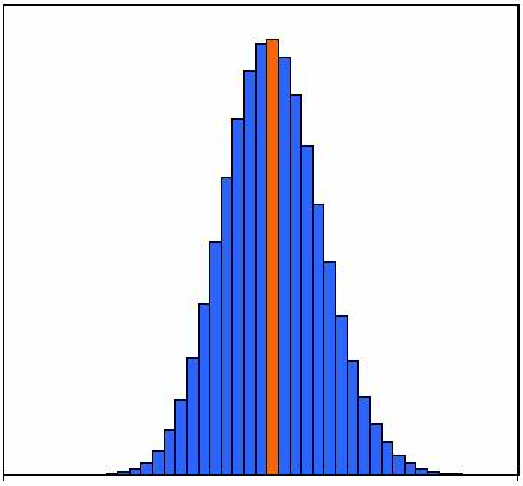
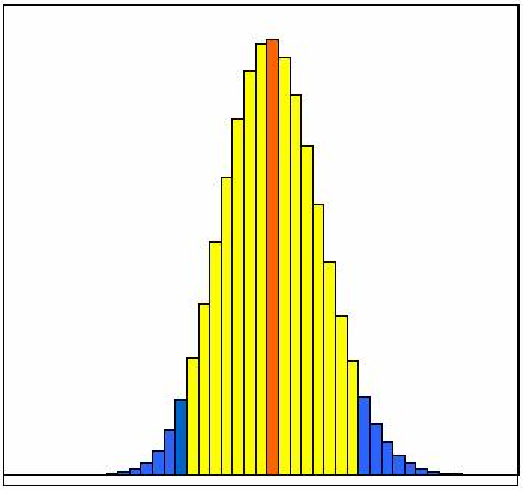
M13 Übungsaufgaben

06.11.2006 / pl

Die Histogramme von Binomialverteilungen werden bei wachsendem Stichprobenumfang *n* immer flacher und breiter. Dem Maximum einer solchen Verteilung kommt daher keine allzu große Wahrscheinlichkeit zu. Vielmehr treten die Nachbarwerte des Maximums mit vergleichbaren Wahrscheinlichkeiten auf (siehe Abbildung).





Da die Histogramme (für große n) nahezu symmetrisch zum Erwartungswert E(X) = μ = n·p sind, betrachtet man symmetrische Bereiche um den Erwartungswert m. Man interessiert sich für z.B. für die symmetrische Umgebung von m, in der man ein Ergebnis mit einer Wahrscheinlichkeit von 90% findet. Man nennt diese Umgebung das 90%-Sicherheitsintervall.

Man könnte die Einzelwahrscheinlichkeiten P(X=k) so lange aufsummieren, bis man die gewünschte Sicherheitswahrscheinlichkeit erreicht hat. Das ist aufwändig! Es gibt aber eine Faustregel, die (bei Vorgabe von n, p und der Sicherheitswahrscheinlichkeit) den Radius des Intervalls liefert. Dabei bezeichnet man als Radius den Abstand der Intervallränder vom Erwartungswert.

Jedem Radius einer Umgebung des Erwartungswertes μ lässt sich eine bestimmte Wahrscheinlichkeit für diese Umgebung zuordnen. Umgekehrt gehören zu bestimmten Wahrscheinlichkeiten um den Erwartungswert bestimmte Radien.
Die folgenden Faustregeln für Binomialverteilungen gelten umso genauer, je größer der Stichprobenumfang n ist, insbesondere falls

$$\sigma = \sqrt{n \cdot p \cdot (1-p)} > 3\sigma \qquad \text{(LAPLACE-Bedingung).}$$

Es gelten folgende Zuordnungen:

Radius der Umgebung	Wahrscheinlichkeit der Umgebung
1σ	68%
2σ	99,5%
3σ	99,7%

Wahrscheinlichkeit der Umgebung	Radius der Umgebung
90%	1,64σ
95%	1,96σ
99%	2,58σ

Beispiel: *Man hat ein 100-stufiges Bernoulli-Experiment mit der Erfolgswahrscheinlichkeit p=0,4. Daraus folgt: Erwartungswert der Zufallsvariable X=Anzahl der Erfolge μ = n·p =40 und Standardabweichung σ mit* σ^2 =n·p·(1-p)=24, d.h. σ ≈ 4,90. *Damit ergibt sich das 90%-Intervall als*

$$[40-1{,}64\cdot\sigma;\ 40+1{,}64\cdot\sigma] = [31{,}96;\ 48{,}03].$$

Man rundet stets "zur sicheren Seite", d.h. zum Erwartungswert hin. Damit bekommt man das Intervall [32; 48]. Mit 90% Wahrscheinlichkeit wird man also zwischen 32 und 48 Erfolge haben. Mit 10% Wahrscheinlichkeit bekommt man Erfolgsanzahlen außerhalb dieses Intervalls.

M13 Übungsaufgaben 06.11.2006 / pl

1) 5% aus der Produktion einer bestimmten Sorte Nägel sind nicht einwandfrei. In einer Packung sind 360 Nägel. Wie viele Nägel sind in 90% [95%; 99%] der Packungen nicht in Ordnung?

2) Die erste Wochenziehung beim Lotto 6 aus 49 fand im Jahre 1955 statt. Die 3000. Ziehung wird voraussichtlich im Jahre 2011 durchgeführt.
Wie oft wird z.B. die Zahl 37 bis zur 3000. Ziehung gezogen worden sein? (Gesucht sind die 90%-, 95%- und 99%-Sicherheitsintervalle.)

3) Die Wahrscheinlichkeit für eine Mädchengeburt beträgt in der Bundesrepublik p=0,487. Ein Krankenhaus gab die Geburtenzahlen des ersten Halbjahres bekannt. Beantworten Sie die folgenden Fragen jeweils auf der Basis einer Sicherheitswahrscheinlichkeit von 95%.

Monat	J	F	M	A	M	J	Summe
Anz. Jungengeburten	57	47	53	57	52	49	315
Anz. Mädchengeburten	53	43	68	50	54	50	318

Wochentag	M	D	M	D	F	S	S	Summe
Anz. Jungengeburten	41	38	54	59	54	42	27	315
Anz. Mädchengeburten	54	46	48	35	49	51	35	318

(a) Ist die relative Häufigkeit der Mädchengeburten in diesem Krankenhaus verträglich mit der Wahrscheinlichkeit für Mädchengeburten in der Bundesrepublik?
(b) Sind einzelne Monatsergebnisse des Krankenhauses ungewöhnlich?
(c) Gibt es einen Wochentag mit ungewöhnlich vielen oder wenigen Mädchen- bzw. Jungengeburten?
(d) Angenommen, in einem Jahr kommen in der Bundesrepublik n=600.000 Kinder zur Welt. Welche Mädchen-Anteile sind mit p=0,487 verträglich

Lösung 1

Aus dem Text ergibt sich: Größe der Stichprobe n=360, Erfolgswahrscheinlichkeit p=0,05.

1) Erwartungswert: $\mu = E(X) = n \cdot p = 360 \cdot 0{,}05 = 18$
2) Standardabweichung $\sigma = \sqrt{n \cdot p \cdot (1-p)} \approx 4{,}14$
3) Laplace-Bedingung: Da $\sigma > 3$ ist, dürfen die Sicherheitsintervalle mit der Faustformel bestimmt werden.
4) "Intervallschätzung":
   Bestimmen des 90%-Sicherheitsintervalls:
   $[\mu - 1{,}64\sigma;\ \mu + 1{,}64\sigma]$
   $= [18 - 6{,}8;\ 18 + 6{,}8]$
   $= [11{,}2;\ 24{,}8]$
   Bestimmen des 95%-Sicherheitsintervalls:
   $[\mu - 1{,}96\sigma;\ \mu + 1{,}96\sigma]$
   $= [18 - 8{,}1;\ 18 + 8{,}1]$
   $= [9{,}9;\ 26{,}1]$
   Bestimmen des 99%-Sicherheitsintervalls:
   $[\mu - 2{,}58\sigma;\ \mu + 2{,}58\sigma]$
   $= [18 - 10{,}7;\ 18 + 10{,}7]$
   $= [7{,}3;\ 28{,}7]$
5) Da beim Stichprobenergebnis nur ganzzahlige Zahlen sinnvoll sind, wird zur "sicheren Seite", also zum Erwartungswert hin gerundet:

90%-Sicherheitsintervall [12; 24]
95%-Sicherheitsintervall [10; 26]
99%-Sicherheitsintervall [8; 28]

Interpretation:
Mit einer Wahrscheinlichkeit von 99% sind in einer Packung 8 bis 28 defekte, also 332 bis 352 einwandfreie Nägel.
Anders ausgedrückt: In 99% der Packungen sind zwischen 2,0% und 8,0% der Nägel nicht in Ordnung. In den anderen Packungen sind es weniger als 2,0% oder mehr als 8,0%.

Lösung 2

Aus dem Text ergibt sich: Größe der Stichprobe n=3000. Die Erfolgswahrscheinlichkeit ist p=6/49, da bei jeder Ziehung 6 von 49 Kugeln gezogen werden.

1) Erwartungswert: $\mu = E(X) = n \cdot p = 3000 \cdot 6 / 49 = 367{,}3$
2) Standardabweichung: $\sigma = \sqrt{n \cdot p \cdot (1-p)} \approx 17{,}95$
3) Überprüfen der Laplace-Bedingung: Da $\sigma > 3$ ist, dürfen die Sicherheitsintervalle mit der Faustformel bestimmt werden.
4) "Intervallschätzung":
   Bestimmen des 90%-Sicherheitsintervalls:
   $[\mu - 1{,}64\sigma;\ \mu + 1{,}64\sigma]$
   $= [367{,}3 - 29{,}4;\ 367{,}3 + 29{,}4]$
   $= [337{,}9;\ 396{,}7]$
   Bestimmen des 95%-Sicherheitsintervalls:
   $[\mu - 1{,}64\sigma;\ \mu + 1{,}64\sigma]$
   $= [367{,}3 - 35{,}2;\ 367{,}3 + 35{,}2]$
   $= [332{,}1;\ 402{,}5]$
   Bestimmen des 99%-Sicherheitsintervalls:
   $[\mu - 1{,}64\sigma;\ \mu + 1{,}64\sigma]$
   $= [367{,}3 - 46{,}3;\ 367{,}3 + 46{,}3]$
   $= [321{,}0;\ 413{,}6]$

5) Da beim Stichprobenergebnis nur ganzzahlige Zahlen sinnvoll sind, wird zur "sicheren Seite", also zum Erwartungswert hin gerundet:

90%-Sicherheitsintervall [338; 396]
95%-Sicherheitsintervall [333; 402]
99%-Sicherheitsintervall [322; 413]

Interpretation

Mit einer Wahrscheinlichkeit von 95% wird die Zahl 37 in 3000 Ziehungen 333 bis 402 mal gezogen. Oder auch: 5% der Ziehungszahlen (also 2 bis 3) werden vermutlich eine Ziehungshäufigkeit außerhalb des Intervalls [333; 402] besitzen.

Lösung 3a

Aus dem Text ergibt sich: Größe der Stichprobe $n = 315+318 = 633$.
Die Erfolgswahrscheinlichkeit (für eine Mädchengeburt) ist $p = 0{,}487$.

1) Erwartungswert $\mu = np = 308{,}3$
2) Standardabweichung $\sigma \approx 12{,}6$ 3.
3) Laplace-Bedingung erfüllt, da $\sigma > 3$
4) 95%-Sicherheitsintervall: $[\mu - 1{,}96\sigma;\ \mu + 1{,}96\sigma] = [283{,}7;\ 332{,}9]$
5) Runden zur sicheren Seite: 95%-Sicherheitsintervall [284; 332]

Interpretation:
Mit einer Wahrscheinlichkeit von 95% werden bei n=633 Geburten zwischen 283 und 332 Mädchen geboren (in Prozent: 48,7%±3,9%). Damit sind die 318 Mädchengeburten in dem betrachteten Fall verträglich mit der Wahrscheinlichkeit für die Gesamtheit.

Lösung 3b

Der wahrscheinlichste Kandidat für ein ungewöhnliches Ergebnis ist der März mit 68 Mädchen bei insgesamt n=121 Geburten. Die Erfolgswahrscheinlichkeit ist nach wie vor p=0,487.

1) Erwartungswert $\mu = np = 58{,}9$
2) Standardabweichung $\sigma \approx 5{,}5$
3) Laplace-Bedingung erfüllt, da $\sigma > 3$
4) 95%-Sicherheitsintervall: $[\mu - 1{,}96\sigma;\ \mu + 1{,}96\sigma] = [48{,}2;\ 69{,}7]$
5) Runden zur sicheren Seite: 95%-Sicherheitsintervall [49; 69]

Antwort:Damit liegt auch Zahl der Mädchengeburten im März knapp im "normalen" Schwankungsbereich und ist somit (im Rahmen der Sicherheitswahrscheinlichkeit 95%) mit der Wahrscheinlichkeit p=0,487 für die Gesamtheit verträglich.

Lösung 3c

Man kann die Frage auf zwei verschiedene Arten verstehen!

Man sollte erwarten, dass die Wahrscheinlichkeit für eine beliebige Geburt für jeden Wochentag gleich ist, also p=1/7. Es fällt aber ins Auge, dass sowohl bei Jungen- als auch bei Mädchengeburten am Sonntag die wenigsten Kinder geboren werden. Wir können nun prüfen, ob das Ergebnis im Rahmen der Sicherheit 95% noch mit der Wahrscheinlichkeit p=1/7 verträglich ist.	Die andere Art die Frage zu verstehen ist die Folgende: Man sollte erwarten, dass an jedem Wochentag die Wahrscheinlichkeit für eine Mädchengeburt im Vergleich zu den Jungengeburten p=48,7% ist. Der "verdächtigste" Kandidat für ein ungewöhnliches Ergebnis ist der Donnerstag mit nur 35 Mädchen bei insgesamt n=94 Geburten.
Wahrscheinlichkeit für eine Geburt für jeden Wochentag p=1/7. Sonntag verdächtig mit X=35 bei n=318 (Mädchen) bzw. X=27 bei n=315 (Jungen).	Wahrscheinlichkeit für eine Mädchengeburt p=48,7%. Donnerstag verdächtig mit X=35 bei n=94.
Das übliche Vorgehen ergibt: Für die Mädchen: $\mu \approx 45{,}4$ und $\sigma \approx 6{,}24$. 95%-Intervall nach Runden zur sicheren Seite: [34; 57]. Für die Jungen: $\mu \approx 45{,}0$ und $\sigma \approx 6{,}21$. 95%-Intervall nach Runden zur sicheren Seite: [33; 57].	Hier folgt wie üblich: $\mu \approx 45{,}8$ und $\sigma \approx 4{,}8$. 95%-Intervall nach Runden zur sicheren Seite: [37;55].
Mädchen: 95%-Intervall: [34; 57]. Das Stichprobenergebnis X=35 liegt also noch knapp im 95%-Intervall. Jungen: 95%-Intervall: [33; 57]. Das Ergebnis X=27 liegt außerhalb des 95%-Intervalls und ist somit nicht mit der erwarteten Gleichverteilung verträglich. Der Grund liegt auf der Hand: Sonntags haben auch die Entbindungsärzte gerne frei!	95%-Intervall [37;55]. Das Ergebnis liegt außerhalb des 95%-Intervalls und ist somit mit p=0,487 nicht verträglich. Ob diese signifikante Abweichung eine besondere Ursache hat, liegt nicht auf der Hand.

Lösung 3d

Aus dem Text ergibt sich: Größe der Stichprobe n = 600.000.
Die Erfolgswahrscheinlichkeit (für eine Mädchengeburt) ist p = 0,487.

1) Erwartungswert $\mu = np = 292.200$
2) Standardabweichung $\sigma \approx 387{,}2$
3) Laplace-Bedingung erfüllt, da $\sigma > 3$
4) 95%-Sicherheitsintervall: $[\mu - 1{,}96\sigma;\ \mu + 1{,}96\sigma]$ = [291.441,2; 292.958,8]
5) Runden zur sicheren Seite: 95%-Sicherheitsintervall [291.442; 292.958]
6) In Prozent lautet das Intervall [48,57%; 48,83%].

Interpretation:
Damit schwankt in Deutschland selbst bei Annahme einer konstanten Wahrscheinlichkeit für eine Mädchengeburt der Mädchenanteil von Jahr zu Jahr noch in einem Bereich von ca. 0,26%.