



Die σ -Umgebung des Erwartungswertes μ

Die Standardabweichung σ ist ein Maß für die Streuung einer Zufallsgröße X um ihren Erwartungswert $\mu = E(X)$. Für binomialverteilte Zufallsgrößen¹ lassen sich Erwartungswert $\mu = n \cdot p$ und Standardabweichung $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$ sehr leicht berechnen. n ist hier die Länge der Bernoulli-Kette und p die Trefferwahrscheinlichkeit.

Beispiel 1

~ Eine faire Münze ($p = \frac{1}{2}$) werde 50-mal geworfen. Die Zufallsvariable X sei die Anzahl der „Zahl“-Würfe. Wie groß ist die Wahrscheinlichkeit, dass X einen Wert annimmt, der höchstens um σ vom Erwartungswert μ abweicht?

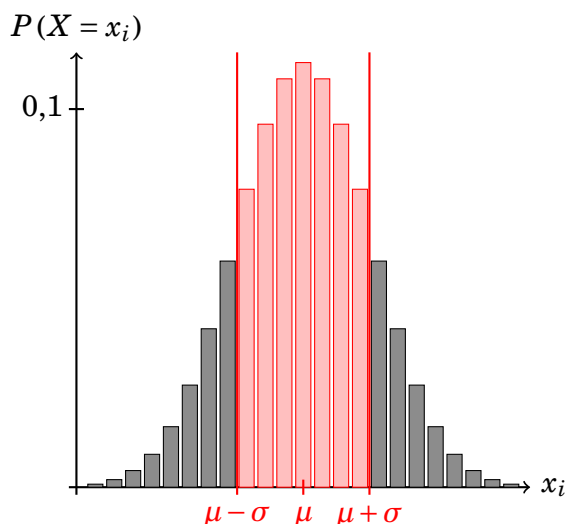
Lösung

Aus $n = 50$ und $p = \frac{1}{2}$ folgt: $\mu = n \cdot p = 50 \cdot \frac{1}{2} = 25$ und damit $\sigma = \sqrt{n \cdot p \cdot (1 - p)} = \sqrt{50 \cdot \frac{1}{2} \cdot \frac{1}{2}} = \sqrt{12,5} \approx 3,5$. Die Abweichung von μ kann sowohl nach oben als auch nach unten ausfallen, daher erhalten wir ein Intervall, also eine untere und eine obere Grenze: $[\mu - \sigma; \mu + \sigma] \approx [25 - 3,5; 25 + 3,5] = [21,5; 28,5]$

Da die Anzahl der Würfe nur eine natürliche Zahl und kein Dezimalbruch sein kann, müssen wir runden. Außerdem hatten wir gefordert, dass die Abweichung höchstens σ betragen soll. Deshalb müssen wir bei der **unteren Grenze immer aufrunden** und bei der **oberen Grenze immer abrunden**. Ansonsten wäre ja die Abweichung größer als σ ! Unser Intervall ist also: $[22; 28]$

Die gesuchte Wahrscheinlichkeit kann man mathematisch sehr kompakt aufschreiben: $P(\underbrace{|X - \mu|}_{\text{Abweichung}} \leq \sigma)$. Da die Abweichung in beide Richtungen möglich ist, verwendet man hier

den Betrag. Bei der konkreten Berechnung hilft uns die kumulierte Binomialverteilung:



$$\begin{aligned}
 P(|X - \mu| \leq \sigma) &= P(|X - 25| \leq 3,5) \\
 &= P(22 \leq X \leq 28) \\
 &= P(X \leq 28) - P(X \leq 21) \\
 &= F(50; 0,5; 28) - F(50; 0,5; 21) \\
 &\approx 0,8389 - 0,1611 \\
 &= 0,6778 \\
 &\approx 68\%
 \end{aligned}$$

Die Zufallsgröße X fällt mit einer Wahrscheinlichkeit von ca. 68% in ein Intervall mit dem Radius σ um den Erwartungswert μ . Dieses Intervall bezeichnet man als **σ -Umgebung des Erwartungswertes μ** .

Ändert sich eigentlich die Wahrscheinlichkeit für die σ -Umgebung des Erwartungswertes μ , wenn

¹ Voraussetzung ist, dass eine Bernoulli-Kette vorliegt: Es gibt genau 2 Ausgänge und die Wahrscheinlichkeit p ist (ungefähr) konstant.



sich n (die Anzahl der Versuche) ändert? Wie wir in der nachfolgenden Tabelle erkennen können, ist die Antwort: Nein! Die Wahrscheinlichkeit $P(|X - \mu| \leq \sigma)$ beträgt (unter bestimmten Voraussetzungen) immer ungefähr 68%, **unabhängig** davon wie groß n ist! Ein entsprechendes Ergebnis erhält man übrigens auch für die 2σ - und 3σ -Umgebung.

n	$P(X - \mu \leq \sigma)$	$P(X - \mu \leq 2\sigma)$	$P(X - \mu \leq 3\sigma)$
50	67,8%	96,7%	99,7%
60	63,4%	94,8%	99,7%
70	71,8%	95,9%	99,7%
80	68,6%	94,3%	99,8%
90	65,7%	95,5%	99,8%
100	72,9%	96,5%	99,8%
200	71,1%	96,0%	99,8%
300	67,4%	95,7%	99,7%
400	70,6%	96,0%	99,8%
500	69,6%	95,6%	99,7%
600	69,3%	95,5%	99,7%
700	69,3%	95,5%	99,7%
800	69,5%	95,6%	99,7%
900	69,9%	95,8%	99,8%
1000	67,3%	95,4%	99,7%
$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\approx 68\%$	$\approx 95,5\%$	$\approx 99,7\%$



Wahrscheinlichkeiten von σ -Umgebungen

Wenn X die Anzahl der Treffer einer Bernoulli-Kette der Länge n mit Erwartungswert μ und Standardabweichung σ ist und die Laplace-Bedingung (Faustregel) $\sigma > 3$ erfüllt ist, dann gilt:

$$P(|X - \mu| \leq 1\sigma) \approx 68\%$$

$$P(|X - \mu| \leq 2\sigma) \approx 95,5\%$$

$$P(|X - \mu| \leq 3\sigma) \approx 99,7\%$$

Ist die Laplace-Bedingung nicht erfüllt, dann können deutlich andere Werte auftreten! Beispielsweise muss beim fairen Münzwurf $n > 36$ gewählt werden, damit die Laplace-Bedingung erfüllt ist ($\sigma = \sqrt{n \cdot \frac{1}{2} \cdot (1 - \frac{1}{2})} = \sqrt{\frac{n}{4}} > 3 \Rightarrow \frac{n}{4} > 9 \Rightarrow n > 36$).

Schluss von der Gesamtheit auf die Stichprobe

Zieht man eine genügend große Stichprobe und ist die Trefferwahrscheinlichkeit p in der Gesamtmenge bekannt, dann kann man mit einer **Sicherheitswahrscheinlichkeit** eine Prognose über die Anzahl der Treffer in der Stichprobe stellen. Besonders häufig werden 2σ -Umgebungen



(95,5%) und 3σ -Umgebungen (99,7%) verwendet, da hier die Sicherheitswahrscheinlichkeit bereits bekannt ist.

Beispiel 2

Ein Blumengroßhandel bestellt 600 Tulpenzwiebeln einer seltenen Sorte, die mit einer Wahrscheinlichkeit von 20% aufgehen. Der Lieferant soll mit einer Sicherheit von über 95% eine Prognose darüber abgeben, wie viele Blumen aufblühen werden.

Lösung

Die Zufallsvariable X sei die Anzahl der aufblühenden Tulpenzwiebeln. Die Wahrscheinlichkeit, dass eine Zwiebel aufgeht, beträgt $p = 0,2$ und die Größe der Stichprobe aus der Grundgesamtheit ist hier: $n = 600$.

Erwartungswert: $\mu = E(X) = 600 \cdot 0,2 = 120$ und Standardabweichung: $\sigma = \sqrt{600 \cdot 0,2 \cdot 0,8} = 4\sqrt{6} > 3$ Damit ist auch die Laplace-Bedingung erfüllt und wir können die Regeln über die σ -Umgebungen anwenden:

Da die Sicherheitswahrscheinlichkeit über 95% betragen soll, verwenden wir eine 2σ -Umgebung: $[\mu - 2\sigma; \mu + 2\sigma] \approx [100,4; 139,6] \Rightarrow [101; 139]$ (Nicht vergessen: nach „innen“ runden!)

Mit einer Wahrscheinlichkeit von ungefähr 95,5% werden daher zwischen 101 und 139 Tulpenzwiebeln aufgehen.

Aufgabe 1

Nach einer Veröffentlichungen des Statistischen Bundesamtes im Jahr 2011 besitzen ungefähr 67% aller Haushalte eine Spülmaschine, 52% einen Laptop, 90% ein Mobiltelefon und 39% ein Navigationsgerät. Eine Stichprobe vom Umfang 700 wird durchgeführt. In wie vielen Haushalten wird sich mit einer Sicherheitswahrscheinlichkeit von 95,5% ein Geschirrspüler, ein Laptop, ein Mobiltelefon bzw. ein Navigationsgerät befinden?

Lösung: Mit einer Sicherheitswahrscheinlichkeit von 95,5% werden zwischen 445 und 493 Haushalte eine Spülmaschine, zwischen 338 und 390 Haushalte einen Laptop, zwischen 615 und 645 Haushalte ein Mobiltelefon und zwischen 248 und 298 Haushalte ein Navigationsgerät besitzen.

Aufgabe 2

Nach einer Veröffentlichungen des Statistischen Bundesamtes im Jahr 2011 besitzen ungefähr 67% aller Haushalte eine Spülmaschine, 52% einen Laptop, 90% ein Mobiltelefon und 39% ein Navigationsgerät. Wie groß muss der Umfang der Stichprobe jeweils mindestens sein?

Lösung: Es müssen mindestens 41 Haushalte bezüglich eines Geschirrspülers, mindestens 37 Haushalte zum Laptop, mindestens 100 Haushalte zum Mobiltelefon und mindestens 38 Haushalte zum Besitz eines Navigationsgerätes befragt werden.

Beurteilung von Stichprobenergebnissen

Verwendet man die σ -Regeln, dann lassen sich statistische Daten ganz leicht beurteilen. Schließlich hat man nicht immer eine passende Tabelle der kumulierten Binomialverteilung zur Hand bzw. kann die Wahrscheinlichkeit mit dem Taschenrechner berechnen. Betrachten wir dazu ein einführendes Beispiel:

**Beispiel 3**

Ein Laplace-Würfel wurde 3600-mal geworfen. Anton behauptet, dass es insgesamt 675 Sechsen waren. Barbara hingegen sagt, dass es nur 540 Sechsen gewesen sind. Welche der Aussagen ist glaubwürdig?

Lösung

Überprüfen wir zuerst die Laplace-Bedingung: $\sigma = \sqrt{3600 \cdot \frac{1}{6} \cdot \frac{5}{6}} = \sqrt{500} \approx 22,4 > 3 \checkmark$

Beginnen wir nun mit der Abschätzung und wählen dazu zuerst ein sehr großes Intervall: Mit einer Wahrscheinlichkeit von 99,7% fällt X (die Anzahl der Sechsen) in eine 3σ -Umgebung von μ , d. h. in das Intervall $[533 ; 667]$. Das Ergebnis von Anton $X = 675$ fällt nicht in dieses Intervall und ist damit sehr unwahrscheinlich! Die Wahrscheinlichkeit, dass ein Ergebnis außerhalb des Intervalls $[533 ; 667]$ auftritt, beträgt schließlich nur 0,3%. Man bezeichnet diese Abweichungen als **hochsignifikante Abweichungen**.

Wählt man ein kleineres Intervall, beispielsweise eine 2σ -Umgebung: $[556 ; 644]$, dann stellt man fest, dass Barbaras genannte Anzahl nicht in dieses Intervall fällt. Da 95,5% aller Ergebnisse in dieses Intervall fallen, liegen 4,5% aller Fälle außerhalb. Man bezeichnet diese Abweichungen als **signifikante Abweichungen**. Die Aussage ist also auch eher zu bezweifeln.

Beispiel 4

Ein Baumarkt erwartet eine sehr große Lieferung an Nägel und möchte die Herstellerangabe, dass höchstens 2% Ausschuss enthalten ist, überprüfen. Dazu entnimmt ein Mitarbeiter des Marktes eine Stichprobe von 600 Stück und lässt diese genau untersuchen. Die Chefin entscheidet nach folgender Regel über die Annahme der Lieferung: Fällt die Anzahl X der unbrauchbaren Nägel in der Stichprobe in eine einfache σ -Umgebung des Erwartungswertes μ von X , dann wird die Lieferung angenommen. Ansonsten schickt sie die Lieferung wieder zurück. Wie lautet die Entscheidung der Chefin, wenn in der Stichprobe genau 18 unbrauchbare Nägel sind?

Lösung

Die Überprüfung der Laplace-Bedingung ergibt: $\sigma = \sqrt{600 \cdot 0,02 \cdot 0,98} \approx 3,4 > 3 \checkmark$

$\mu + \sigma \approx 600 \cdot 0,02 + 3,4 = 15,4$ damit liegt die beobachtete Anzahl an unbrauchbaren Nägeln rechts außerhalb der einfachen σ -Umgebung und die Lieferung wird reklamiert.

Abschätzung von relativen Häufigkeiten: $\frac{\sigma}{n}$ -Umgebungen

Bisher wurden nur absolute Häufigkeiten abgeschätzt. Doch wie sieht es bei relativen Häufigkeiten aus? Dazu teilt man einfach die absolute Häufigkeit X durch die Gesamtzahl n und erhält die relative Häufigkeit $\frac{X}{n}$. Zur Abschätzung können wir dann die Formel aus den vorangegangenen Aufgaben nehmen. Wir gehen zuerst von „ X liegt in einer σ -Umgebung von μ “ aus, nehmen dann ein paar Umformungen vor:

$$|X - \mu| \leq \sigma$$



Beide Seiten durch n teilen

$$\left| \frac{X}{n} - \frac{\mu}{n} \right| \leq \frac{\sigma}{n}$$

μ durch $n \cdot p$ ersetzen

$$\left| \frac{X}{n} - \frac{n \cdot p}{n} \right| \leq \frac{\sigma}{n}$$

Kürzen mit n

$$\left| \frac{X}{n} - p \right| \leq \frac{\sigma}{n}$$

Und erhalten schließlich in der letzten Zeile: „ $\frac{X}{n}$ liegt in einer $\frac{\sigma}{n}$ -Umgebung von p “.



Wahrscheinlichkeiten von $\frac{\sigma}{n}$ -Umgebungen

Wenn X die Anzahl der Treffer einer Bernoulli-Kette der Länge n mit Erwartungswert μ und Standardabweichung σ ist und die Laplace-Bedingung $\sigma > 3$ erfüllt ist, dann fällt $\frac{X}{n}$ ungefähr mit:

$$\begin{aligned} &68\% \text{ ins Intervall } \left[p - \frac{\sigma}{n}; p + \frac{\sigma}{n} \right], \\ &95,5\% \text{ ins Intervall } \left[p - 2\frac{\sigma}{n}; p + 2\frac{\sigma}{n} \right], \\ &99,7\% \text{ ins Intervall } \left[p - 3\frac{\sigma}{n}; p + 3\frac{\sigma}{n} \right]. \end{aligned}$$



Beispiel 5

- ~ Eine Münze wird 500-mal geworfen. In welches Intervall um den erwarteten Wert von $p = 0,5$
- ~ fällt die relative Häufigkeit für „Zahl“ mit einer Sicherheitswahrscheinlichkeit von 99,7%?

Lösung

X sei die Anzahl der „Zahl“-Würfe bei $n = 500$ Versuche, dann ist die Standardabweichung $\sigma = \sqrt{500 \cdot \frac{1}{2} \cdot \frac{1}{2}} = \sqrt{125} \approx 11,2 > 3$ ✓ Die Laplace-Bedingung mit $\sigma > 3$ ist also erfüllt.

Der Radius der Umgebung um die Trefferwahrscheinlichkeit $p = 0,5$ beträgt hier: $3\frac{\sigma}{n} \approx 0,0671$, d. h. die relative Häufigkeit $\frac{X}{n}$ fällt mit einer Wahrscheinlichkeit von 99,7% in das Intervall $[0,4329; 0,5671]$

Zum Weiterdenken: Warum darf hier nicht nach „innen“ auf die nächste ganze Zahl gerundet werden?!

**Beispiel 6**

Ein Würfel wird geworfen und man gewinnt, wenn eine Eins gewürfelt wird. Wie viele Versuche sind mindestens notwendig, wenn bei einer Sicherheitswahrscheinlichkeit von 95,5% die relative Gewinnhäufigkeit von der theoretischen Gewinnwahrscheinlichkeit p höchstens um 2% abweichen soll?

Lösung

Man geht von einer $2\frac{\sigma}{n}$ -Umgebung um die Gewinnwahrscheinlichkeit $p = \frac{1}{6}$ aus und muss dann sicherstellen, dass...

$$\begin{aligned}2\frac{\sigma}{n} &\leq 0,02 \\2\frac{\sqrt{n \cdot \frac{1}{6} \cdot \frac{5}{6}}}{n} &\leq 0,02 \\ \frac{\sqrt{n \cdot \frac{5}{36}}}{n} &\leq 0,01 \\ \frac{n \cdot \frac{5}{36}}{n^2} &\leq 0,0001 \\ \frac{5}{36n} &\leq 0,0001 \\ \frac{5}{36 \cdot 0,0001} &\leq n \\ n &\geq 1388,\bar{8}\end{aligned}$$

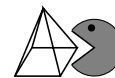
Es sind also mindestens 1389 Versuche notwendig.

Konfidenzintervalle

In der statistischen Praxis kommt es sehr häufig vor, dass die Trefferwahrscheinlichkeit p für das Eintreten eines bestimmten Ereignisses nicht bekannt ist. Im Folgenden wollen wir einen Schätzwert für p bestimmen, genauer gesagt ein Intervall, in dem p mit einer vorgegebenen Wahrscheinlichkeit liegt.

Beispiel 7

Von einem Würfel sei *nicht* bekannt, ob er gefälscht ist. Zur Probe wird er 4000-mal geworfen, dabei fällt genau 680-mal die Eins. Die Wahrscheinlichkeit für das Fallen einer Eins soll mit einer Sicherheitswahrscheinlichkeit von 95,5 % abgeschätzt werden.



Lösung

Übersicht der Bezeichnungen:

n : Länge der Bernoulli-Kette

X : Anzahl der Einsen in der Stichprobe (absolute Häufigkeit)

$\frac{X}{n}$: relative Häufigkeit der Eins

σ : Standardabweichung

p : unbekannte Wahrscheinlichkeit für eine Eins

Für die unbekannte Wahrscheinlichkeit p wird ein **Konfidenzintervall** (Vertrauensintervall) gesucht. Die relative Häufigkeit ist bekannt, daher können wir den Ansatz einer $\frac{\sigma}{n}$ -Umgebung übernehmen. Gehen wir also von der Annahme aus, dass der Würfel echt ist:

$$\left| \frac{X}{n} - p \right| \leq 2 \cdot \frac{\sigma}{n}$$

Setzen wir die angegebenen Werte ein:

$$\left| \frac{680}{4000} - p \right| \leq 2 \cdot \frac{\sigma}{4000}$$

σ kann man nun durch $\sqrt{n \cdot p \cdot (1-p)}$ ersetzen:

$$\left| \frac{680}{4000} - p \right| \leq 2 \cdot \frac{\sqrt{4000 \cdot p \cdot (1-p)}}{4000}$$

Jetzt werden beide Seiten quadriert. Dadurch fällt der Betrag und die Wurzel weg:

$$\left(\frac{680}{4000} - p \right)^2 \leq 2^2 \cdot \frac{4000 \cdot p \cdot (1-p)}{4000^2}$$

Vereinfachen ergibt:

$$\left(\frac{17}{100} - p \right)^2 \leq \frac{p \cdot (1-p)}{1000}$$

Links steht die zweite binomische Formel...

$$\begin{aligned} \frac{289}{10000} - \frac{17}{50}p + p^2 &\leq \frac{p \cdot (1-p)}{1000} && | \cdot 1000 \\ \frac{289}{10} - 340p + 1000p^2 &\leq p - p^2 && | + p^2 - p \\ 1001p^2 - 341p + \frac{289}{10} &\leq 0 && | : 1001 \\ p^2 - \frac{31}{91}p + \frac{289}{10010} &\leq 0 \end{aligned}$$

Fortsetzung auf der nächsten Seite ~>

**Fortsetzung der Lösung**

Anwenden der *pq-Formel* ergibt die Randwerte der Ungleichung:

$$p_{1,2} = \frac{31}{182} \pm \sqrt{\left(\frac{31}{182}\right)^2 - \frac{289}{10010}}$$
$$p_1 \approx 0,1585$$
$$p_2 \approx 0,1822$$

Damit erhalten wir als Konfidenzintervall für p :

$$0,1585 \leq p \leq 0,1822$$

Da die theoretische Wahrscheinlichkeit $\frac{1}{6} = 0,1\bar{6}$ für eine Eins eines Laplace-Würfels im Konfidenzintervall von p liegt, kann die Annahme, dass der untersuchte Würfel echt ist, nicht abgelehnt werden. Läge die theoretische Wahrscheinlichkeit außerhalb des Intervalls, dann würden wir die Annahme, dass der Würfel echt ist, verwerfen und mit einer Sicherheit von 95,5% von einer Fälschung ausgehen.

**Aufgabe 3**

Der Anteil (Prozentsatz) p der Hörer, die eine bestimmte Radiosendung regelmäßig hören, soll mit einer Sicherheitswahrscheinlichkeit von 99,7% abgeschätzt werden. Von einem Marktforschungsinstitut wird eine Stichprobe von 1000 Zuschauern befragt, von denen 750 die Sendung regelmäßig hören.

Lösung: Mit einer Sicherheitswahrscheinlichkeit von 99,7% liegt der Anteil der Zuschauer zwischen 70,68% und 78,87%.

**Aufgabe 4**

Von einem Würfel sei nicht bekannt, ob er gefälscht ist. Zur Probe wird er 9000-mal geworfen, wobei 1650-mal die Eins fällt. Ein 95,5%-Konfidenzintervall für die Wahrscheinlichkeit der Eins ist gesucht.

Lösung: $0,1753 \leq p \leq 0,1916$. $p = \frac{1}{6}$ liegt links außerhalb, daher ist mit einer Sicherheitswahrscheinlichkeit von 95,5% von einer Fälschung auszugehen. Wie eine zusätzliche Rechnung zeigt, ist sogar mit einer Sicherheitswahrscheinlichkeit von 99,7% von einer Fälschung auszugehen ($0,1714 \leq p \leq 0,1959$)!

**Aufgabe 5**

2000 Personen wurden gefragt welchen Moderator einer Fernsehsendung sie bevorzugen würden. 790 entschieden sich für Moderator A und 880 wählten Moderator B. Daraufhin behauptet Moderator B, der beliebteste Moderator zu sein. Dieser Anspruch soll auf einem Sicherheitsniveau von 95,5% geprüft werden.

Tipp: Vergleiche, ob sich die beiden Konfidenzintervalle überschneiden.

Lösung: p_a sei die Wahrscheinlichkeit, dass Moderator A der beliebteste ist und p_b die Wahrscheinlichkeit, dass es Moderator B ist. Dann ergeben sich folgende Konfidenzintervalle: $0,3734 \leq p_a \leq 0,4171$ und $0,4179 \leq p_b \leq 0,4623$. Da sich die beiden Intervalle nicht überschneiden, kann Moderator B mit einer Wahrscheinlichkeit von 95,5% behaupten, der beliebteste Moderator zu sein. Hätten sich jedoch beispielsweise 800 Personen für Moderator A entschieden, dann ließe sich der Anspruch auf diesem Sicherheitsniveau nicht aufrechterhalten, da sich dann die Konfidenzintervalle überschneiden!

**Beispiel 8**

Braunbären gehören zu den bedrohten Tierarten. Um den Bestand in einer bestimmten Region abschätzen zu können, wurden dort 150 Tiere eingefangen, mit einer Markierung versehen und wieder freigelassen. Nach einiger Zeit wurden 120 Tiere in freier Wildbahn beobachtet, von denen 40 markiert waren. Berechnen Sie ein 95,5 %-Konfidenzintervall für den unbekannten Tierbestand N .

**Lösung**

$$\begin{aligned}
 \left| \frac{X}{n} - p \right| &\leq 2 \cdot \frac{\sigma}{n} \\
 \left| \frac{40}{120} - p \right| &\leq 2 \cdot \frac{\sqrt{120 \cdot p \cdot (1-p)}}{120} \\
 \left(\frac{1}{3} - p \right)^2 &\leq 2^2 \cdot \frac{120 \cdot p \cdot (1-p)}{120^2} \\
 p^2 - \frac{2}{3}p + \frac{1}{9} &\leq \frac{1}{30}p - \frac{1}{30}p^2 \\
 p^2 - \frac{21}{31}p + \frac{10}{93} &\leq 0 \\
 0,2539 &\leq p \leq 0,4235
 \end{aligned}$$

Der Erwartungswert $\mu = N \cdot p$ für die Anzahl der markierten Tiere im Gesamtbestand ist bekannt und hier 150. Daher gilt: $150 = N \cdot p$ bzw. $N = \frac{150}{p}$. Setzen wir die beiden Wahrscheinlichkeiten aus dem Konfidenzintervall ein, dann erhalten wir $\frac{150}{0,4235} \leq N \leq \frac{150}{0,2539}$, also $354,2 \leq N \leq 590,8$. Der Bestand an Braunbären liegt daher mit einer Wahrscheinlichkeit von 95,5 % zwischen $N = 355$ und $N = 590$ Tieren (nach „innen“ runden!).



Ein Näherungsverfahren für Konfidenzintervalle

Beispiel 9

Bei einer Umfrage werden 2600 Personen befragt. 1092 von ihnen geben an, eine bestimmte Tageszeitung zu lesen. Mit welchem Marktanteil kann der Verlag rechnen, wenn eine Sicherheitswahrscheinlichkeit von 95,5 % gefordert wird. Vergleichen Sie die Näherungslösung mit der exakten Lösung.

Lösung

Die relative Häufigkeit $h_n = \frac{X}{n}$ ist hier: $h_n = \frac{1092}{2600} = \frac{21}{50} = 0,42$.

Näherungslösung:

$$\begin{aligned} \left| \frac{X}{n} - p \right| &\leq 2 \cdot \frac{\sigma}{n} \\ \left| \frac{1092}{2600} - p \right| &\leq 2 \cdot \frac{\sqrt{2600 \cdot p \cdot (1-p)}}{2600} \\ \left| \frac{21}{50} - p \right| &\leq 2 \cdot \frac{\sqrt{2600 \cdot h_n \cdot (1-h_n)}}{2600} \\ |0,42 - p| &\leq 2 \cdot \frac{\sqrt{2600 \cdot 0,42 \cdot 0,58}}{2600} \\ |0,42 - p| &\leq 0,019359 \\ 0,4006 &\leq p \leq 0,4394 \end{aligned}$$

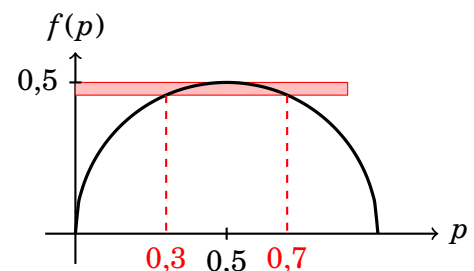
Exakte Lösung:

$$\begin{aligned} \left| \frac{X}{n} - p \right| &\leq 2 \cdot \frac{\sigma}{n} \\ \left| \frac{1092}{2600} - p \right| &\leq 2 \cdot \frac{\sqrt{2600 \cdot p \cdot (1-p)}}{2600} \\ \left(\frac{21}{50} - p \right)^2 &\leq 2^2 \cdot \frac{2600 \cdot p \cdot (1-p)}{2600^2} \\ p^2 - \frac{21}{25}p + \frac{441}{2500} &\leq \frac{1}{650}p - \frac{1}{650}p^2 \\ p^2 - \frac{547}{651}p + \frac{273}{1550} &\leq 0 \\ 0,4008 &\leq p \leq 0,4395 \end{aligned}$$

Der Verlag kann mit einem Marktanteil zwischen 40 % und 44 % rechnen. Das Näherungsverfahren liefert in diesem Fall fast das gleiche Konfidenzintervall wie die exakte Lösung.

Die Rechnung in der rechten Spalte ist – bis auf ein paar ausgelassene Zwischenschritte – die gleiche wie in Beispiel 7 auf Seite 6. In der linken Spalte wird p durch h_n ersetzt. Dadurch vereinfacht sich die rechte Seite sehr und die Abweichung ist vernachlässigbar gering.

In der Lösung wird angedeutet, dass das Näherungsverfahren nicht immer annähernd gleiche Werte liefert. Betrachten wir dazu einmal den abgeänderten Term: $f(p) = \sqrt{p \cdot (1-p)}$. Der Graph von $f(p)$ verläuft im Intervall $[0,3; 0,7]$ sehr flach und daher sind die Abweichungen nur gering. Die Näherungslösung weicht für $0,3 \leq h_n \leq 0,7$ nur wenig von der exakten Lösung ab.



Faustregel zur Anwendung der Näherungslösung

Gilt für die relative Häufigkeit: $0,3 \leq h_n \leq 0,7$, dann liefert die Näherungslösung annähernd exakte Ergebnisse. Für kleinere bzw. größere Werte können deutlich größere Abweichung von der exakten Lösung auftreten.