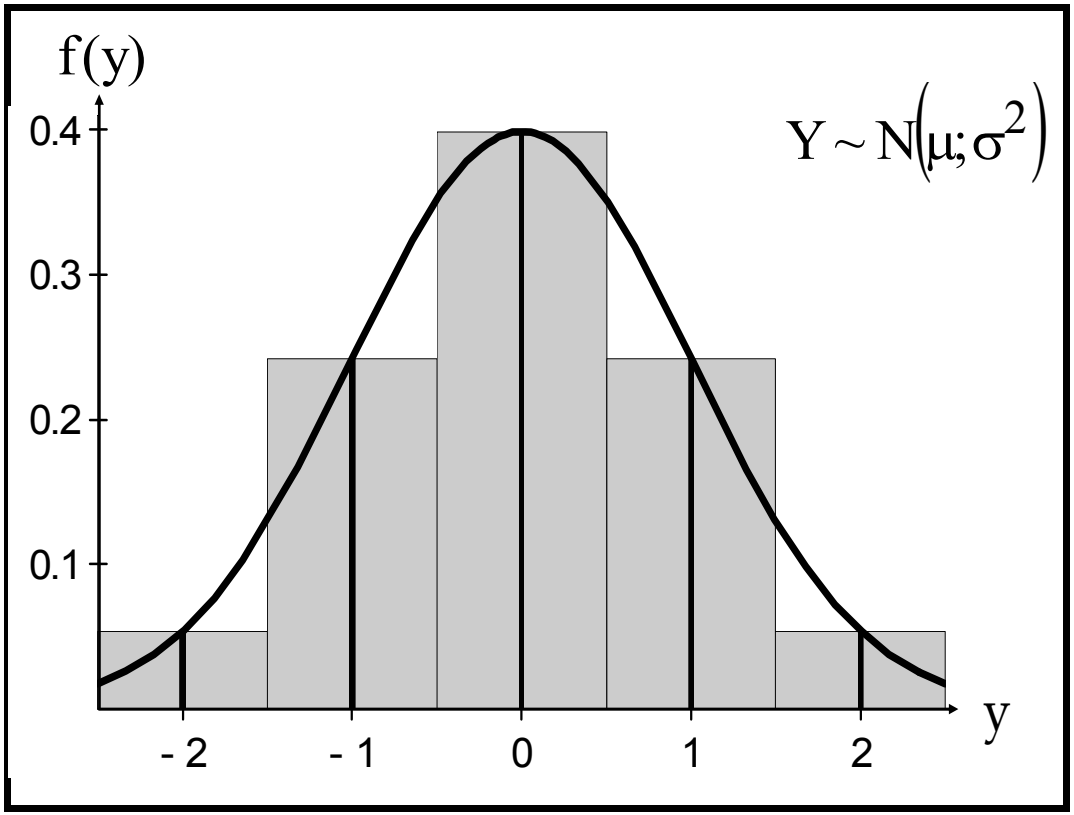
Danach muss n um so größer sein, je asymmetrischer die Binomialverteilung ausgeprägt ist:

- bei Symmetrie (p=0,5): n > 36
- bei merklicher Asymmetrie z.B. p=0,2: n > 56
- bei starker Asymmetrie z.B. p=0,1: n > 100

Stetigkeitskorrektur

Bei der Approximation der Binomial- durch die Normalverteilung nimmt man häufig noch eine **Stetigkeitskorrektur** vor, die in der unten stehenden Abbildung verdeutlicht wird. Eine binomialverteilte Zufallsvariable Y kann nur die Werte y = 0, 1, ..., n mit den Wahrscheinlichkeiten P(Y = y) annehmen. Um eine kontinuierliche Fläche zu erhalten, konstruiert man Säulen mit einer Breite von 1 um die Stäbe.

Abbildung: Visualisierung der Stetigkeitskorrektur



Wahrscheinlichkeit, dass die binomialverteilte Zufallsvariable Y einen bestimmten Wert a annimmt:

(7.31)

$$P(Y = a) = F(a + 0{,}5) - F(a - 0{,}5)$$

F: Verteilungsfunktion der Normalverteilung

(Fläche des Rechtecks über dem Intervall [a-0,5; a+0,5])

Wahrscheinlichkeit, dass eine binomialverteilte Zufallsvariable zwischen den Grenzen a und b liegt (mit Stetigkeitskorrektur):

(7.31) $$P(a \le Y \le b) = F(b + 0{,}5) - F(a - 0{,}5).$$

Da F(y) die Verteilungsfunktion einer Normalverteilung mit den Parametern

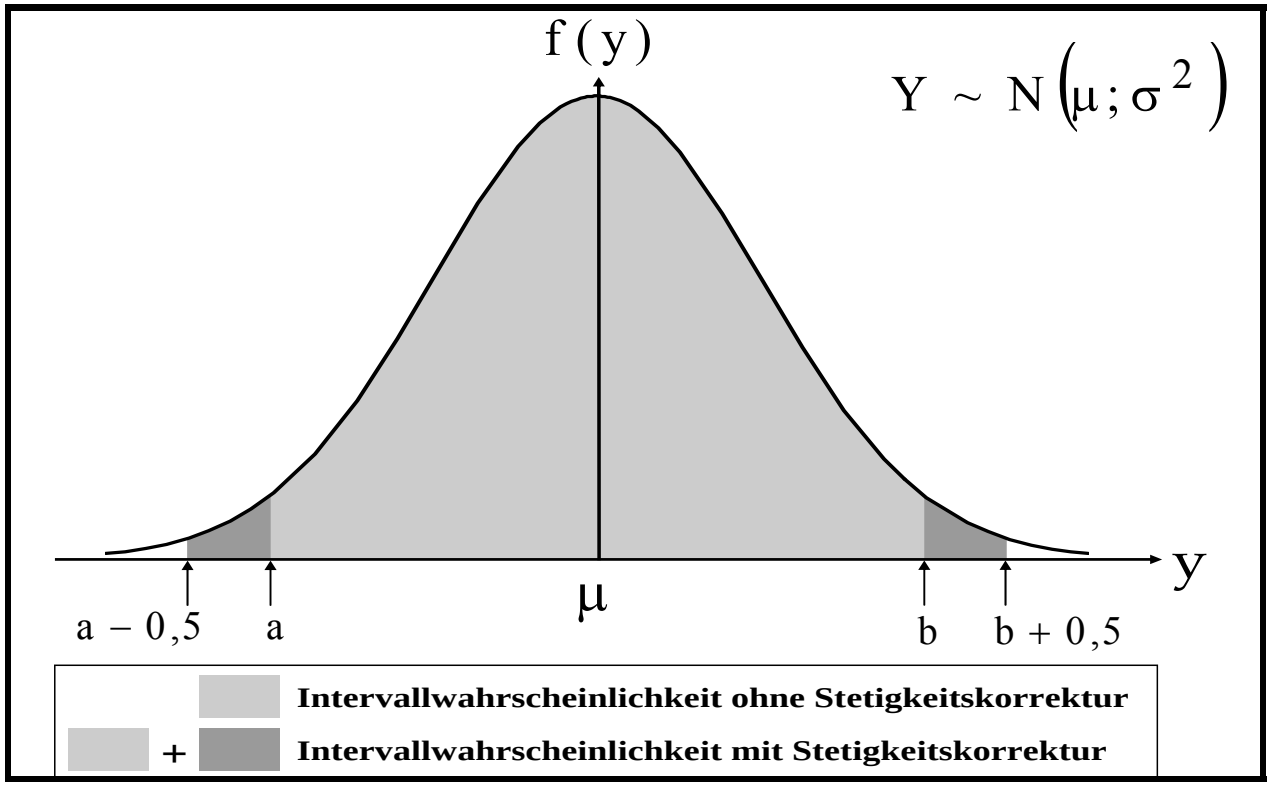
$$E(Y_n) = n \cdot p \quad \text{und} \quad V(Y_n) = n \cdot p \cdot (1-p)$$

ist, lässt sich die Intervallwahrscheinlichkeit (7.32) aus

(7.32) $$P(a \le Y \le b) = \Phi\left(\frac{b + 0{,}5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 0{,}5 - np}{\sqrt{np(1-p)}}\right)$$

.

Abbildung: Intervallwahrscheinlichkeit mit Stetigkeitskorrektur



Die Berücksichtigung von ±0,5 im Argument der Verteilungsfunktionen der Normalverteilung, F(y) und $\Phi(z)$, wird als **Stetigkeitskorrektur** bezeichnet. Je größer n ist, desto weniger wirkt sich die Stetigkeitskorrektur auf die Berechnung der Wahrscheinlichkeiten aus.

Beispiel 7.5:

Ein Betrieb liefert Glühlampen in Kartons zu je 1000 Stück. Aus früheren Untersuchungen ist bekannt, dass der Betrieb im Mittel 3 % Ausschuss produziert.

Wie groß ist die Wahrscheinlichkeit, dass sich in einem zufällig ausgewählten Karton zwischen 20 und 40 defekte Glühlampen befinden?

Es sei A das Ereignis, dass eine entnommene Glühlampe defekt ist. Wenn A eintritt, d.h. die entnommene Glühbirne defekt ist, nimmt die Zufallsvariable den Wert 1 an, sonst 0. Die Zufallsvariable Y bezeichnet die Anzahl der defekten Glühlampen in dem zufällig ausgewählten Karton:

$$Y = \sum_{i=1}^{1000} X_i$$

Dann wissen wir, daß Y binomialverteilt ist mit den Parametern n=1000 und p=0,03. Die gesuchte Wahrscheinlichkeit dafür, dass Y in das Intervall [20,40] fällt, ergibt sich dann aus

$$P(20 \le Y \le 40) = \sum_{y=20}^{40} \binom{1000}{y} \cdot 0{,}03^{y} \cdot 0{,}97^{1000-y}$$

Eine Berechnung der Wahrscheinlichkeit mit dieser Formel ist jedoch zu umständlich. Nach dem Zentralen Grenzwertsatz von de Moivre und Laplace gilt jedoch, dass die standardisierte Zufallsvariable

$$Z = \frac{Y - np}{\sqrt{np(1-p)}}$$

für große n näherungsweise standardnormalverteilt ist. Da die Faustregel (7.30)

$$(1.000 =)n > \frac{9}{p \cdot (1-p)} \left(= \frac{9}{0{,}03 \cdot (1-0{,}03)} = 309{,}3 \right)$$

erfüllt ist, können wir die gesuchte Wahrscheinlichkeit mit der Normalverteilung bestimmen. Mit den Parametern

$$E(Y) = n \cdot p = 1000 \cdot 0{,}03 = 30$$

und

$$V(Y) = n \cdot p(1-p) = 1000 \cdot 0{,}03 \cdot 0{,}97 = 29{,}1$$

erhalten wir

$$P(20 \le Y \le 40) = \Phi\left(\frac{40-30}{\sqrt{29{,}1}}\right) - \Phi\left(\frac{20-30}{\sqrt{29{,}1}}\right) = \Phi(1{,}85) - \Phi(-1{,}85).$$

Aufgrund der Symmetrie der Standardnormalverteilung ist

$$\Phi(-1{,}85) = 1 - \Phi(1{,}85),$$

so dass wir nach Ablesen des Wertes aus der Verteilungstabelle der Standardnormalverteilung,

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	$\boxed{0,9678}$	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767

die gesuchte Wahrscheinlichkeit erhalten:

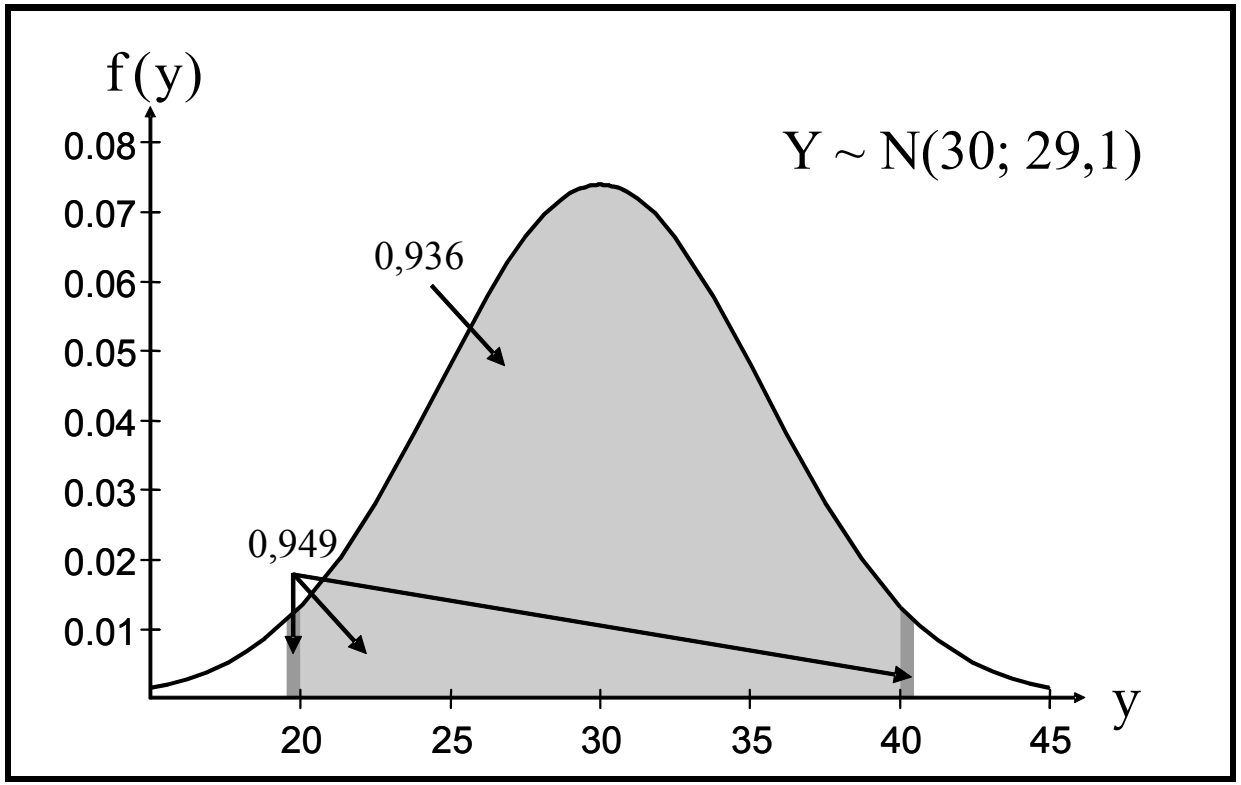
$$\begin{aligned} P(20 \le Y \le 40) &= \Phi(1{,}85) - \Phi(-1{,}85) = \Phi(1{,}85) - [1 - \Phi(1{,}85)] \\ &= 0{,}9678 - (1 - 0{,}9678) = 0{,}9356. \end{aligned}$$

Danach beträgt die Wahrscheinlichkeit, dass in einem zufällig ausgewählten Karton von 1000 Glühlampen zwischen 20 und 40 Ausschussstücke vorhanden sind, 93,6%.

Allerdings haben wir dieses Ergebnis ermittelt, ohne zu berücksichtigen, dass bei der Approximation der Binomial- durch die Normalverteilung eine diskrete durch eine stetige Verteilung angenähert wird. Wenn die gesuchte Wahrscheinlichkeit unter Berücksichtigung der **Stetigkeitskorrektur** berechnet wird, ergibt sich

$$\begin{aligned} P(20 \le Y \le 40) &= \Phi\left(\frac{40 + 0{,}5 - 30}{\sqrt{29{,}1}}\right) - \Phi\left(\frac{20 - 0{,}5 - 30}{\sqrt{29{,}1}}\right) \\ &= \Phi(1{,}95) - \Phi(-1{,}95) = \Phi(1{,}95) - [1 - \Phi(1{,}95)] \\ &= 0{,}9744 - (1 - 0{,}9744) = 0{,}9488. \end{aligned}$$

Wenn also die Stetigkeitskorrektur vorgenommen wird, erhöht sich die Wahrscheinlichkeit um 1,3 Prozentpunkte.



♦

Approximative Normalverteilung des Anteilwerts $\overline{P}_n$

Der standardisierte Anteilswert (relative Häufigkeit)

(7.33) $$Z_n^* = \frac{\overline{P}_n - p}{\sqrt{\frac{p(1-p)}{n}}}$$

unterscheidet sich von der standardisierten Summenvariablen (7.28) nur um den Faktor n im Zähler und im Nenner. Der **Zentrale Grenzwertsatz von de Moivre und Laplace** gilt daher auch für den standardisierten Anteilswert (7.33):

(7.34) $$Z_n^* \overset{a}{\sim} N(0,1)$$

Der Anteilswert (relative Häufigkeit) $\overline{P}_n$ ist daher bei großen n (Faustregel: (7.30)) unter den VorausSetzungen des Zentralen Grenzwertsatzes von de Moivre und Laplace approximativ normalverteilt mit den Parametern

$$E(\overline{P}_n) = p \qquad \text{und} \qquad V(\overline{P}_n) = \frac{p(1-p)}{n}$$

- **Weitere Grenzwertsätze**

Für den Zentralen Grenzwertsatz von Lindeberg und Lévy gibt es Verallgemeinerungen.

Grenzwertsatz von Ljapunoff:

Gibt hinreichende Bedingungen an, unter denen der Zentrale Grenzwertsatz von Lindeberg und Lévy auch im Fall **unabhängiger, aber nicht notwendig identisch verteilter Zufallsvariablen** gilt.

Grenzwersatz von Lindeberg-Feller:

Gibt für den beim Grenzwertsatz von Ljapunoff genannten Fall darüber hinaus die notwendigen Bedingungen an.

Hierbei zeigt sich, dass unter recht allgemeinen Bedingungen die Summe unabhängiger und beliebig (nicht notwendig identisch) verteilter Zufallsvariablen stets asymptotisch normalverteilt ist. Insbesondere lässt sich aus dieser Verallgemeinerung zeigen, dass eine Summenvariable Y, die als additive Überlagerung vieler kleiner unabhängiger Zufallseinflüsse mit relativ kleinen Varianzen interpretierbar ist, asymptotisch einer Normalverteilung folgt. Geringfügige Abweichungen von der Unabhängigkeitsannahme schränken die Anwendbarkeit nicht ein.