

STV – Statistische Verteilung

Auswertung

Nachbesserung

Yudong Sun, Andreas Artz
Gruppe F2

3. März 2020

Teilversuch 1: Generierung einer Binomialverteilung mit dem Galton-Brett

Annahme: Die Verteilung der Kugeln ist bekannt

Bei jeder Weggabelung kann die Kugel entweder links oder rechts fallen. Damit die Kugel in den 0. Kanal fällt muss sie bei jeder Weggabelung im Galton-Brett nach links fallen, also insgesamt 10 mal.

$$\begin{aligned} p(\text{Eine Kugel in Kanal } 0) &= \binom{10}{0} \times \frac{1}{2}^{10} \times \frac{1}{2}^0 \\ &= \frac{1}{2^{10}} = 0,0977\% \end{aligned}$$

Damit die Kugel in den 5. Kanal, der genau in der Mitte liegt, fällt, muss sie bei den Weggabelungen insgesamt 5 mal nach links und 5 mal nach rechts fallen. Die Reihenfolge ist dabei egal.

$$\begin{aligned} p(\text{Eine Kugel in Kanal } 5) &= \binom{10}{5} \times \frac{1}{2}^5 \times \frac{1}{2}^5 \\ &= 24,6\% \end{aligned}$$

Annahme: Die Verteilung der Kugeln ist unbekannt

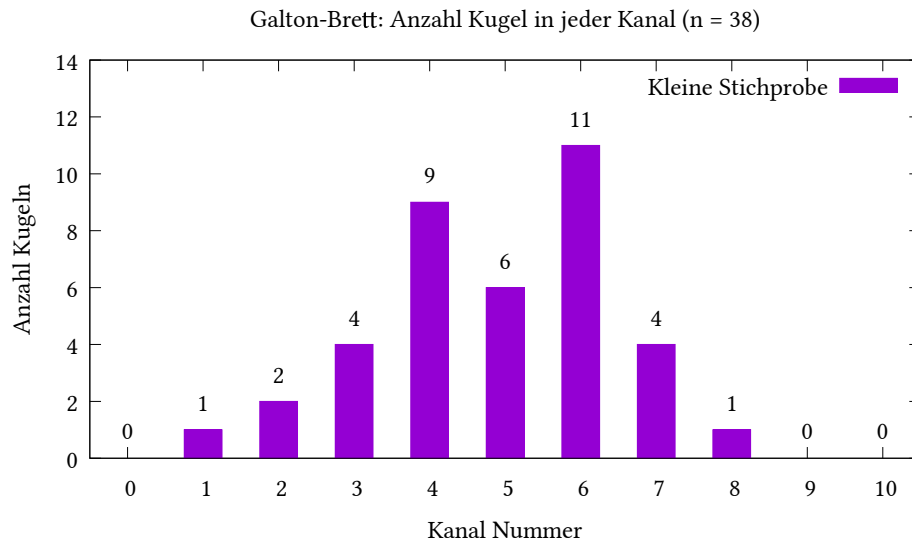


Abbildung 1: Histogramm der kleinen Statistik

Die Mittelwert und Standardabweichung wird mithilfe eines Python-Skripts berechnet (Siehe Appendix A):

Mittelwert	4,8684
Standardabweichung	1,5968

Man kann vom Histogramm nicht auf eine Binomialverteilung schließen. Zumindest ist das Histogramm nicht symmetrisch, und das Maximum ist nicht beim Mittelwert.

Diskussion

Die Häufigkeitsverteilungen sollten sich mit wachsender Stichprobengröße der Wahrscheinlichkeitsfunktion der Binomialverteilung annähern. Das Ergebnis zeigt dies, da die Verteilung der großen Statistik deutlich symmetrischer ist als die der mittleren Statistik. Zudem ist der Mittelwert der großen Statistik näher am Erwartungswert.

Mit $N = 10$ und $p = 1/2$ lässt sich die theoretische Erwartungswert und Standardabweichung σ einer binomischen Verteilung wie folgt berechnen:

$$E(x) = Np = 10 \times \frac{1}{2} = 5 \quad (1)$$

$$\sigma = \sqrt{Np(1-p)} = \sqrt{10 \left(\frac{1}{2}\right) \left(1 - \frac{1}{2}\right)} = 1,581 \quad (2)$$

Die empirische Mittelwert und Standardabweichung liegen trotz des Histogrammes nahe an die theoretischen Werten.

Diskussion der Histogramme der mittleren und großen Statistik

Stichprobe	\bar{x}	s
Mittel	4,832	1,6327
Groß	4,915	1,6237

Im Vergleich zu der mittleren Statistik liegen der Mittelwert und Standardabweichung der großen Statistik näher an den theoretischen Werten. Die Verbesserung ist aber numerisch gering und könnte theoretisch zu normaler Abweichungen beim Experiment zurückgeführt werden. Die zwei Histogrammen haben einigermaßen auch die gleiche Form. Insgesamt ist die Erhöhung der Anzahl der Kugeln nach 2560 kaum eine bessere Annäherung im Vergleich zu dem Versuch mit nur 256 Kugeln.

Vergleich alle drei Statistiken

Wir bezeichnen \bar{x} als der Mittelwert und s als die empirische Standardabweichung einer Stichprobe. Der Fehler des Mittelwerts ist dann gegeben durch s/\sqrt{n} , wobei n die Größe der Stichprobe ist.

Stichprobe	n	\bar{x}	s	Fehler
Klein	38	4,868	1,5968	0,259
Mittel	256	4,832	1,6327	0,102
Groß	2560	4,915	1,6237	0,0321

Mit wachsende Anzahl der Kugeln nähert sich der Mittelwert immer besser an den theoretischen Erwartungswert. Der Fehler nimmt auch ab. Theoretisch ist das zu erwarten. Bei beliebig großer Stichprobe sollten \bar{x} und $E(x)$ sich beliebig nah annähern.

Bei der Standardabweichung ist aber die kleine Stichprobe hier die beste Annäherung. Das ist vermutlich wegen zufälliger experimenteller Schwankungen. Zumindest haben wir schon diskutiert, dass das Histogramm keine gute Annäherung ist.

Teilversuch 3: Zentrale Grenzwertsatz

Die Standardabweichung aus Teilversuch 2 wird mit dem vorgenannten Python Skript berechnet.¹

Teilversuch	n	\bar{x}	s	s^2	Verteilung
2	50	2.16	1.1314	1.28	Poisson
2	100	2,36	1,3681	1,872	Poisson
3	100	77,54	8,9594	80.271	Normal

wobei n die Messungen je 2 s ist.

Variante 1

Diese Variante wurde wegen zeitlicher Gründe nicht untersucht.

Variante 2

Der zentrale Grenzwertsatz besagt, dass die Verteilung jeder beliebig verteilten Zufallsvariable in einer Stichprobe aus n unabhängigen Messungen sich mit wachsendem n einer Normalverteilung annähern wird.

In diesem Fall, wird das ROI bzw. Energieintervall im Vergleich zu vorher vergrößert. Die Zufallsvariable in diesem Fall ist die Anzahl der Impulse je 2 Sekunden. Wenn man diese Anzahl von alle Kanäle summiert, hat man nun durchschnittlich 77,54 Impulse pro Zeiteinheit, statt durchschnittlich nur 2,36 Impulse pro Zeiteinheit im Teilversuch 2. Das entspricht die Situation, indem man ein Poisson-verteilte Variable mehrmals messen, sodass man laut ZGS eine Normalverteilung bekommen.

Für 100 Messungen je 2 Sekunden sollen wir theoretisch beim Teilversuch 3 die folgende Werte bekommen:

$$E(x) = n\mu \iff n = \frac{77,54}{2,36} = 32.86 \quad (3)$$

$$\text{Var}(x) = n\sigma^2 = \frac{77,54}{2,36} \times 1,872 = 61,506 \quad (\text{Vgl. } s^2 = 80,271) \quad (4)$$

Diese zwei Varianzen sind zumindest in der gleichen GröÙeordnung. Die Abweichungen liegt vermutlich daran, dass der Skalierungsfaktor n von dem Mittelwert abgeschätzt wurden und nicht den echten Wert von n entspricht.

¹Mit dieser neuen Berechnung ist das Problem beim Ablesen im MATLAB während des Versuchs auch bekannt gemacht: Die Dateien sind im UTF-16LE und nicht UTF-8 codiert.

A Quellcode zur Auswertung von Teilversuch 1

gnuplot Code für Abbildung 1

```

1  set term epscairo font 'Linux Libertine, 14'
2  set output "kleine.eps"
3
4  set xrange [-0.5:10.5]
5  set yrange [0:14]
6  set xtics 1
7  set title "Galton-Brett: Anzahl Kugel in jeder Kanal (n = 38)"
8  set ylabel "Anzahl Kugeln"
9  set xlabel "Kanal Nummer"
10
11 set boxwidth 0.5
12 set style fill solid
13 plot "kleine.dat" w boxes title "Kleine Stichprobe", "" u 1:2:2 with labels
    ↪ offset char 0,1 notitle

```

Python Code für die Rechnung der Mittelwert und Standardabweichung:

```

1  #!/usr/bin/env python3
2
3  import numpy as np
4
5  with open("kleine.dat") as f:
6      A = np.loadtxt(f, dtype=int)
7      kanale = A[:,0] # Erste Spalte
8      anzahl = A[:,1] # Zweite Spalte
9
10 # Mittelwert
11 mit = np.dot(kanale, anzahl)/np.sum(anzahl)
12
13 # Standardabweichung
14 sab = np.sqrt((1/(np.sum(anzahl) - 1))*np.dot((kanale - mit)**2, anzahl))
15
16 print("Mittelwert: ", mit)
17 print("Standardabweichung: ", sab)

```

kleine.dat:

```

1  0    0
2  1    1
3  2    2
4  3    4
5  4    9
6  5    6
7  6   11
8  7    4
9  8    1
10 9    0
11 10   0

```