Gewerbeschule Lörrach	Hypergeometrische Verteilung	

Idee: Baumann, LFB QS 1996

Aufgabe:

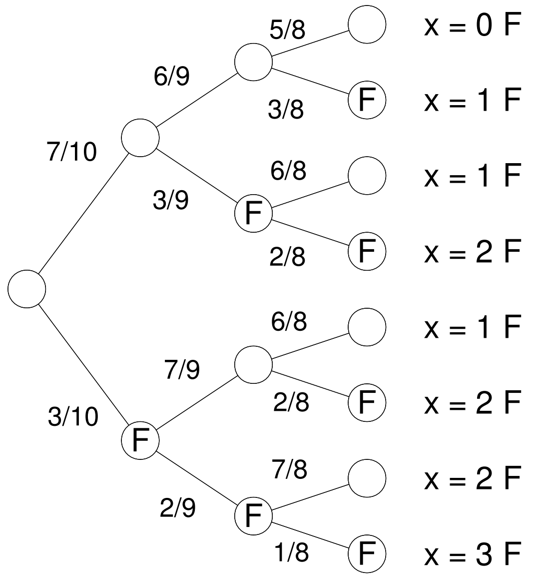
Wie groß sind die Wahrscheinlichkeiten g(x), aus einer Grundgesamtheit mit N = 10 Teilen, davon d = 3 fehlerhaften Teilen, in einer Stichprobe vom Umfang n = 3 genau x = 0, 1, 2 oder 3 fehlerhafte Teile zu finden ?

Lösungen:

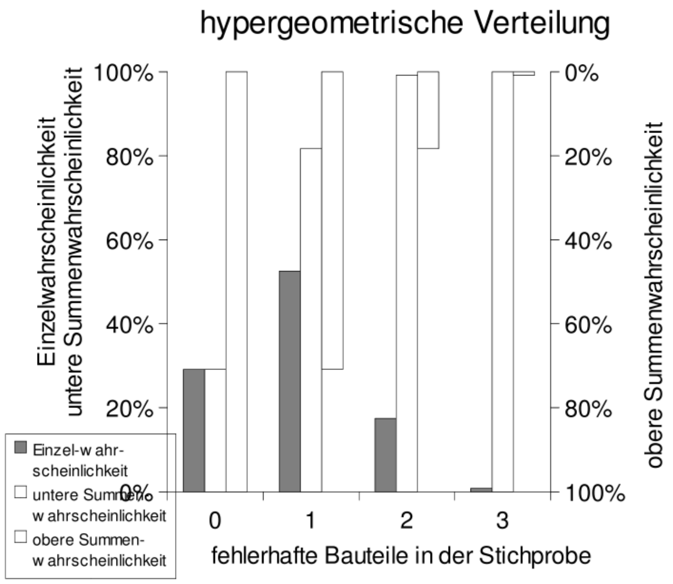
g(x = 0) = g(x = 1) =

g(x = 2) = g(x = 3) =

Darstellung im Wahrscheinlichkeitsbaum
(kaum geeignet für größere Stichproben)



Darstellung als Wahrscheinlichkeitsfunktion



Berechnung der hypergeometrischen Verteilung mit Tabellenkalkulationen
Excel & Co berechnen für die hypergeometrischen Verteilung nur die Einzelhäufigkeiten g(x = 0); g(x = 1) usw.

Die Funktion lautet (Parameter siehe unten):
=HYPGEOMVERT(x; n; d; N) in Excel, bzw.
=HYPGEOMVERT(X; N_Stich; M; N_Gesamt) in OOo.

Um Summenhäufigkeiten zu ermitteln, berechnet man alle Einzelhäufigkeiten und summiert sie, was in Tabellenkalkulationen nicht aufwändig ist.

Die untere Summenhäufigkeit G(x) gibt an, wie groß die Wahrscheinlichkeit ist, den Wert x oder einen kleineren zu ziehen.

Notwendige Parameter
Verteilungsart: hypergeometrisch

Abkürzungen:	Anzahl alles	Anzahl Ausschuss
Im ganzen Los	N / N_Gesamt	d / M
In der Stichprobe	n / N_Stich	x / X

linke Abkürzung:Excel rechts: OpenOffice / StarOffice

Übungen

1) Wie groß sind die Wahrscheinlichkeiten g(x) und G(x), aus einem Paket mit N = 100 Schrauben, davon d = 8 fehlerhaft, in einer Stichprobe vom Umfang n = 5 zufällig genau x = 0, 1, 2, 3, 4 oder 5 fehlerhafte Schrauben zu finden ?
2) Wie groß sind im Lotto die Wahrscheinlichkeiten für 0, 1, 2, .. 6 Richtige ?
3) Stellen Sie die oben abgebildeten grafischen Funktionen mit einer Tabellenkalkulation nach.

Anwendungen
Stichproben, Lotto

Lösungen:

0) g(x = 0) = 0,29167; g(x = 1) = 0,525; g(x = 2) = 0,175; g(x = 3) = 0,0083333
1) g(x = 0) = 0,65319; g(x = 1) = 0,296906; g(x = 2) = 0,0467041; g(x = 3) = 0,0031127; g(x = 4) = 85,539E-6; g(x = 5) = 0,75482E-6
2) g(x = 0) = 0,435965; g(x = 1) = 0,4130195; g(x = 2) = 0,132,3780; g(x = 3) = 0,0176504 g(x = 4) = 0,0009686197; g(x = 5) = 0,0000184499, g(x = 6) = 0,0000000715112

Quelle: Baumann, LFB QS 1996

Aufgabe:

Wie groß sind die Wahrscheinlichkeiten g(x), aus einem Paket mit N = 100 Schrauben, davon d = 8 fehlerhaft, in einer Stichprobe vom Umfang n = 5 zufällig genau x = 0, 1, 2, 3, 4 oder 5 fehlerhafte Schrauben zu finden, wenn jede gezogene Schraube sofort <u>zurückgelegt und untergemischt</u> wird ?

Lösung

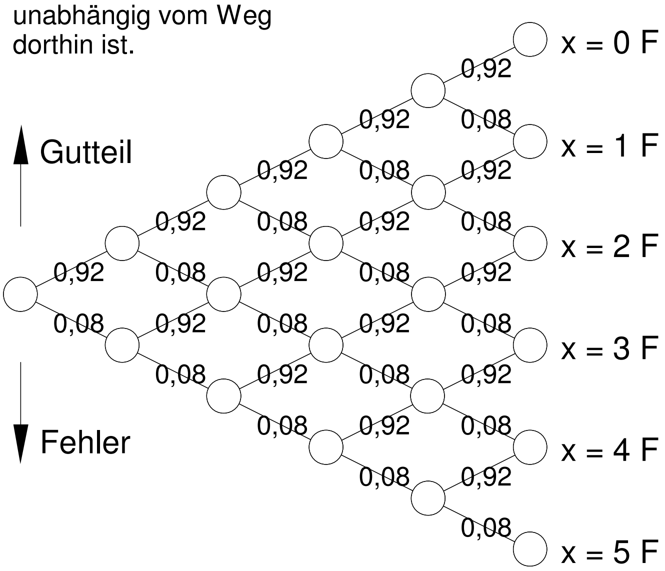
g(x = 0) = 　　　g(x = 1) =

g(x = 2) = 　　　g(x = 3) =

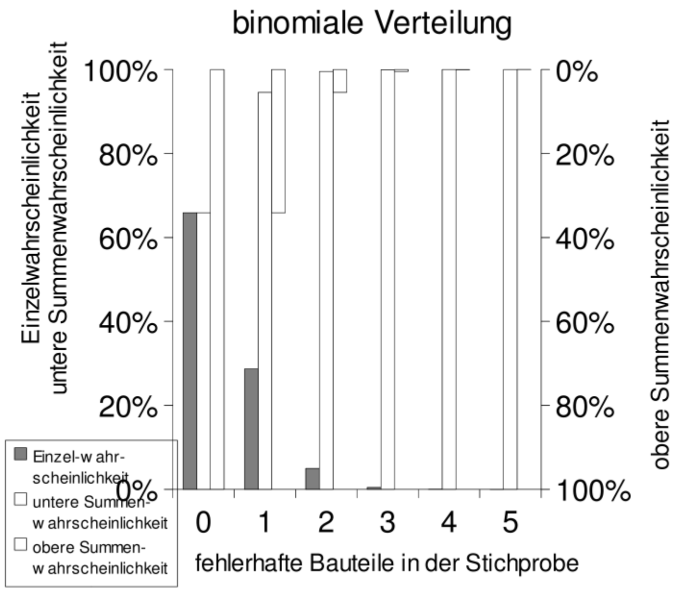
g(x = 4) = 　　　g(x = 5) =

Darstellung im vereinfachten W.-baum

Die Knoten können zusammengefasst werden, weil die Wahrscheinlichkeit der Knoten unabhängig vom Weg dorthin ist.



Darstellung als Wahrscheinlichkeitsfunktion



Berechnung der binomialen Verteilung mit Tabellenkalkulationen

Excel & Co berechnen Einzelhäufigkeiten g(x = 0); g(x = 1) usw. der binomialen Verteilung, wenn als letzter Parameter eine „0“ angeben ist[1].

Die Funktion lautet (Parameter siehe unten):

=BINOMVERT(x; n; p; 0) in Excel bzw.

=BINOMVERT(X; N; W; 0) in OpenOffice.org.

Die unteren Summenhäufigkeiten G(x ≤ 0); G(x ≤ 1) usw. werden mit dem letzten Parameter „1“ berechnet:

=BINOMVERT(x; n; p; 1) in Excel bzw.

=BINOMVERT(X; N; W; 1) in OpenOffice.org.

Obere Summenhäufigkeiten $G_{ob}(x)$:

=1 – BINOMVERT(X;N;W;1) + BINOMVERT(X;N;W;0)

Notwendige Parameter

Verteilungsart: binomial

Excel	Abkürzungen:	OpenOffice / StarOffice
p:	(prozentualer) Anteil der Fehler in der Grundgesamtheit	:W
n:	Umfang der Stichprobe	:N
x:	Anzahl der gesuchten Fehler in der Stichprobe	:X

Übungen

1) Wie groß sind die Wahrscheinlichkeiten g(x) und G(x), aus einem Paket mit N = 10 Schrauben, davon d = 3 fehlerhaft, in einer Stichprobe vom Umfang n = 3 zufällig genau x = 0, 1, 2 oder 3 fehlerhafte Schrauben zu finden, wenn jede gezogene Schraube zurückgelegt und sofort wieder untergemischt wird ?
2) Wie groß sind beim Kniffel die Wahrscheinlichkeiten, mit 5 Würfeln x = 0, 1, 2, .. oder 5 mal die Sechs zu würfeln ?
3) Stellen Sie die oben abgebildeten grafischen Funktionen mit einer Tabellenkalkulation nach.

Anwendungen

Schätzwert für hypergeometrische Verteilungen (Stichproben) mit N/n>10

- Wareneingang
- n-c - Anweisungen

Lösungen:

0) Aufgabe: g(x = 0) = 0,6590815; g(x = 1) = 0,2865572; g(x = 2) = 0,0498360; g(x = 3) = 0,0043336; g(x = 4) = 0,0001884160; g(x = 5) = 0,0000032768
1) g(x = 0) = 0,343; g(x = 1) = 0,441; g(x = 2) = 0,189; g(x = 3) = 0,027
2) g(x = 0) = 0,4018776; g(x = 1) = 0,4018776, g(x = 2) = 0,1607510; g(x = 3) = 0,0321502; g(x = 4) = 0,0032150; g(x = 5) = 0,0001286008

[1] 0 bedeutet „nicht kumuliert“

Berechnung der hypergeometrischen Verteilung

Wie groß sind die Wahrscheinlichkeiten g(x), aus einer Grundgesamtheit mit N = 10 Teilen, davon d = 3 fehlerhaften Teilen, in einer Stichprobe vom Umfang n = 3 genau x = 0, 1, 2 oder 3 fehlerhafte Teile zu finden ?

Für diese einfache Aufgabe kann man das Ergebnis aus dem Wahrscheinlichkeitsbaum (s.o.) ermitteln. Für x = 1 fehlerhaftes Teil gibt es drei mögliche Zweige, deren Einzelergebnis addiert werden:

$$g(x=1)=\frac{7}{10}\cdot\frac{6}{9}\cdot\frac{3}{8}+\frac{7}{10}\cdot\frac{3}{9}\cdot\frac{6}{8}+\frac{3}{10}\cdot\frac{7}{9}\cdot\frac{6}{8} \qquad (1a)$$

Man erkennt, dass die Koeffizienten der Summanden gleich sind, auch wenn ihre Reihenfolge vertauscht ist. Die Einzelwahrscheinlichkeiten für jeden der Zweige sind also gleich groß und können zusammengefasst werden:

$$g(x=1)=3\cdot\frac{3\cdot 7\cdot 6}{10\cdot 9\cdot 8} \qquad (1b)$$

Die Gleichung (1b) besteht aus 4 Teilen:

1) Die Koeffizienten $10\cdot 9\cdot 8$ unter dem Bruchstrich stehen für die Anzahl der Teile, die bei jedem Zug gezogen werden können. Beim ersten Zug sind dies $N=10$, beim zweiten Zug ist es ein Teil weniger, da schon ein es gezogen wurde und so weiter mit insgesamt $n=3$ Koeffizienten.
   Um dies für alle N und n möglichst kurz auszudrücken, verwendet man die Schreibweise „N!“ bzw. „(N-n)!“. N!, gesprochen N-Fakultät, ist das Produkt aller ganzen Zahl von 1 bis N. (N-n)! unter dem Bruchstrich kürzt überflüssige Zahlen wieder heraus. Die Koeffizienten $10\cdot 9\cdot 8$ können also wie folgt geschrieben werden:

$$\frac{N!}{(N-n)!}=\frac{10!}{(10-3)!}=\frac{\cancel{1}\cdot\cancel{2}\cdot\cancel{3}\cdot\cancel{4}\cdot\cancel{5}\cdot\cancel{6}\cdot\cancel{7}\cdot 8\cdot 9\cdot 10}{\cancel{1}\cdot\cancel{2}\cdot\cancel{3}\cdot\cancel{4}\cdot\cancel{5}\cdot\cancel{6}\cdot\cancel{7}}=8\cdot 9\cdot 10 \qquad (2a)$$

2) Die Koeffizienten $6\cdot 7$ über dem Bruchstrich stehen für die Anzahl der guten Teile, die bei jedem Zug gezogen werden können. Beim ersten Zug sind dies $N-d=10-3=7$, beim zweiten Zug wieder eines weniger und so weiter mit insgesamt $n-x=3-1=2$ Koeffizienten. 2 Koeffizienten deshalb, weil bei unter Stichprobe mit $n=3$ Teilen nur 2 Gutteile sein können, wenn man den Fall $x=1$ Ausschussteile untersucht.
   Die Koeffizienten können wieder mit Fakultäten geschrieben werden:

$$\frac{(N-d)!}{([N-d]-[n-x])!}=\frac{(10-3)!}{([10-3]-[3-1])!}=\frac{\cancel{1}\cdot\cancel{2}\cdot\cancel{3}\cdot\cancel{4}\cdot\cancel{5}\cdot 6\cdot 7}{\cancel{1}\cdot\cancel{2}\cdot\cancel{3}\cdot\cancel{4}\cdot\cancel{5}}=6\cdot 7 \qquad (2b)$$

3) Der Koeffizient 3 <u>über</u> dem Bruchstrich steht für die Anzahl der Ausschussteile, die bei jedem Zug gezogen werden können. Beim ersten Zug sind dies $d=3$ und es sind $x=1$ Koeffizienten. Für den allgemeinen Fall können die Koeffizienten geschrieben werden:

$$\frac{d!}{(d-x)!}=\frac{3!}{(3-1)!}=\frac{\cancel{1}\cdot\cancel{2}\cdot 3}{\cancel{1}\cdot\cancel{2}}=3 \qquad (2c)$$

4) Der Koeffizient 3 <u>vor</u> dem Bruchstrich ist der so genannte Binomialkoeffizient und gibt an, wie viele Wege durch den Wahrscheinlichkeitsbaum zum Ergebnis $x=1$ führen. Auch den Binomialkoeffizienten kann man mit Fakultäten berechnen (ohne Erklärung):

$$\frac{n!}{x!\cdot(n-x)!}=\frac{3!}{1!\cdot(3-1)!}=\frac{\cancel{1}\cdot\cancel{2}\cdot 3}{1\cdot\cancel{1}\cdot\cancel{2}}=3 \qquad (2d)$$

Um die Gleichung (1b) zu verallgemeinern, kann man die Bestandteile (2a) bis (2d) zusammenwerfen.

$$g(x)=\frac{n!}{x!\cdot(n-x)!}\cdot\frac{\dfrac{d!}{(d-x)!}\cdot\dfrac{(N-d)!}{([N-d]-[n-x])!}}{\dfrac{N!}{(N-n)!}} \qquad (3a)$$

Ein bisschen umgeformt

$$g(x)=\frac{\dfrac{d!}{x!\cdot(d-x)!}\cdot\dfrac{(N-d)!}{(n-x)!\cdot([N-d]-[n-x])!}}{\dfrac{N!}{n!\cdot(N-n)!}} \tag{4a}$$

und die Schreibweise „N über n“ eingeführt

$$\binom{N}{n}=\frac{N!}{n!\cdot(N-n)!} \tag{4b}$$

ergibt die übliche Schreibweise für die Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung:

$$g(x)=\frac{\binom{d}{x}\cdot\binom{N-d}{n-x}}{\binom{N}{n}} \tag{4c}$$

Berechnung der binomialen Verteilung

Wie groß sind die Wahrscheinlichkeiten g(x), aus einem Paket mit N = 100 Schrauben, davon d = 8 fehlerhaft, in einer Stichprobe vom Umfang n = 5 zufällig genau x = 0, 1, 2, 3, 4 oder 5 fehlerhafte Schrauben zu finden, wenn jede gezogene Schraube sofort <u>zurückgelegt und untergemischt</u> wird ?

Die Wahrscheinlichkeitsfunktion der binomialen Verteilung ist wesentlich einfacher als die der hypergeometrischen Verteilung, weil sich die Wahrscheinlichkeiten nach erfolgtem Zug nicht ändern. Deshalb ist hier die Gleichung vorgegeben und anschließend am Beispiel x = 1 erklärt.

$$g(x)=\frac{n!}{x!\cdot(1-x)!}\cdot p^x\cdot(1-p)^{n-x} \tag{5a}$$

bzw.

$$g(x)=\binom{n}{x}\cdot p^x\cdot(1-p)^{n-x} \tag{5b}$$

Die Gleichung (5) besteht aus 3 Teilen:

1) Im ersten Teil erkennt man wieder den Binomialkoeffizienten, der angibt, wie viele Wege es zu x = 1 fehlerhafte Teile gibt. Im gegebenen Fall sind es 5 Wege, da das fehlerhafte Teil im ersten, zweiten .. oder fünften Zug gezogen werden kann. Gerechnet sieht es aus wie folgt:

$$\frac{n!}{x!\cdot(n-x)!}=\frac{5!}{1!\cdot(5-1)!}=\frac{\cancel{1}\cdot\cancel{2}\cdot\cancel{3}\cdot\cancel{4}\cdot 5}{1\cdot\cancel{1}\cdot\cancel{2}\cdot\cancel{3}\cdot\cancel{4}}=5 \tag{6a}$$

2) Der zweiten Teil wird die Wahrscheinlichkeit für einen Zweig berechnet. Mit Zahlen sollte die Struktur leicht erkennbar sein:

$$p^x\cdot(1-p)^{n-x}=(8\,\%)^1\cdot(92\,\%)^4=0{,}08^1\cdot 0{,}92^4=0{,}057 \tag{6b}$$

Bevor es elektronische Rechenhilfen gab, hat man Wahrscheinlichkeiten der binomialen Verteilungen oft grafisch mit Hilfe des Larson-Nomogrammes gelöst.

Binomiale Verteilung als Näherung für die hypergeometrische Verteilung

Bei der hypergeometrischen Verteilung ändern sich die Einzelwahrscheinlichkeiten wie beim Lotto, wo mit jedem Zug eine Kugel entnommen wird. Bei der binomialen Verteilung bleiben die Einzelwahrscheinlichkeiten gleich wie beim Roulette, bei dem die letzten Würfe allen Gerüchten zum Trotz nichts über die zukünftigen Würfe verraten. Dieser Unterschied bewirkt natürlich auch unterschiedliche Ergebnisse für binomiale und hypergeometrische Verteilungen.

Wenn aber große Gesamtmengen N mit kleinen Stichprobenumfängen n geprüft werden, fällt die Änderung der Wahrscheinlichkeit (z.B. 1/999 statt 1/1000) kaum ins Gewicht, sodass die Berechnung mit der binomialen und mit der hypergeometrischen Wahrscheinlichkeitsfunktion praktisch gleiche Ergebnisse liefert.

Da binomiale Verteilungen ohne Rechenhilfe wesentlich leichter zu berechnen sind als hypergeometrische Verteilungen, nutzt man die Ähnlichkeiten der Ergebnisse aus und berechnet Verteilungen mit der binomialen Formel, wenn der Stichprobenumfang n nicht größer als 10% der Gesamtmenge N ist.

$$n<0{,}1\cdot N$$