

Wilfried Rohm

wrohman@aon.at

Simulation einer Binomialverteilung



- Mathematische / Fachliche Inhalte in Stichworten:

Binomialverteilung, Simulation einer Stichprobe

- Kurzzusammenfassung

Ausgehend von einer praktischen Problemstellung aus dem Gebiet der Qualitätssicherung wird das Entstehen der Binomialverteilung simuliert und mit den berechneten Werten verglichen. Es erfolgen auch Hinweise auf den Einsatz im Unterricht. In der Literaturangabe wird auf einen AMMU-File verwiesen, in dem der Unterrichtseinsatz noch näher erläutert wird.

- Didaktische Überlegungen / Zeitaufwand:

Das Beispiel eignet sich erfahrungsgemäß sehr gut als Einstieg in statistische Problemstellungen. Ich führe es durch, noch bevor irgendwelche Kenntnisse über Verteilungsfunktionen vorhanden sind.

- Lehrplanbezug (bzw. Gegenstand / Abteilung / Jahrgang):

Angewandte Mathematik: Statistische Methoden des Qualitätsmanagements

- Mathcad-Version:

Mathcad 2001

- Literaturangaben:

**AMMU-Artikel des Autors in AMMU-15 (www.ammu.at)
"Ein Einstiegsbeispiel zur Wahrscheinlichkeitsrechnung und Statistik",**



Die folgende Fragestellung stammt aus dem Bereich Stichprobensysteme und erscheint geeignet, in das statistische Denken und in Grundlagen der Wahrscheinlichkeitsrechnung einführen soll:

In einer Firma werden Bauteile gefertigt, der Ausschußanteil liegt gemäß den laufenden Protokollen ziemlich konstant bei 2%. In der Eingangsprüfung einer Firma, die ständig große Mengen dieser Bauteile bezieht, soll mittels regelmäßig durchgeführter Stichprobenprüfungen sichergestellt werden, dass keine Lieferungen mit einem höheren Ausschußanteil als 2% angenommen werden.

Es wird vereinbart: Stichprobenumfang $n = 100$

Bei wieviel schlechten Bauteilen in der Stichprobe würdest Du als Prüfer annehmen, daß der Ausschußanteil "ziemlich sicher" größer geworden ist?

Bestätige bzw. widerlege die Sinnhaftigkeit Deiner Vermutung durch entsprechende Simulationen. Ausserdem sollen die (auf Grund der Simulationen) ermittelten relativen Häufigkeiten für das Auftreten von $x = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 \dots$ schlechten Bauteilen in der Stichprobe mit den "theoretischen" Wahrscheinlichkeiten veralichen werden.

Nach einer entsprechenden "händischen" Simulation könnte diese mittels Software erfolgen:

$n := 100$ $p := 0.02$

$X := \text{rbinom}(N, n, p)$ N binomialverteilte Zufallszahlen (N wird neben der Grafik [unten] global festgelegt!)

$N = 50$

Die folgende Tabelle zeigt die ersten 20 Versuche und als Ergebnis, wieviel "schlechte" in der "Lieferung" jeweils gefunden wurden

$$X^T =$$

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0	1	2	2	2	0	2	1	5	4	3	3	4	1	3	2	1	2	2	1

Nun sollen die Ergebnisse zusammengefaßt und mit dem theoretischen Ergebnis der Binomialverteilung mit den Parametern n und p verglichen werden. Dazu wird mit dem "hist"-Befehl eine Klasseneinteilung für die Anzahl der schlechten in der Stichprobe ($j=0,1,\dots,7$) durchgeführt.

theoretische Wahrscheinlichkeiten: $g_{BI}(x) := \text{dbinom}(x, n, p)$

$j := 0..n$ $\text{bezug}_j := j$

$h := \frac{\text{hist}(\text{bezug}, X)}{N}$

Klasseneinteilung bezogen auf $j=0..n$

Der Vektor h_j enthält also die relativen Häufigkeiten, wie oft j schlechte unter den N Versuchen vorgekommen sind

$$h^T =$$

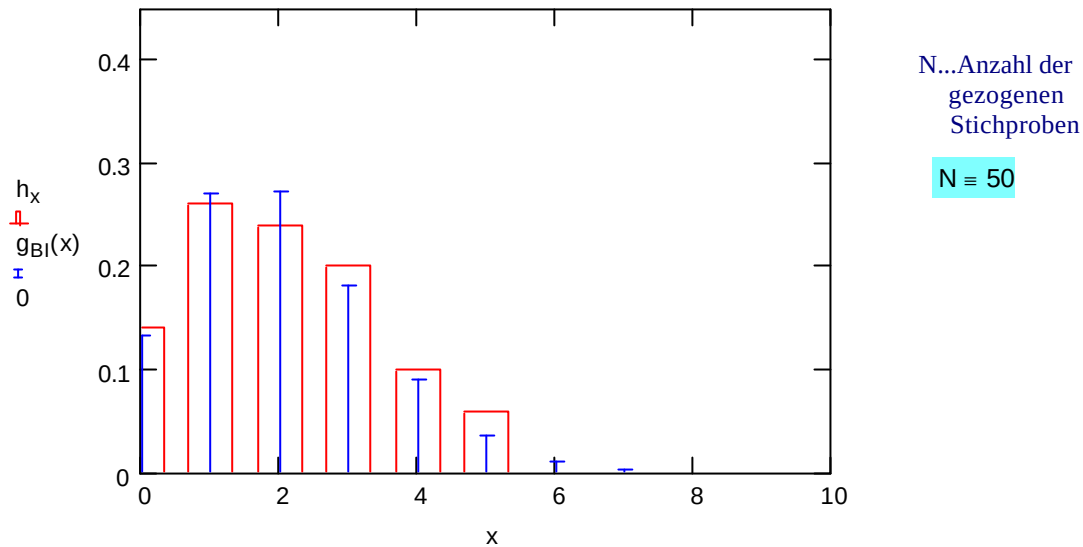
	0	1	2	3	4	5	6	7	8	9	10
0	0.14	0.26	0.24	0.2	0.1	0.06	0	0	0	0	0

Tabellarischer Vergleich zwischen theoretischen und simulierten Werten

$h(j) := h_j$ nur aus optischen Gründen

$k := 0..10$

k =	(k) =	$g_{BI}(k) =$
0	0.14	0.133
1	0.26	0.271
2	0.24	0.273
3	0.2	0.182
4	0.1	0.09
5	0.06	0.035
6	0	0.011
7	0	$3.13 \cdot 10^{-3}$
8	0	$7.426 \cdot 10^{-4}$
9	0	$1.549 \cdot 10^{-4}$
10	0	$2.877 \cdot 10^{-5}$

Zeichnerischer Vergleich zwischen theoretischen und simulierten Werten $x := 0..10$ **HINWEISE zum Einsatz im Unterricht: (Region anklicken)**

Man beachte: Die Schüler werden hier aufgefordert, ohne jede wahrscheinlichkeitstheoretische Vorbildung sozusagen "aus dem Bauch" heraus auf die oben gestellte Frage zu antworten. Die üblichen Antworten offenbaren zum Teil eine völlige Hilflosigkeit, wie man an diese Frage herangehen könnte. Eine typische spontane Antwort ist daher:

"Zwei Prozent von 100 sind 2. Daher nehme ich bei mehr als 2 schlechten Stück bereits an, dass der Ausschußanteil größer als 2% ist."

Im Gegensatz dazu treten auch Meinungen auf, die sehr "vorsichtig" sind und erst bei 10-20 (!) schlechten Bauteilen in der Stichprobe wirklich von einem höheren Ausschußanteil als 2% ausgehen möchten. Es gibt meist nur wenige Schüler, die den Einfluß des Zufalls bei dieser Fragestellung wirklich berücksichtigen und noch weniger können ihn einigermaßen richtig einschätzen!

Bevor die Computersimulation durchgeführt wird, empfiehlt sich daher auch, eine "händische" Simulation vorzunehmen. Ein geeignetes Simulationsgerät sind z.B. Kugelkästen, wie sie am TGM hergestellt werden. Diese enthalten 1000 verschiedenfarbige Kugeln, die ein Simulieren unterschiedlicher Fehleranteile erlauben und durch die gleichzeitige Entnahme von 100 Kugeln auch eine einigermaßen zeitsparende Simulation ermöglichen.

Als Variante sind auch "Kugelkästen" mit 2% irgendwie markierten "Kugeln" denkbar, aus denen (je nach Gesamtanzahl der Kugeln) mit Zurücklegen (oder ohne Zurücklegen bei sehr vielen Kugeln) gezogen wird. Allerdings steigt dann der Zeitaufwand enorm, wenn man so viele Daten erzeugen will, damit eine einigermaßen aussagekräftige Auswertung der Daten möglich ist.

Ich führe in der Regel 50 Versuche durch und lasse die Schüler ein Häufigkeitsdiagramm erstellen. Anschließend stelle ich die Daten den "theoretischen Wahrscheinlichkeiten" gegenüber (mit dem Hinweis, dass wir diese ohnehin bald selbst berechnen werden können). Erst anschließend führe ich die Simulation am Computer durch.

Zur Computersimulation:

- * Neue Simulation mit "**Rechnen-Arbeitsblatt berechnen**" oder: den Cursor auf das Eingabefeld für N stellen und **F9** drücken.
- * Die Simulation zeigt, dass für kleines N (z.B. N=50) ein großer Unterschied zwischen den relativen Häufigkeiten und den theoretischen Wahrscheinlichkeiten besteht. Für größer werdendes N beobachtet man eine immer bessere Übereinstimmung (**Gesetz der großen Zahlen**)



Interpretation des Ergebnisses in Bezug auf die gestellte Frage

Für die Beantwortung der Frage, ab wieviel schlechten man annimmt, dass der Ausschußanteil "ziemlich sicher" größer geworden ist, ist es sinnvoll, die Werte $P(x) = 1 - G(x-1)$ auszurechnen. Dies ist die Wahrscheinlichkeit, dass die **Anzahl der schlechten Stücke in der Stichprobe größer oder gleich x** ist (bei unverändertem p).

$$G_{BI}(x) := \sum_{k=0}^x g_{BI}(k)$$

$G_{BI}(x)$ kann auch mit der eingebauten Funktion `pbinom(x, n, p)` ermittelt werden.

$$P(x) := \begin{cases} 1 & \text{if } x = 0 \\ 1 - G_{BI}(x-1) & \text{otherwise} \end{cases}$$

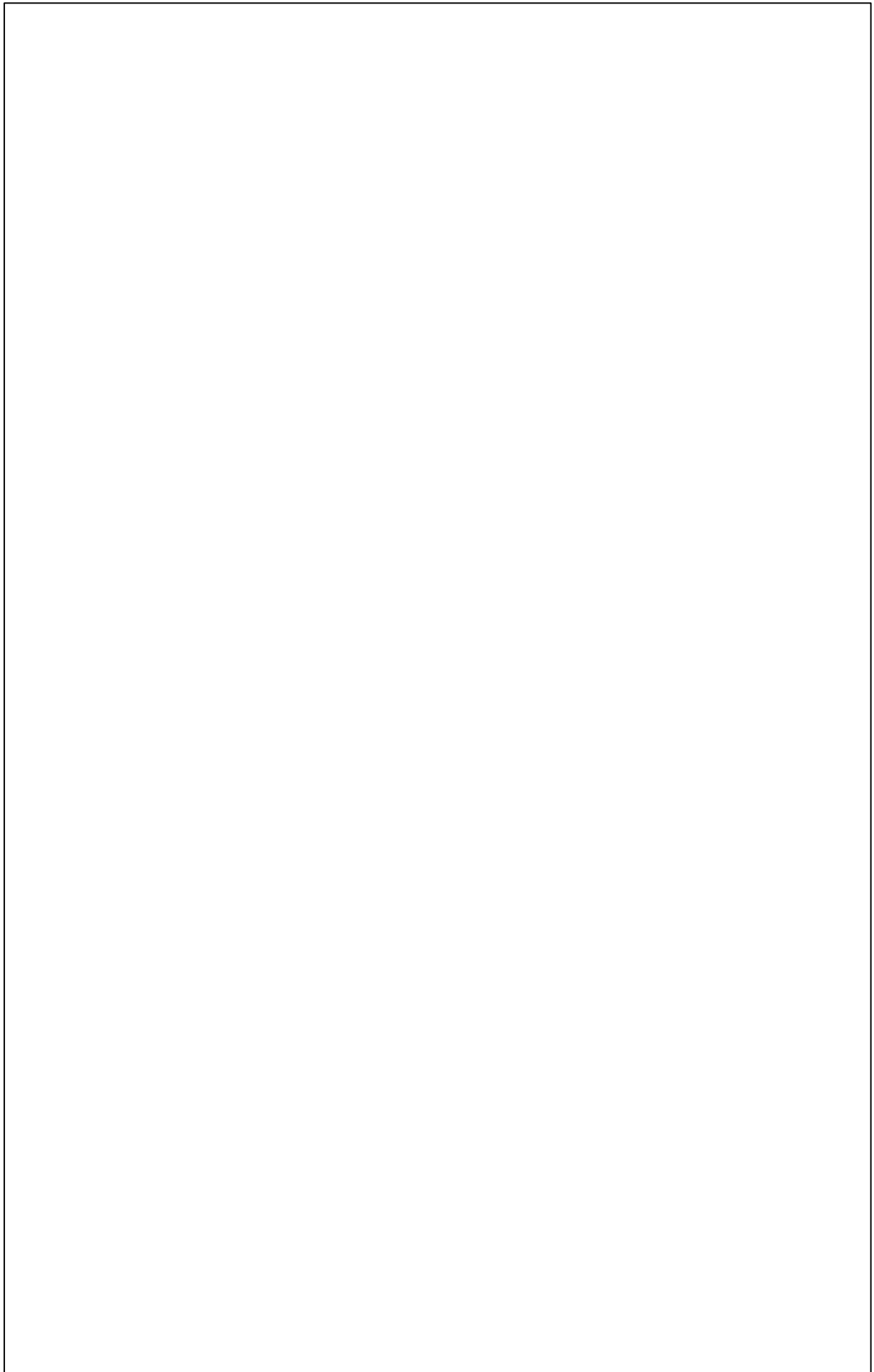
$x := 0..10$

$x = \quad G_{BI}(x) = \quad P(x) =$

0	0.13262	1
1	0.40327	0.86738
2	0.67669	0.59673
3	0.85896	0.32331
4	0.94917	0.14104
5	0.98452	0.05083
6	0.99594	0.01548
7	0.99907	0.00406
8	0.99981	0.00093
9	0.99997	0.00019
10	0.99999	0.00003

Beispiel: Es werden 7 "schlechte" in der Stichprobe gefunden. Die Wahrscheinlichkeit, dass dieser Wert bei $p=2\%$ auftritt, ist höchstens $P(7)=0,4\%$.

Wenn ich daher annehme, dass "in Wirklichkeit" p größer als 2% ist, kann es zwar sein, dass ich mich irre - doch diese "Irrtumswahrscheinlichkeit" ist kleiner als 1/2 Prozent!!



[illegible]

