Kurzprotokoll „Statistik“
05.04.2017
Kassandra Kunz, Helmut Hörner

1. Versuch „Messfehler“

Die simulierten Messdaten lagen im Bereich zwischen 185 und 3692. Daraus ergeben sich 3508 Klassen der Breite 1. Da der Wert von 3508 ungünstig für andere Unterteilungen ist, wurden letztlich 3509 „elementare“ Klassen der Breite 1 gebildet; und in weiterer Folge Klassen der Breite 11, 29 und 121.

Die entsprechenden Histogramme sind in den folgenden vier Abbildungen zu sehen. Man erkennt, dass sich die Histogramme immer weiter der Dichtefunktion (einer Gaußverteilung) annähern, je größer die Klassenbreite gewählt wird.

Offensichtlich ist weder eine zu feine, noch eine zu grobe Unterteilung sinnvoll. Im Fall der zu feinen Unterteilung erhält man stark streuenden Ergebnisse; im anderen Fall (wenn man z.B. nur zwei oder gar nur mehr eine Klasse hätte), wäre die Aussagekraft stark beschränkt.

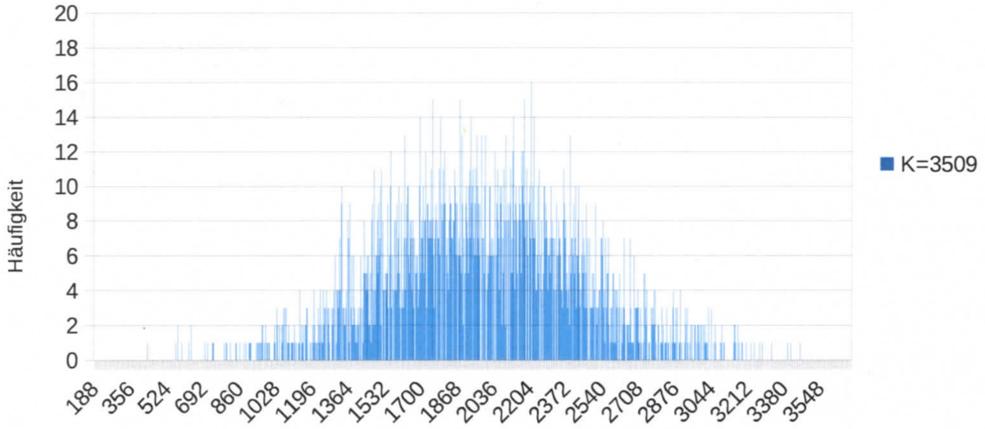
Wenn man zu den Werten aus dem Datensatz eine feste Zahl hinzuaddiert, ändert sich der Mittelwert (wie erwartet) um genau diesen additive Konstante. Die Standardabweichung bleibt bleibt unverändert.

Wenn alle Werte aus dem Datensatz durch eine feste Zahl a dividiert, wird auch der Mittelwert (wie erwartet) durch diese Zahl dividiert. Die Standardabweichung ändert sich um $1/a^2$.

Um aus den gegebenen Daten eine Standard-Normalverteilung zu erhalten, muss zunächst von allen Messwerten der Mittelwert abgezogen werden (da die Standard-Normalverteilung ihr Maximum bei $x=0$ hat). Anschließend müssen die Messwerte mit einem Faktor multipliziert werden, so dass die Fläche unter der Verteilungskurve=1 wird.

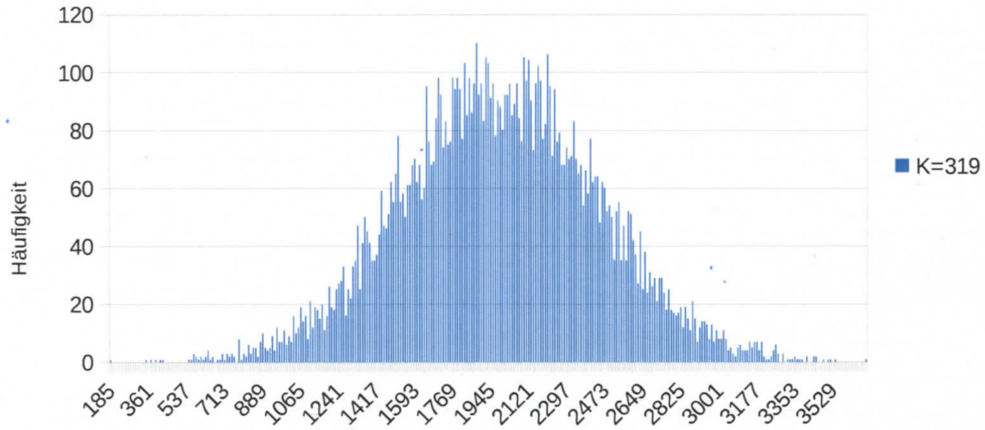
Mittelwert: 1991, Standardabw.: 460,3

3509 Klassen der Breite 1



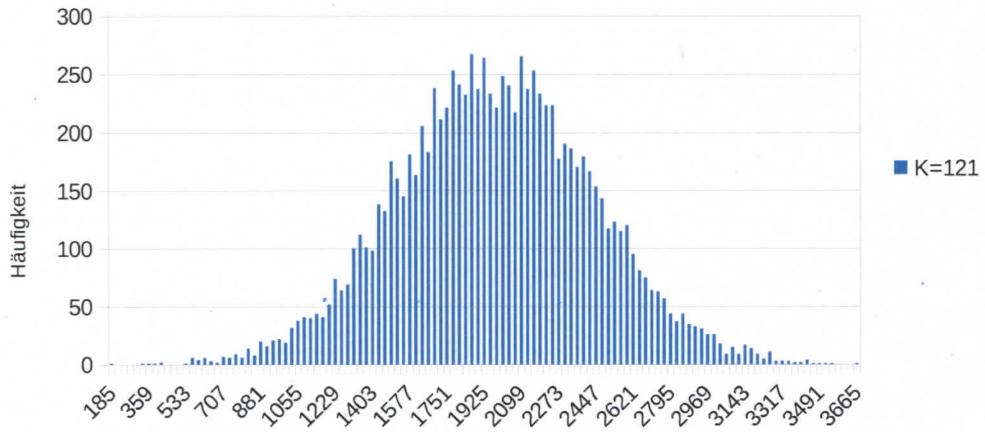
Mittelwert: 1991; Standardabweichung 460,3

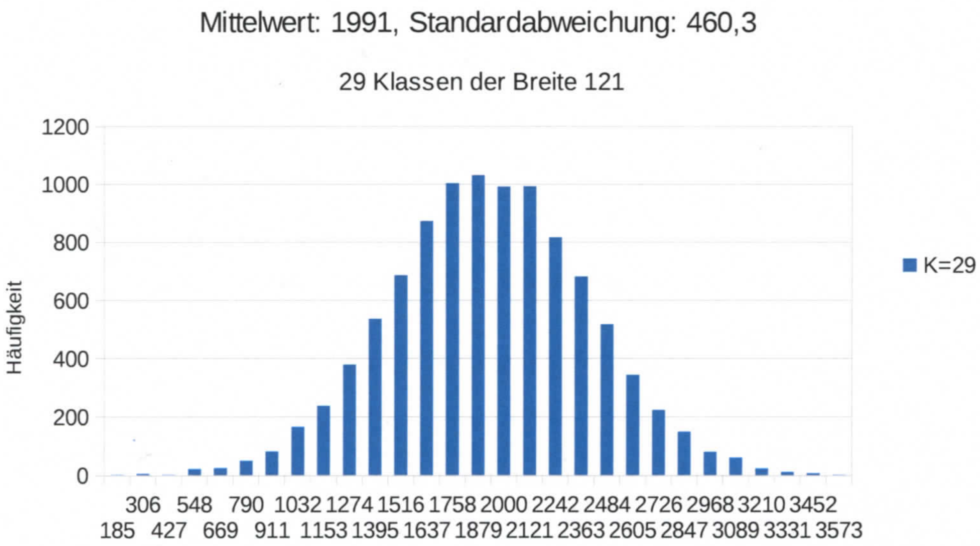
319 Klassen der Breite 11



Mittelwert: 191; Standardabweichung: 460,3

121 Klassen der Breite 29




Mittelwert: 1991, Standardabweichung: 460,3
29 Klassen der Breite 121
Häufigkeit
1200
1000
800
600
400
200
0
306 548 790 1032 1274 1516 1758 2000 2242 2484 2726 2968 3210 3452
185 427 669 911 1153 1395 1637 1879 2121 2363 2605 2847 3089 3331 3573
K=29

2. Versuch „Würfel“

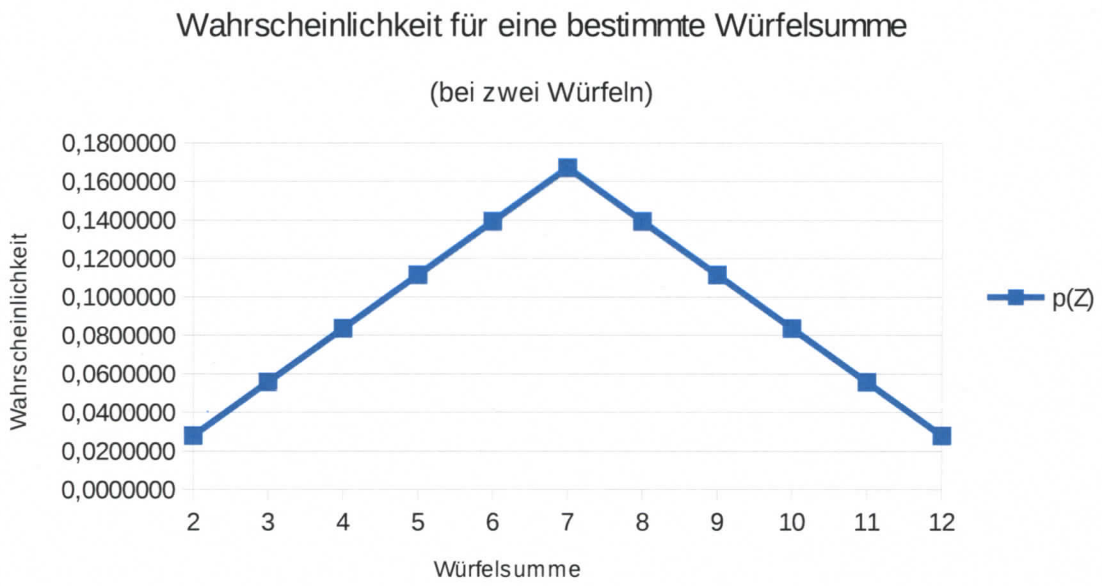
Es wurde der wiederholte Wurf mit zwei Würfeln simuliert, wobei die Würfelsumme als Zufallsvariable definiert wurde. Dies führt zu einer Laplaceverteilung.

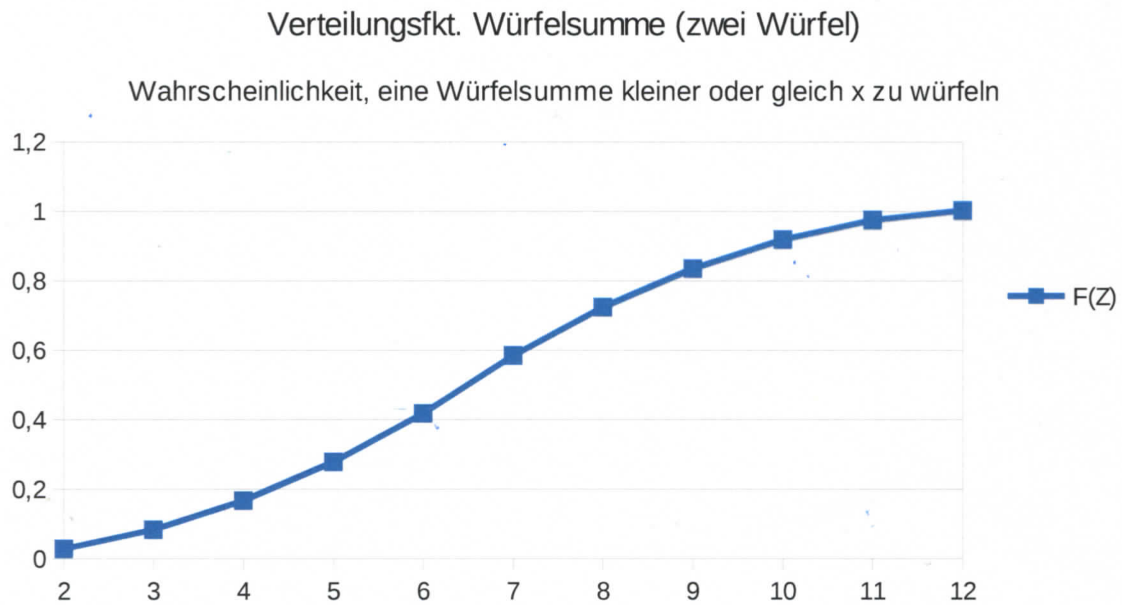
Die Werte für f(W1,W2) können aus der folgenden Tabelle entnommen werden.

Es ergeben sich folgende Werte für Wahrscheinlichkeit, Häufigkeit, Erwartungswert, Varianz und Verteilungsfunktion:

Z=f(W1,W2)	p(Z)	Z*p(Z)	p(Z)*(<Z>-Z)^2	F(Z)
2	0,0277778	0,0555556	0,6944444	0,0277778
3	0,0555556	0,1666667	0,8888889	0,0833333
4	0,0833333	0,3333333	0,7500000	0,1666667
5	0,1111111	0,5555556	0,4444444	0,2777778
6	0,1388889	0,8333333	0,1388889	0,4166667
7	0,1666667	1,1666667	0,0000000	0,5833333
8	0,1388889	1,1111111	0,1388889	0,7222222
9	0,1111111	1,0000000	0,4444444	0,8333333
10	0,0833333	0,8333333	0,7500000	0,9166667
11	0,0555556	0,6111111	0,8888889	0,9722222
12	0,0277778	0,3333333	0,6944444	1,0000000
SUMME	**1,0000000**	**7,0000000**	**5,8333333**	
		Erwartungsw.	**Varianz**	

Die folgenden Abbildungen zeigen den Verlauf für p(Z) und für die Verteilungsfunktion F(Z).


Wahrscheinlichkeit für eine bestimmte Würfelsumme
(bei zwei Würfeln)
Wahrscheinlichkeit
0,1800000
0,1600000
0,1400000
0,1200000
0,1000000
0,0800000
0,0600000
0,0400000
0,0200000
0,0000000
2 3 4 5 6 7 8 9 10 11 12
Würfelsumme
p(Z)


Verteilungsfkt. Würfelsumme (zwei Würfel)
Wahrscheinlichkeit, eine Würfelsumme kleiner oder gleich x zu würfeln
1,2
1
0,8
0,6
0,4
0,2
0
2 3 4 5 6 7 8 9 10 11 12
F(Z)

Es wurde in 5, 10, 25, 50, 100, 250, 500, 1000, 2500 und 5000 Schritten Durchführungen des Zufallsexperiments simuliert.

Wie in der folgenden Grafik zu erkennen ist, konvergiert der Mittelwert bei hinreichend großer Wiederholung (hier ab n=250) letztlich gegen Erwartungswert.

Zuletzt wurde ein Chi-Quadrat-Test für eine Stichprobe n=100 durchgeführt. Die folgende Tabelle zeigt das Ergebnis:

Z=f(W1,W2)	Abs. Häufigkeit H(Z)	Kalk. Häufigkeit e(Z)	(H(Z)-e(Z))^2/e(Z)	
2	1	2,7777778	1,1377777778	
3	5	5,5555556	0,0555555556	
4	6	8,3333333	0,6533333333	
5	8	11,1111111	0,8711111111	
6	14	13,8888889	0,0008888889	
7	15	16,6666667	0,1666666667	
8	17	13,8888889	0,6968888889	
9	11	11,1111111	0,0011111111	
10	9	8,3333333	0,0533333333	
11	9	5,5555556	2,1355555556	
12	5	2,7777778	1,7777777778	
SUMME	**100**	**100**	**7,55**	**χ^2**
MW	7,540		**19,68**	**X^2-0.95 Fraktil**
Sigma	2,435			

Man erkennt, dass der aus der Stichprobe berechnet Wert von χ^2 kleiner ist als der tabellierte Wert von χ^2 für γ=0,95. Das bedeutet, dass die Hypothese, dass die Daten aus der angenommenen Verteilung stammen, mit 95%iger Sicherheit nicht verworfen werden kann.

Dies ist jedoch kein geeignetes Mittel, um auf die Zufälligkeit von W1, W2 bzw. Z zu schließen, da Tests auf „echte“ Zufälligkeit eine wesentlich komplexere Herausforderung darstellen.

Daten

	n=5	n=10	n=25	n=50	n=100	n=250	n=500	n=1000	n=2500	n=5000
Mittelwert	7,800	7,100	8,400	8,040	7,540	6,912	6,974	6,945	6,974	6,997
Sigma	2,490	2,885	2,363	2,490	2,435	2,565	2,485	2,422	2,432	2,429
<Z>	7,000	7,000	7,000	7,000	7,000	7,000	7,000	7,000	7,000	7,000