Stochastik

Barth, Friedrich

München, [20]03

14. 8. Anwendungen der Ungleichung von Bienaymé-Tschebyschow

unsinnig anzunehmen, daß eine startende Kugel weiß, in welchem Fach gerade Defizit herrscht, um bevorzugt dorthin zu springen.
Unterstellt man dem schwachen Gesetz der großen Zahlen also einen Buchhaltercharakter, so müßte man wider alle Vernunft annehmen, daß stochastische Geräte Gewissen und Gedächtnis hätten, wie es *Joseph Bertrand* (1822–1900) einmal treffend formulierte*. Wäre dem so, entgegnete 1785 *Leonhard Euler* (1707 bis 1783) in seinen *Opuscula Analytica*** der Auffassung *d'Alemberts*,

»dann müßte jeder nach einem Jahr, ja nach einem Jahrhundert stattfindende Zug vom Ergebnis aller Züge abhängen, die seit undenklichen Zeiten an irgendwelchen Orten dieser Erde stattgefunden haben; Absurderes kann sicherlich kaum gedacht werden.«

14.8. Anwendungen der Ungleichung von *Bienaymé-Tschebyschow*

Die Ungleichung von *Bienaymé-Tschebyschow* kann, je nach Bedarf, unterschiedlich formuliert werden. Wir stellen die drei häufigsten Formulierungen der *Bienaymé-Tschebyschow*-Ungleichung in der Form, in der sie sich am leichtesten merken lassen, zusammen:

1) Ist X eine Zufallsgröße mit $\mathscr{E}X = \mu$ und ist $a > 0$, dann gilt

$$P(|X - \mu| \geqq a) \leqq \frac{\operatorname{Var} X}{a^2}. \qquad \text{(Satz 184.1)}$$

2) Ist H_n die relative Häufigkeit eines Ereignisses mit der Wahrscheinlichkeit p in einer *Bernoulli*-Kette der Länge n und ist $\varepsilon > 0$, dann gilt

$$P(|H_n - p| \geqq \varepsilon) \leqq \frac{pq}{n\varepsilon^2} \leqq \frac{1}{4n\varepsilon^2}. \qquad \text{(Satz 248.1)}$$

3) Ist $\bar{X}_n$ das arithmetische Mittel n gleichverteilter, paarweise unabhängiger Zufallsgrößen X_i mit $\mathscr{E}X_i = \mu$ und $\operatorname{Var} X_i = \sigma^2$ und ist $a > 0$, dann gilt

$$P(|\bar{X}_n - \mu| \geqq a) \leqq \frac{\sigma^2}{na^2}. \qquad \text{(Aufgabe 271/71)}$$

Viele Aufgaben der Wahrscheinlichkeitsrechnung handeln davon, daß das wahre Risiko, d.h., daß die Wahrscheinlichkeit dafür, daß die Werte einer Zufallsgröße X von ihrem Erwartungswert μ um mindestens a abweichen, eine gewisse Schranke η nicht überschreiten soll, kurz, daß

$$P(|X - \mu| \geqq a) \leqq \eta \tag{1}$$

sein soll. Anders ausgedrückt: Die Wahrscheinlichkeit, daß die Werte von X sich um weniger als a von μ unterscheiden, soll einen gewissen Mindestwert besitzen, d.h.,

* »On fait trop d'honneur à la roulette: elle n'a ni conscience ni mémoire.« (*Calcul des Probabilités*, p. XXII, 1889)
** Die Abhandlung lautet *Solutio quarundam quaestionum difficiliorum in Calculo Probabilium*. – *Friedrich II.* bat *Euler* 1749 und 1763 um Rat bezüglich der Errichtung von Lotterien, um die Finanznot seines Staates zu beheben. Aus der Beschäftigung mit diesem Problem entstanden *Eulers* wahrscheinlichkeitstheoretische Arbeiten.

$$P(|X - \mu| < a) \geqq 1 - \eta. \quad (2)$$

Da man nun auf Grund von Satz 184.1 weiß, daß das wahre Risiko höchstens so groß wie das *Tschebyschow*-Risiko r_T ist, ist Bedingung (1) für das wahre Risiko sicher erfüllt, wenn man das *Tschebyschow*-Risiko r_T höchstens so groß wie die Schranke η werden läßt, also (meist) weniger fordert, nämlich

$$P(|X - \mu| \geqq a) \leqq r_T \leqq \eta.$$

Es ist uns natürlich bewußt, daß man dadurch unter Umständen viel zu grobe Abschätzungen erhält. Wo möglich, wird man außerdem versuchen, mit $r_T = \eta$ auszukommen.

Nun zu den Aufgaben! Der einfachste Aufgabentyp ist derjenige, bei dem aus gegebenen Daten eine Schranke für das wahre Risiko gesucht wird.

Beispiel 1: Wie groß ist die Mindestwahrscheinlichkeit dafür, daß die relative Häufigkeit für eine Sechs beim 100fachen Wurf eines L-Würfels um weniger als 0,05 von der Wahrscheinlichkeit für eine Sechs abweicht?

Lösung: An sich könnte man die gesuchte Wahrscheinlichkeit direkt berechnen. Mit $X :=$ »Anzahl der Sechsen bei 100 Würfen« erhalten wir

$$\begin{aligned} P(|H_{100} - \tfrac{1}{6}| < 0{,}05) &= P(|\tfrac{X}{100} - \tfrac{1}{6}| < \tfrac{1}{20}) = P(|X - \tfrac{100}{6}| < 5) = \\ &= P(11\tfrac{2}{3} < X < 21\tfrac{2}{3}) = \\ &= \sum_{k=12}^{21} B(100; \tfrac{1}{6}; k) = F_{1/6}^{100}(21) - F_{1/6}^{100}(11) = \\ &= 0{,}89982 - 0{,}07772 = 0{,}82210. \end{aligned}$$

Hätten wir keine Tabellen, z.B. wenn $n = 80$ wäre, so müßten wir eine sehr mühsame Rechnung durchführen. Da ist man dann oft froh, wenn man die gesuchte Wahrscheinlichkeit durch eine untere Schranke abschätzen kann. Wir suchen nun also eine untere Schranke für $P(|H_{100} - \frac{1}{6}| < 0{,}05)$. Dazu gehen wir zum Gegenereignis über und suchen eine obere Schranke für $P(|H_{100} - \frac{1}{6}| \geqq 0{,}05)$.

Das *Tschebyschow*-Risiko $r_T = \dfrac{\frac{1}{6} \cdot \frac{5}{6}}{100 \cdot 0{,}05^2}$ ist eine solche obere Schranke. Wir erhalten $r_T = \frac{5}{9} < 0{,}556$. Also ist

$P(|H_{100} - \frac{1}{6}| < 0{,}05) \geqq 1 - \frac{5}{9} = \frac{4}{9} > 44{,}4\%$.

Das bedeutet:

Mit einer Wahrscheinlichkeit von mehr als 44,4% liegen beim 100fachen Wurf eines L-Würfels die Werte h_{100} (»Sechs«) der relativen Häufigkeit H_{100} (»Sechs«) im Intervall $]\frac{1}{6} - 0{,}05; \frac{1}{6} + 0{,}05[=]\frac{7}{60}; \frac{13}{60}[$, was durch Figur 254.1 veranschaulicht wird.

In einer Vielzahl von Aufgaben wird nach der Zahl n der Versuche gefragt, die nötig sind, um das wahre Risiko nicht größer als η werden zu lassen.

Beispiel 2: Wie oft muß ein L-Würfel mindestens geworfen werden, damit mit einer Sicherheit von mindestens 60% das arithmetische Mittel der Augenzahlen um weniger als 0,25 vom Erwartungswert 3,5 abweicht?

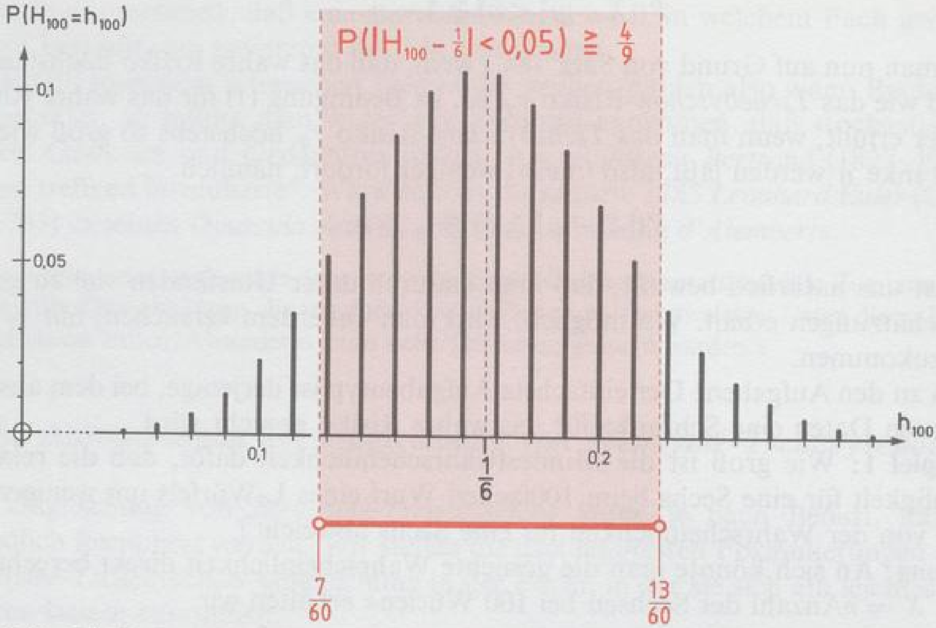

Fig. 254.1 Die Wahrscheinlichkeit, daß beim 100maligem Werfen eines L-Würfels die relative Häufigkeit der Sechs um weniger als $\frac{1}{20}$ von ihrer Wahrscheinlichkeit $\frac{1}{6}$ abweicht, ist mindestens $\frac{4}{9}$.

Lösung: Gesucht ist ein kleinstes n, so daß $P(|\bar{X}_n - 3{,}5| < 0{,}25) \geqq 60\% = 1 - \eta$. Da die Varianz der Zufallsgröße Augenzahl den Wert $\frac{35}{12}$ hat (Aufgabe 194/**44**), erhalten wir aus der *Tschebyschow*-Ungleichung

$$P(|\bar{X}_n - 3{,}5| \geqq 0{,}25) \leqq \frac{\frac{35}{12}}{n \cdot 0{,}25^2}.$$

Setzen wir das rechts stehende *Tschebyschow*-Risiko höchstens gleich der Schranke $\eta (= 40\%)$, dann gewinnen wir für n die folgende Abschätzung

$$\frac{35}{12 \cdot 0{,}25^2 \cdot n} \leqq 0{,}4 \Leftrightarrow n \geqq \tfrac{350}{3} = 116\tfrac{2}{3}, \qquad \text{also } n \geqq 117.$$

Somit gilt: Wirft man mindestens 117mal einen L-Würfel, so ist die Wahrscheinlichkeit dafür, daß das arithmetische Mittel der Augenzahlen vom Erwartungswert 3,5 um weniger als 0,25 abweicht, mindestens 60%, was Figur 255.1 veranschaulichen soll.

Schwieriger als diese beiden Aufgabentypen sind diejenigen, in denen ε- bzw. a-Intervalle gesucht sind. Dabei sind zwei Fragestellungen zu unterscheiden.

1. Fragestellung: Es ist dasjenige Intervall um p (bzw. μ) gesucht, in das die relative Häufigkeit H_n (bzw. das arithmetische Mittel $\bar{X}$) mit einer vorgegebenen Sicherheitswahrscheinlichkeit von mindestens $1 - \eta$ trifft. Man sucht also ein ε, so daß die Bedingung

$$|H_n - p| < \varepsilon \Leftrightarrow p - \varepsilon < H_n < p + \varepsilon$$

mit einer vorgegebenen Mindestwahrscheinlichkeit $1 - \eta$ erfüllt wird.

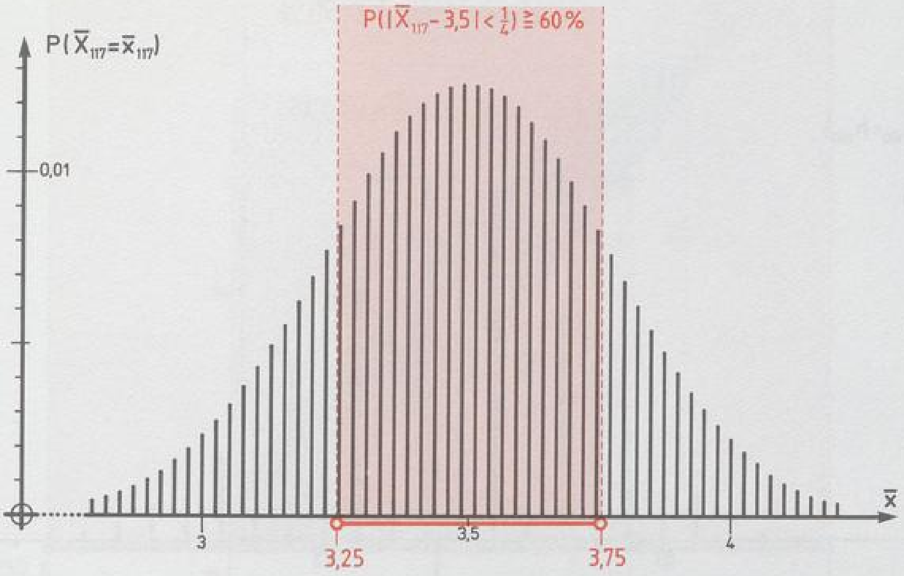

Fig. 255.1 Soll die Wahrscheinlichkeit, daß das arithmetische Mittel der Augenzahlen eines L-Würfels vom Erwartungswert 3,5 um weniger als $\frac{1}{4}$ abweicht, mindestens 60% betragen, so muß mindestens 117mal gewürfelt werden.
Gezeichnet ist vom Stabdiagramm der Wahrscheinlichkeitsverteilung $P(\bar{X}_{117} = \bar{x})$ nur jeder dritte der 586 Stäbe (die bei $\bar{x} \in \{1, \frac{118}{117}, \frac{119}{117}, \ldots, 6\}$ liegen), sofern er mindestens $5 \cdot 10^{-5}$ mißt.

Beispiel 3: In welchem Intervall um $p = \frac{1}{6}$ liegt bei 100maligem Werfen eines L-Würfels die relative Häufigkeit für die Augenzahl 6 mit einer Mindestwahrscheinlichkeit von 60%?
Lösung: Gesucht ist ein ε, so daß

$$P(|H_{100} - \tfrac{1}{6}| < \varepsilon) = P(\tfrac{1}{6} - \varepsilon < H_{100} < \tfrac{1}{6} + \varepsilon) \geqq 60\%$$

wird. Statt dessen können wir auch

$$P(|H_{100} - \tfrac{1}{6}| \geqq \varepsilon) \leqq 40\%$$

fordern. Das ist sicher erfüllt, wenn das *Tschebyschow*-Risiko höchstens 40% wird, also

$$\frac{\frac{1}{6} \cdot \frac{5}{6}}{100\,\varepsilon^2} \leqq 0{,}4 \Leftrightarrow \varepsilon \geqq \tfrac{1}{24}\sqrt{2} = 0{,}0589\ldots$$

Für $\varepsilon = 0{,}059$ ist die Bedingung sicherlich erfüllt, d.h., mit einer Wahrscheinlichkeit von mindestens 60% ergeben sich Werte h_{100}(»6«) der relativen Häufigkeit H_{100}(»6«) zwischen 0,107 und 0,226. Figur 256.1 veranschaulicht diesen Sachverhalt. – Bedenkt man noch, daß H_{100} nur Werte aus $\{0, \frac{1}{100}, \frac{2}{100}, \ldots, \frac{99}{100}, 1\}$ annehmen kann, so läßt sich verschärfend sagen, daß mit einer Wahrscheinlichkeit von mindestens 60% die relative Häufigkeit H_{100}(»Sechs«) Werte im Intervall [0,11; 0,22] annimmt.

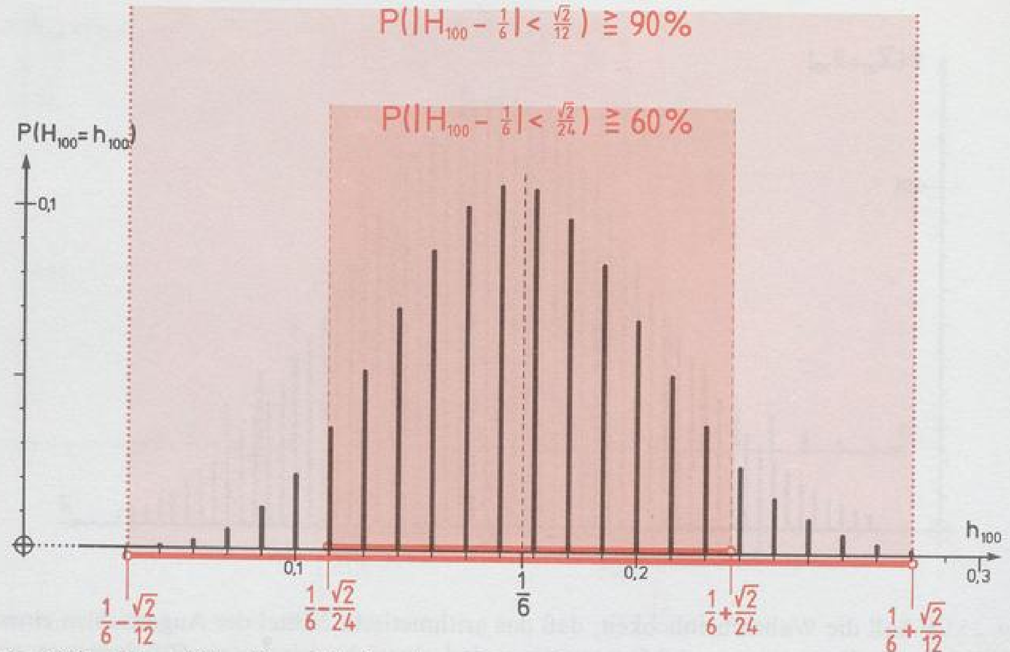

Fig. 256.1 Beim 100maligen Werfen eines L-Würfels ist die Wahrscheinlichkeit mindestens 60%, daß die relative Häufigkeit des Ereignisses »Sechs« von seiner Wahrscheinlichkeit $\frac{1}{6}$ um weniger als $\varepsilon = \frac{1}{24}\sqrt{2}$ abweicht. – Punktiert ist dasjenige ε-Intervall angegeben, das man wählen muß, falls man eine Wahrscheinlichkeit von mindestens 90% fordert.

Die Aufgabenstellung von Beispiel 3 lautet allgemein $P(|H_n - p| < \varepsilon) \geqq 1 - \eta$ bzw. $P(|H_n - p| \geqq \varepsilon) \leqq \eta$. Mit dem Ansatz $r_T = \eta$, also $\frac{pq}{n\varepsilon^2} = \eta$, erhält man

$$\varepsilon = \sqrt{\frac{pq}{n\eta}}, \text{ was zum Intervall } \quad I(p) := \left]p - \sqrt{\frac{pq}{n\eta}};\ p + \sqrt{\frac{pq}{n\eta}}\right[$$

führt. Es wird also jedem p ein Intervall $I(p)$ zugeordnet, in das die Werte h_n der relativen Häufigkeit H_n mindestens mit der Wahrscheinlichkeit $1 - \eta$ hineinfallen. Figur 257.1 veranschaulicht diesen Zusammenhang $p \mapsto I(p)$. Die Hüllkurve all dieser Intervalle ist eine Ellipse mit der Gleichung

$$|h_n - p| = \sqrt{\frac{p(1-p)}{n\eta}}.$$

2. Fragestellung: Der andere Fall der Intervallbestimmung besteht darin, daß man bei einer Versuchsserie der Länge n einen Wert h_n der relativen Häufigkeit H_n ermittelt hat und nun ein ε-Intervall um diesen Wert h_n angeben möchte, von dem man mit einer vorgegebenen Mindestwahrscheinlichkeit sagen kann, daß es die unbekannte, aber feste Wahrscheinlichkeit p enthält. Solche Zufallsintervalle nannte 1934 *Jerzy Neyman** (1894–1981) **Vertrauensintervall** oder **Konfidenzintervall** für p.

* gesprochen jɛʒɨ nɛjman

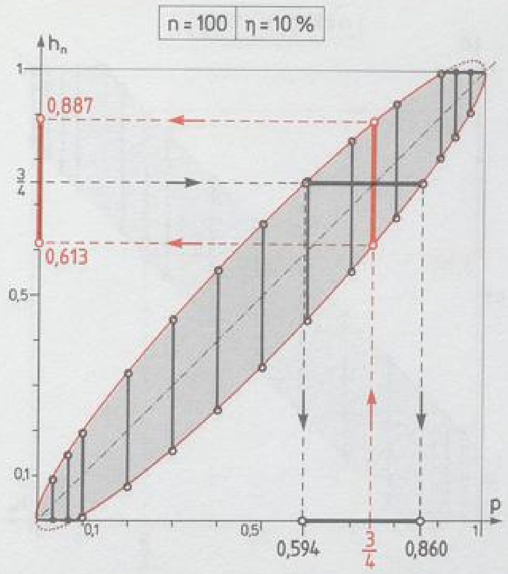

Fig. 257.1
Der Graph der Relation $p \mapsto I(p)$ ist die Punktmenge $\{(p|h_n) | h_n \in I(p) \cap [0; 1] \wedge p \in [0; 1]\}$, also das grau unterlegte Gebiet einschließlich des schwarzen Randes.
Für $p = \frac{3}{4}$ ist $I(\frac{3}{4}) =]\frac{3}{4} - \frac{1}{40}\sqrt{30}; \frac{3}{4} + \frac{1}{40}\sqrt{30}[\subset]0{,}613; 0{,}887[$ rot hervorgehoben.
In dieses Intervall fällt die relative Trefferhäufigkeit mindestens mit der Wahrscheinlichkeit 90%, wenn P(»Treffer«) $= \frac{3}{4}$ ist.
---: Geht man bei $h_n = \frac{3}{4}$ ein, so erhält man das zugehörige echte Konfidenzintervall auf der p-Achse (vgl. Figur 260.1).

Man konstruiert dazu *vor* der Ausführung des Zufallsexperiments ein möglichst enges Zufallsintervall $]H_n - \varepsilon; H_n + \varepsilon[$, das die unbekannte, aber feste Wahrscheinlichkeit p mindestens mit der Wahrscheinlichkeit $1 - \eta$ überdeckt, für das also $P(|H_n - p| < \varepsilon) = P(H_n - \varepsilon < p < H_n + \varepsilon) \geqq 1 - \eta$ gilt.
Bei der 1. Fragestellung lag das ε-Intervall um den bekannten Wert p fest. Der Zufall steckte im Hineintreffen der relativen Häufigkeit H_n in dieses Intervall. Bei der 2. Fragestellung ist zwar auch p fest, aber nicht bekannt. Der Zufall bestimmt jetzt den Wert h_n der relativen Häufigkeit H_n und damit mindestens mit der Wahrscheinlichkeit $1 - \eta$ das ε-Intervall um h_n, das so auf der Zahlengeraden liegt, daß es den gesuchten p-Wert überdeckt. Dabei hängt der Radius ε natürlich von η ab. (Das Verfahren ähnelt also dem Jagen einer Fliege mit einer Fliegenklatsche: Die Fliege ist das p, die Klatsche das ε-Intervall, die Klatschenmitte trifft zufallsgesteuert bei jedem Schlag auf das jeweilige h_n.)

Beispiel 4: Die ersten 100 Würfe von Tabelle 10.1 ergaben $h_{100}(\{6\}) = 0{,}18$. Für welches Intervall kann man mit einer Sicherheit von mindestens 90% schließen, daß es die Wahrscheinlichkeit p für eine Sechs enthält?

Lösung: Gesucht ist ein ε, so daß $P(H_{100}(\{6\}) - \varepsilon < p < H_{100}(\{6\}) + \varepsilon) \geqq 90\%$ wird. Dazu betrachten wir wieder das Gegenereignis, also

$P(|H_{100} - p| \geqq \varepsilon) \leqq 10\%$,

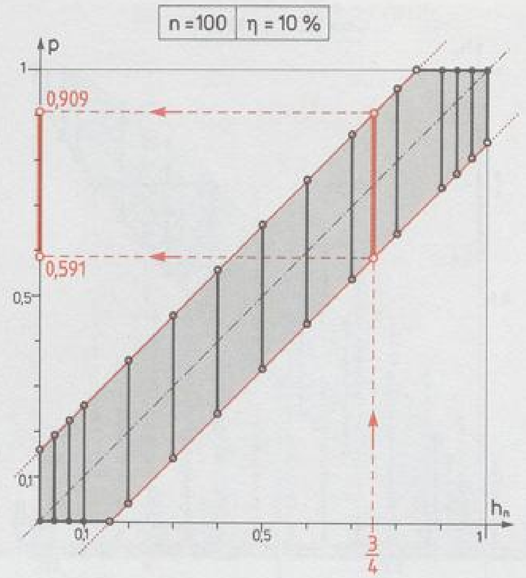

Fig. 258.1 Grobe Konfidenzintervalle. Der Graph der Relation $h_n \mapsto I(h_n)$ ist die Punktmenge $\{(h_n|p)|p \in I(h_n) \cap [0;1] \wedge h_n \in [0;1]\}$, also das grau unterlegte Gebiet einschließlich des schwarzen Randes. Für $h_{100} = \frac{3}{4}$ ist $I(\frac{3}{4}) =]\frac{3}{4} - \frac{1}{20}\sqrt{10}; \frac{3}{4} + \frac{1}{20}\sqrt{10}[\subset]0{,}591; 0{,}909[$ rot hervorgehoben. Man kann mit einer Sicherheit von mindestens 90% darauf vertrauen, daß dieses Intervall die Wahrscheinlichkeit $p = P$(»Treffer«) überdeckt, wenn die relative Häufigkeit des Treffers zu $h_{100} = \frac{3}{4}$ gemessen wurde.

was sicherlich erfüllt ist, wenn $\frac{pq}{100\,\varepsilon^2} \leqq \frac{1}{400\,\varepsilon^2} \leqq 10\%$.

Wir erhalten $\varepsilon \geqq \frac{1}{20}\sqrt{10} = 0{,}158\ldots$ Zum Zufallsergebnis $h_{100}(\{6\}) = 0{,}18$ gehört also das Intervall $]0{,}021; 0{,}339[$, von dem wir sagen können, es wurde auf Grund eines Verfahrens erhalten, das mit einer Wahrscheinlichkeit von mindestens 90% zu einem Intervall führt, das die wahre Wahrscheinlichkeit p für die Augenzahl 6 bei diesem Würfel enthält.

Löst man die Aufgabenstellung von Beispiel 4 allgemein mit dem Ansatz $P(|H_n - p| \geqq \varepsilon) \leqq \frac{pq}{n\varepsilon^2} \leqq \frac{1}{4n\varepsilon^2} = \eta$, so erhält man $\varepsilon = \frac{1}{2\sqrt{n\eta}}$ und damit das **grobe Konfidenzintervall** $I(h_n) = \left]h_n - \frac{1}{2\sqrt{n\eta}}; h_n + \frac{1}{2\sqrt{n\eta}}\right[$.

Es wird also jedem Wert h_n ein Intervall $I(h_n)$ zugeordnet, das den unbekannten Wert p mindestens mit der Wahrscheinlichkeit $1 - \eta$ enthält. Figur 258.1 veranschaulicht diesen Zusammenhang $h_n \mapsto I(h_n)$. Die Hüllkurve dieser groben Konfidenzintervalle ist ein Parallelenpaar mit der Gleichung $|p - h_n| = \frac{1}{2\sqrt{n\eta}}$.

»Genauere« Näherung. Weil p unbekannt ist, mußten wir den Ausdruck pq aus r_T durch den Wert $\frac{1}{4}$ abschätzen. Kennte man p, so wäre für $p \neq \frac{1}{2}$ eine genauere ε-Bestimmung durch $\frac{pq}{n\varepsilon^2} = \eta$ möglich. Man erhielte $\varepsilon = \sqrt{\frac{pq}{n\eta}}$. Nach dem

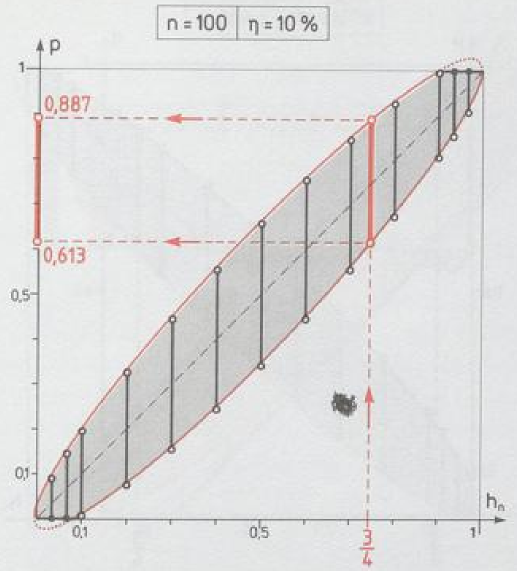

Fig. 259.1 Näherungskonfidenzintervalle. Der Graph der Relation $h_n \mapsto I(h_n)$ ist die Punktmenge $\{(h_n|p) | p \in \tilde{I}(h_n) \cap [0; 1] \wedge h_n \in [0; 1]\}$, also das grau unterlegte Gebiet einschließlich des schwarzen Randes. Für $h_{100} = \frac{3}{4}$ ist $\tilde{I}(\frac{3}{4}) =]\frac{3}{4} - \frac{1}{40}\sqrt{30}; \frac{3}{4} + \frac{1}{40}\sqrt{30}[\subset]0{,}613; 0{,}887[$ hervorgehoben. Man kann mit einer Sicherheit von etwa 90% darauf vertrauen, daß dieses Intervall die Wahrscheinlichkeit $p = P(»\text{Treffer}«)$ enthält, wenn die relative Häufigkeit des Treffers zu $h_{100} = \frac{3}{4}$ gemessen wurde.

schwachen Gesetz der großen Zahlen ist aber h_n ein Näherungswert für p. Ersetzen wir also p durch h_n, so wird $\varepsilon \approx \sqrt{\frac{h_n(1-h_n)}{n\eta}}$.

Mit den Werten aus Beispiel 4 gewinnen wir $\varepsilon \approx \sqrt{\frac{0{,}18 \cdot 0{,}82}{100 \cdot 0{,}1}} = 0{,}121\ldots$, also wie erwartet, ein kleineres Konfidenzintervall um 0,18 für $p = P(\{6\})$. Wir können damit sagen: Das Intervall $]0{,}059; 0{,}301[$ wurde durch ein Verfahren ermittelt, das mit einer Sicherheit von *ungefähr* mindestens 90% zu einem Intervall führt, das die wahre Wahrscheinlichkeit für die Augenzahl 6 bei diesem Würfel enthält.

Die genauere Näherung führt im allgemeinen Fall also zu einem

Näherungskonfidenzintervall $\tilde{I}(h_n) = \left]h_n - \sqrt{\frac{h_n(1-h_n)}{n\eta}};\ h_n + \sqrt{\frac{h_n(1-h_n)}{n\eta}}\right[$.

Figur 259.1 zeigt den Zusammenhang $h_n \mapsto \tilde{I}(h_n)$. Die Hüllkurve dieser Näherungskonfidenzintervalle ist eine Ellipse mit der Gleichung $|p - h_n| = \sqrt{\frac{h_n(1-h_n)}{n\eta}}$, die mit der Ellipse aus Figur 257.1 übereinstimmt, wenn man die Achsenbezeichnungen p und h_n miteinander vertauscht. Diese Näherung ist vor allem für sehr kleine und sehr große h_n nicht sehr sinnvoll. In Figur 259.1 entartet z. B. für

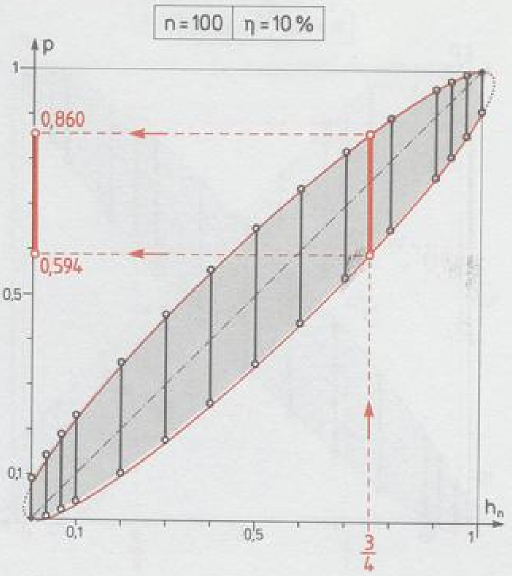

Fig. 260.1 Echte Konfidenzintervalle und die Konfidenzellipse $|p - h_n| = \sqrt{\frac{pq}{n\eta}}$. Das grau unterlegte Gebiet einschließlich des schwarzen Randes ist die Menge der Konfidenzintervalle. Für $h_{100} = \frac{3}{4}$ ist das zugehörige Konfidenzintervall $\mathfrak{I}(\frac{3}{4}) =]\frac{8}{11} - \frac{1}{44}\sqrt{34}; \frac{8}{11} + \frac{1}{44}\sqrt{34}[\subset]0{,}594; 0{,}860[$ rot hervorgehoben. Man kann mit einer Sicherheit von mindestens 90% darauf vertrauen, daß dieses Intervall die Wahrscheinlichkeit $p = P$(»Treffer«) überdeckt, wenn die relative Häufigkeit des Treffers zu $h_{100} = \frac{3}{4}$ gemessen wurde.

$h_n = 0$ das Vertrauensintervall für p zu einem Punkt. Das würde heißen, daß für $h_n = 0$ die Wahrscheinlichkeit p mit der Sicherheit $1 - \eta$ (in unserem Beispiel also 90%) den Wert 0 hätte, was sicher zuviel gesagt ist, wie die grobe Abschätzung von Figur 258.1 zeigt, die als grobes Konfidenzintervall für diesen Fall noch das Intervall $\left[0; \frac{1}{2\sqrt{n\eta}}\right[$ zuläßt.

Das **echte Konfidenzintervall** erhält man, wenn man das oben gefundene $\varepsilon = \sqrt{\frac{pq}{n\eta}}$ verwendet und damit die Ungleichung $|h_n - p| < \varepsilon$ löst. Die Grenzen dieses offenen Intervalls sind somit die Lösungen der Gleichung

$$|h_n - p| = \sqrt{\frac{pq}{n\eta}}.$$

Bezeichnen wir die beiden Lösungen dieser quadratischen Gleichung für p mit p_1 und p_2 (wobei $p_1 < p_2$ sein soll), dann wird jedem h_n
das **echte Konfidenzintervall** $\mathfrak{I}(h_n) =]p_1; p_2[$
zugeordnet.
Man gewinnt dieses echte Konfidenzintervall übrigens graphisch, wenn man die Relation zwischen h_n und p aus Figur 257.1 von der h_n-Achse her liest. Zeichnet man die h_n-Achse, wie üblich, als Rechtswertachse, dann wird die Hüllellipse von Figur 257.1 an der Winkelhalbierenden gespiegelt. Es entsteht Figur 260.1, die

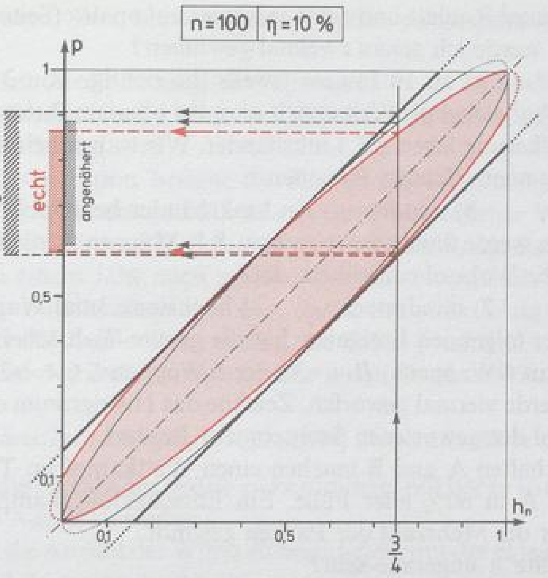

Fig. 261.1 Der Zusammenhang zwischen grobem, Näherungs- und echtem Konfidenzintervall einschließlich der Hüllkurven Parallelenpaar, Näherungskonfidenzellipse (schwarz) und echte Konfidenzellipse (rot). – Hervorgehoben ist der Wert $h_n = \frac{3}{4}$.

die echten Konfidenzintervalle samt der Konfidenzellipse mit der Gleichung $|p - h_n| = \sqrt{\frac{p(1-p)}{n\eta}}$ als Hüllkurve zeigt.

In unserem konkreten Beispiel finden wir das echte Konfidenzintervall durch Lösen der quadratischen Gleichung $|0{,}18 - p| = \sqrt{\frac{p(1-p)}{100 \cdot 0{,}1}}$. Eine leichte Rechnung liefert $p_1 = 0{,}08965\ldots$ und $p_2 = 0{,}32852\ldots$ Damit können wir sagen: Das 90%-Konfidenzintervall]0,089; 0,329[wurde durch ein Verfahren ermittelt, das mit einer Sicherheit von mindestens 90% zu einem Intervall führt, das die wahre Wahrscheinlichkeit für die Augenzahl 6 bei diesem Würfel enthält. Das bedeutet: Führt man sehr oft dieses Verfahren durch, so werden mindestens 90% der so gefundenen Intervalle p enthalten. (Vgl. Aufgaben 275/**96** und **97**.)

Die vermeintlich genauere Schranke 0,301 von $\bar{I}(0{,}18)$ darf uns nicht täuschen! Sie ist ja nur ein Näherungswert. Zur Klärung zeigt Figur 261.1 den Zusammenhang zwischen dem Parallelenpaar der groben Abschätzung, der Hüllellipse sog. »genaueren« Näherung und der Konfidenzellipse.

Aufgaben

Zu 14.1.

1. Eine Urne enthält 6 schwarze, 8 weiße und 10 rote Kugeln. Mit welcher Wahrscheinlichkeit erhält man bei 6maligem Ziehen mit Zurücklegen genau 3 rote Kugeln?
2. Eine Maschine stellt Stanzteile mit einem Ausschußanteil von 5% her. Wie groß ist die Wahrscheinlichkeit, daß 4 zufällig ausgewählte Teile ausnahmslos in Ordnung sind?